

Risk Forecasting with GARCH, Skewed t Distributions, and Multiple Timescales

Alec N. Kercheval

Department of Mathematics
Florida State University

Yang Liu

Department of Mathematics
Florida State University

DRAFT of 8/29/2010

Abstract

Historical time series of asset returns are commonly used to derive forecasts of risk, such as value at risk (VaR). Provided there is enough data, this can be done successfully even though asset returns are typically heavy-tailed, heteroskedastic, and serially dependent. We describe how the historical data can first be GARCH filtered and then used to calibrate parameters of the heavy-tailed skewed t distribution. Sufficient recent data is available if the forecasting horizon is short enough, for example for daily VaR forecasts. When the horizon is weekly or monthly, however, a sufficiently long weekly or monthly returns series extends too far into the past to be practical.

To address this we introduce a multiple timescale approach, where risk forecasts at a longer timescale, such as weekly or monthly, can be made with the more abundant data available at a shorter timescale, such as daily or weekly. The method is analyzed both theoretically and empirically using the last few decades of daily S&P500 returns. Since this method is not tied to a particular timescale, it can be used as well for intraday data; we illustrate with a set of 1-minute bond index futures returns.

The advantages of this multiscale approach are that it is more adaptable to longer horizons, it increases the quality of forecasts by virtue of increasing the number of recent observations that can be used, and it makes risk

forecasts more quickly reactive to recent events that occur late in the most recent period.

Contents

1	Introduction	3
2	The Skewed t Distributions	5
2.1	Normal Mean-Variance Mixture Distributions	5
2.2	Skewed t : A Special Case of GH Distributions	6
2.3	The EM Algorithm	11
3	Risk Forecasts on a Fixed Timescale	15
3.1	Value at Risk	16
3.2	Data and Stylized Facts	16
3.3	GARCH Filter	17
3.4	VaR Forecasting	23
3.5	Drawbacks of the Fixed Frequency Approach	23
4	Multiple Timescale Forecasts	27
4.1	Simulation of the GARCH Sum	28
4.2	Confidence Intervals for the Forecasts	29
5	Backtesting	30
5.1	Independence of Violation Indicators	30
5.2	Backtest Algorithm	31
5.3	Statistical Tests	32
5.4	n -Day Horizon	34
5.5	A Variant Backtest: Forecasting a One Day Return, n Days Ahead	40
5.6	A Note on the Monthly Forecast Horizon	44
5.7	Stability of the Likelihood Ratio Test and Algorithm Speed	46
5.8	A Quick Look at Intraday Data	49
6	Further Analysis: Long-Term GARCH and Comparisons using Simulated Data	55
6.1	Long-Term Behavior of GARCH	55

6.2	High-Low vs Fixed Frequency with Abundant Data	58
6.3	A Simulated Weekly Returns Scenario	64
6.4	A Simulated Daily Returns Scenario	68
7	Conclusion	70

1 Introduction

This article is about forecasting risk. The vague word “risk” refers to the degree of future variability of a quantity of interest, such as price return. A risk model is a quantitative approach to making a numerical risk forecast based on observed data, and such models are central to the practice of investing.

The classical risk forecast, as developed by [Markowitz \(1952\)](#) and [Sharpe \(1964\)](#), is a forecast of the standard deviation (StD) of portfolio return at a fixed time horizon, but there are several other measures of risk in common use, such as value-at-risk (VaR), expected shortfall (ES), and others (see [Artzner et al. \(1999\)](#); [Rockafellar and Uryasev \(2002\)](#)). Each of these is a kind of measure of the width of a probability density function describing the future return. Ultimately, the real underlying risk forecast is a forecast of the full probability distribution of returns, from which any numerical risk measure is determined.

The historical emphasis on a single number to measure risk has tended to hide the fact that most risk models in fact implicitly generate the forecast of a full distribution, and therefore represent an implicit choice of a family of distributions from which the forecast is to be made. This choice is difficult to avoid, even if it is not explicit.

For example, given a historical time series of monthly returns for a stock index, one could compute the sample standard deviation over the history and use that as a forecast for the coming month. However, this is approximately equivalent to a maximum likelihood fitting of the data to a Normal distribution, and therefore implicitly uses a Normal model for the returns distribution forecast.

It is now well-acknowledged that financial returns are poorly described by Normal distributions, even in one dimension, because of the prevalence of extreme outcomes (fat tails of the probability density function). What other choices do we have? The empirical distribution defined by the data is usually a poor choice because it has inadequate tail behavior. However, there are heavy-tailed parametric families in common use now, such as Variance Gamma, Hyperbolic, Student t , skewed t , and Normal Inverse Gaussian. (These will be defined later.)

See [Hu and Kercheval \(2007, 2008, 2010\)](#); [Hu \(2005\)](#); [McNeil et al. \(2005\)](#); [Aas and Hobaek Haff \(2006\)](#); [Keel and Geering \(2006\)](#).

Two questions immediately arise for these more complicated distributions. Does it matter for practical risk management? And, is it computationally practical to use these families?

The answer to both questions is yes. For example, the composition of portfolios on the efficient frontier, whether risk is measured via StD, VaR, or ES, depends on the choice of distribution family. Moreover, the use of these heavier-tailed distributions leads to much better fit of the returns data, and has become practical with the application of the EM algorithm (described below) to the maximum likelihood problem.

The method of [Hu and Kercheval \(2007\)](#), in brief, is to use a GARCH filter to remove serial dependence in a financial returns series, fit the filtered returns to a heavy-tailed distribution family using the EM algorithm, and then de-filter the resulting pdf to get a risk forecast conditional on current information.

This works well, but a drawback of this method is that a relatively large amount of data is required for numerical stability of the estimate – in our experiments, around 750 to 1000 observations are required for reliable results. This is fine for daily returns, but impractical for a monthly risk forecast. Here we address this difficulty by introducing a way to use higher frequency data, which is in more plentiful supply, to estimate risk on a lower frequency horizon. In this way, weekly or monthly risk forecasts can be made with daily or weekly data. The method we describe applies to any frequency, so we also illustrate it briefly with intraday data at high frequency.

The article is organized as follows. In [Section 2](#) we define a useful general class of probability distributions called the Generalized Hyperbolic (GH) distributions. These include as special cases the Variance Gamma, Hyperbolic, Normal Inverse Gaussian, Student t , and skewed t distributions; we focus primarily on the last of these in the remainder of the article. We discuss in detail how the distribution parameters can be estimated from data using the EM algorithm.

[Section 3](#) describes the GARCH filter and how it can be used to forecast risk on a fixed timescale, with a specific emphasis on VaR (though the methods apply equally well to StD, ES, or any other common risk measure). In [Section 4](#) we introduce a method for using high frequency data to forecast risk on a lower frequency horizon, using the GARCH methodology. The method proves viable through backtesting in [Section 5](#), which includes multiple-day and monthly VaR forecasts as well as some experiments with high frequency intra-day data. [Section 6](#) provides some further discussion of the long term behavior of the GARCH

process, along with additional analysis of the comparison between the multiscale approach and the fixed scale approach.

2 The Skewed t Distributions

The Generalized Hyperbolic (GH) distributions are becoming well-used to describe financial data. This family of probability distributions was introduced in [Barndorff-Nielsen \(1977, 1978\)](#) and further explored in [Barndorff-Nielsen and Blæsild \(1981\)](#). See also [McNeil et al. \(2005\)](#). It includes as subfamilies or limiting subfamilies many of the popular distributions in current modeling use, including Gaussian, Student t , skewed t , Variance Gamma, Normal Inverse Gaussian, and Hyperbolic distributions. Our primary interest is in the skewed t distribution because of its efficient fitting of equity returns data (see [Hu and Kercheval \(2007, 2010\)](#)). We first describe the GH distributions and some of their properties, and then specialize to the skewed t distributions. A detailed description of parameter estimation using the EM algorithm follows.

2.1 Normal Mean-Variance Mixture Distributions

Definition 2.1. *The random vector \mathbf{X} is said to have a (multivariate) normal mean-variance mixture distribution if*

$$\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W}A\mathbf{Z}, \quad (2.1)$$

where

- (i) $\mathbf{Z} \sim N_k(\mathbf{0}, I_k)$, the standard k -dimensional normal distribution;
- (ii) $W \geq 0$ is a nonnegative, scalar-valued random variable (r.v.) independent of \mathbf{Z} ;
- (iii) $A \in \mathbb{R}^{d \times k}$ is a matrix of constants, and
- (iv) $\boldsymbol{\mu} \in \mathbb{R}^d$ and $\boldsymbol{\gamma} \in \mathbb{R}^d$ are vectors of constants.

From the definition, we can see that

$$\mathbf{X}|W \sim N_d(\boldsymbol{\mu} + W\boldsymbol{\gamma}, W\Sigma), \quad (2.2)$$

where $\Sigma = AA'$, the covariance of $N(0, A\mathbf{Z})$. This is also why the distribution is called a normal mean-variance mixture. Simple calculation yields the following moment formulas

$$E(\mathbf{X}) = E(E(\mathbf{X}|W)) = \boldsymbol{\mu} + E(W)\boldsymbol{\gamma}, \quad (2.3)$$

$$\text{cov}(\mathbf{X}) = E(\text{cov}(\mathbf{X}|W)) + \text{cov}(E(\mathbf{X}|W)) = E(W)\Sigma + \text{var}(W)\boldsymbol{\gamma}\boldsymbol{\gamma}'. \quad (2.4)$$

In the context of modeling risk-factor returns, the mixing variable W can be interpreted as a shock that arises from new information and impacts the mean and volatility of stocks.

2.2 Skewed t: A Special Case of GH Distributions

Definition 2.2. Modified Bessel Function of the Third Kind. *The modified Bessel function of the third kind with index λ is defined by the integral*

$$K_\lambda(x) = \frac{1}{2} \int_0^\infty y^{\lambda-1} e^{-\frac{x}{2}(y+y^{-1})} dy, \quad x > 0. \quad (2.5)$$

When $\lambda < 0$, the following asymptotic property of the Bessel function is useful for computing the limiting density of GIG and GH,

$$K_\lambda(x) \sim \Gamma(-\lambda)2^{-\lambda-1}x^\lambda \quad \text{as } x \rightarrow 0^+. \quad (2.6)$$

Definition 2.3. The Generalized Inverse Gaussian Distribution (GIG). *The random variable X has a generalized inverse Gaussian (GIG) distribution, written $X \sim N^-(\lambda, \chi, \psi)$, if its density is*

$$h(x; \lambda, \chi, \psi) = \frac{\chi^{-\lambda}(\sqrt{\chi\psi})^\lambda}{2K_\lambda(\sqrt{\chi\psi})} x^{\lambda-1} \exp\left(-\frac{1}{2}(\chi x^{-1} + \psi x)\right), \quad x > 0, \quad (2.7)$$

where K_λ is a modified Bessel function of the third kind with index λ and the parameters satisfy

$$\begin{cases} \chi > 0, \psi \geq 0 & \text{if } \lambda < 0; \\ \chi > 0, \psi > 0 & \text{if } \lambda = 0; \\ \chi \geq 0, \psi > 0 & \text{if } \lambda > 0. \end{cases}$$

The following formulas for a GIG distributed r.v. X will be used later,

$$E(X^\alpha) = \left(\frac{\chi}{\psi}\right)^{\alpha/2} \frac{K_{\lambda+\alpha}(\sqrt{\chi\psi})}{K_\lambda(\sqrt{\chi\psi})} \quad (2.8)$$

and

$$E(\log X) = \frac{dE(X^\alpha)}{d\alpha} \Big|_{\alpha=0}, \quad (2.9)$$

where (2.9) needs to be evaluated numerically. More details on the properties of GIG can be found in Jørgensen (1982).

Theorem 2.4. The Generalized Hyperbolic Distribution (GH). *If a random vector \mathbf{X} has a normal mean-variance mixture distribution $\mathbf{X} \stackrel{d}{=} \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W}A\mathbf{Z}$ and the mixing variable $W \sim N^-(\lambda, \chi, \psi)$, then \mathbf{X} is said to have a generalized hyperbolic (GH) distribution, denoted by $\mathbf{X} \sim GH_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$. Its density is given by*

$$f(\mathbf{x}) = c \frac{K_{\lambda-\frac{d}{2}}\left(\sqrt{[\chi + (\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})](\psi + \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})}\right) e^{(\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}}}{\left(\sqrt{[\chi + (\mathbf{x} - \boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x} - \boldsymbol{\mu})](\psi + \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})}\right)^{\frac{d}{2}-\lambda}}, \quad (2.10)$$

where the normalizing constant c is

$$c = \frac{(\sqrt{\chi\psi})^{-\lambda} \psi^\lambda (\psi + \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})^{\frac{d}{2}-\lambda}}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} K_\lambda(\sqrt{\chi\psi})}$$

and $|\cdot|$ denotes the determinant.

Proof. From the definition of the normal mean-variance mixture distribution, the density of \mathbf{X} is given by

$$\begin{aligned} f(\mathbf{x}) &= \int_0^\infty f_{\mathbf{X}|W}(\mathbf{x}|w) h(w) dw \\ &= \int_0^\infty \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} w^{\frac{d}{2}}} \exp\left\{-\frac{(\mathbf{x} - \boldsymbol{\mu} - w\boldsymbol{\gamma})'(w\Sigma)^{-1}(\mathbf{x} - \boldsymbol{\mu} - w\boldsymbol{\gamma})}{2}\right\} h(w) dw, \end{aligned}$$

where $h(w)$ is the density of W .

$f(\mathbf{x})$ can be rewritten as

$$f(\mathbf{x}) = \int_0^\infty \frac{e^{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}w^{\frac{d}{2}}} \exp\left\{-\frac{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}{2w} - \frac{\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}}{2/w}\right\} h(w)dw. \quad (2.11)$$

Using (2.7) and some rearrangements, we get

$$f(\mathbf{x}) = \frac{(\sqrt{\chi\psi})^{-\lambda}\psi^\lambda e^{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}}}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}K_\lambda(\sqrt{\chi\psi})} \times \frac{1}{2} \int_0^\infty w^{\lambda-\frac{d}{2}-1} \exp\left\{-\frac{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}) + \chi}{2w} - \frac{\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma} + \psi}{2/w}\right\} dw.$$

By setting

$$y = w \frac{\sqrt{\psi + \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma}}}{\sqrt{\chi + (\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu})}} \quad (2.12)$$

and further rearrangements, we obtain

$$f(\mathbf{x}) = c \frac{e^{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}}}{\left(\sqrt{(\chi + (\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}))(\psi + \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})}\right)^{\frac{d}{2}}} \times \frac{1}{2} \int_0^\infty y^{\lambda-\frac{d}{2}-1} \exp\left\{-\frac{1}{2}\sqrt{(\chi + (\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}(\mathbf{x}-\boldsymbol{\mu}))(\psi + \boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})} \left[\frac{1}{y} + y\right]\right\} dy.$$

By (2.5), we can get the density of GH distributions. \square

For a d -dimensional normal random variable $\mathbf{X} \sim N_d(\boldsymbol{\mu}, \Sigma)$, it is well known that its characteristic function is

$$\phi_{\mathbf{X}}(\mathbf{t}) = E(e^{i\mathbf{t}'\mathbf{X}}) = \exp(i\mathbf{t}'\boldsymbol{\mu} - \frac{\mathbf{t}'\Sigma\mathbf{t}}{2}). \quad (2.13)$$

From the mean-variance mixture definition, we obtain the characteristic function of the GH random variable $\mathbf{X} \sim GH_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$:

$$\begin{aligned} \phi_{\mathbf{X}}(\mathbf{t}) &= E(E(\exp(i\mathbf{t}'\mathbf{X})|W)) = E(\exp(i\mathbf{t}'\boldsymbol{\mu} + W\mathbf{t}'\boldsymbol{\gamma} - \frac{1}{2}W\mathbf{t}'\Sigma\mathbf{t})) \\ &= \exp(i\mathbf{t}'\boldsymbol{\mu})\widehat{H}(\mathbf{t}'\Sigma\mathbf{t}/2 - i\mathbf{t}'\boldsymbol{\gamma}), \end{aligned}$$

where $\widehat{H}(\theta) = E(e^{-\theta W})$ is the Laplace transform of the density function h of W .

With the help of the characteristic function, we can show that GH distributions are closed under linear transformations.

Proposition 2.5. *If $\mathbf{X} \sim GH_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ and $\mathbf{Y} = B\mathbf{X} + \mathbf{b}$, where $B \in \mathbb{R}^{k \times d}$ and $\mathbf{b} \in \mathbb{R}^k$, then $\mathbf{Y} \sim GH_k(\lambda, \chi, \psi, B\boldsymbol{\mu} + \mathbf{b}, B\Sigma B', B\boldsymbol{\gamma})$.*

Proof.

$$\phi_{\mathbf{Y}}(\mathbf{t}) = E(e^{it'(B\mathbf{X}+\mathbf{b})}) = e^{it'\mathbf{b}}\phi_{\mathbf{X}}(B'\mathbf{t}) = e^{it'(B\boldsymbol{\mu}+\mathbf{b})}\widehat{H}(\mathbf{t}'B\Sigma B'\mathbf{t}/2 - it'B\boldsymbol{\gamma}).$$

□

This proposition shows that linear transformations of GH distributions remain in the class of GH distributions generated by the same GIG distribution $N^-(\lambda, \chi, \psi)$, which is a useful property in portfolio management.

Corollary 2.6. *If $B = \boldsymbol{\omega}' = (\omega_1, \dots, \omega_d)'$ and $\mathbf{b} = \mathbf{0}$, then $y = \boldsymbol{\omega}'\mathbf{X}$ is a one dimensional GH distribution, and*

$$y \sim GH_1(\lambda, \chi, \psi, \boldsymbol{\omega}'\boldsymbol{\mu}, \boldsymbol{\omega}'\Sigma\boldsymbol{\omega}, \boldsymbol{\omega}'\boldsymbol{\gamma}).$$

More specifically, the margins of \mathbf{X} is

$$X_i \sim GH_1(\lambda, \chi, \psi, \mu_i, \Sigma_{ii}, \gamma_i).$$

This corollary shows that the method used in portfolio risk management based on multivariate normal distribution is also applicable to GH distributions.

When $\psi = 0$ and $\lambda < 0$, a GIG distribution becomes the so-called inverse gamma distribution and the corresponding limiting case of the GH distribution is known as the skewed t distribution. Using the GH density and the asymptotic formula in (2.6), we can get the density of the skewed t distribution.

Definition 2.7. Skewed t Distribution. *If $\mathbf{X} \sim GH_d(\lambda, \chi, \psi, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$, $\lambda = -\frac{1}{2}\nu$, $\chi = \nu$ and $\psi = 0$, X is of skewed t distribution, denoted as $SkewT(\nu, \mu, \gamma, \sigma)$. Its density is given by*

$$f(\mathbf{x}) = c \frac{K_{\frac{\nu+d}{2}} \left(\sqrt{(\nu + \rho_x)(\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})} \right) e^{(\mathbf{x}-\boldsymbol{\mu})'\Sigma^{-1}\boldsymbol{\gamma}}}{\left(\sqrt{(\nu + \rho_x)(\boldsymbol{\gamma}'\Sigma^{-1}\boldsymbol{\gamma})} \right)^{\frac{\nu+d}{2}} \left(1 + \frac{\rho_x}{\nu} \right)^{\frac{\nu+d}{2}}}, \quad (2.14)$$

where the normalizing constant c is

$$c = \frac{2^{1-\frac{\nu+d}{2}}}{\Gamma(\frac{\nu}{2})(\pi\nu)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}} \quad (2.15)$$

and

$$\rho_x = (\mathbf{x} - \boldsymbol{\mu})' \Sigma^{-1} (\mathbf{x} - \boldsymbol{\mu}). \quad (2.16)$$

The mean and covariance of a skewed t distributed random variable \mathbf{X} are

$$E(\mathbf{X}) = \boldsymbol{\mu} + \gamma \frac{\nu}{\nu - 2}, \quad (2.17)$$

$$\text{cov}(\mathbf{X}) = \frac{\nu}{\nu - 2} \Sigma + \gamma \gamma' \frac{2\nu^2}{(\nu - 2)^2(\nu - 4)}, \quad (2.18)$$

where the covariance matrix is only defined when $\nu > 4$.

Moreover, when $\gamma = 0$, the skewed t distribution degenerates into the Student t distribution.

As implied by its name, an inverse gamma random variable is the inverse of a gamma random variable. Together with the mean-variance mixture definition, we can generate a skewed t random variable accordingly.

Algorithm 2.8. Simulation of the Skewed t Distribution.

1. Generate Y from a $Gamma(\frac{\nu}{2}, \frac{\nu}{2})$ distribution.
2. Set $W = Y^{-1}$. By definition, $W \sim InverseGamma(\frac{\nu}{2}, \frac{\nu}{2})$.
3. Generate a d -dimensional normal random vector $\mathbf{Z} \sim N_d(\mathbf{0}, I_d)$
4. Let

$$\mathbf{X} = \boldsymbol{\mu} + W\boldsymbol{\gamma} + \sqrt{W}A\mathbf{Z}.$$

Then $X \sim SkewT(\nu, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$.

Other subfamilies of the GH distribution include:

Hyperbolic Distributions:

If $\lambda = (d + 1)/2$, we refer to the distribution as a d -dimensional hyperbolic distribution.

If $\lambda = 1$, we get the multivariate distribution whose univariate margins are one-dimensional hyperbolic distributions.

Normal Inverse Gaussian Distributions (NIG):

If $\lambda = -1/2$, GIG becomes the inverse Gaussian distribution. The corresponding GH distribution is known as the normal inverse Gaussian (NIG) distribution.

Variance Gamma Distributions (VG):

If $\lambda > 0$ and $\chi = 0$, GIG becomes the gamma distribution. The corresponding GH limiting distribution is known as the variance gamma (VG) distribution.

2.3 The EM Algorithm

GH distributions can be fitted with an iterative procedure known as the EM (expectation-maximization) algorithm. To illustrate the idea behind it, we'll present a compact derivation of the formulas used in the EM algorithm for the skewed t distribution. Further details and an array of formulas for other subfamilies of the GH distributions can be found in [Hu \(2005\)](#).

Definition 2.9. Likelihood Function. Let $f(x|\theta_1, \dots, \theta_k)$ denote the probability density function (pdf) or the probability mass function (pmf, the discrete version of a pdf) of an i.i.d. sample X_1, \dots, X_n , with parameters $\theta = (\theta_1, \dots, \theta_k)$. Given an observation $\mathbf{x} = \{x_1, \dots, x_n\}$, the function of θ defined by

$$L(\theta|\mathbf{x}) = \prod_{i=1}^n f(x_i|\theta_1, \dots, \theta_k)$$

is called the likelihood function.

Definition 2.10. Maximum Likelihood Estimator (MLE). For each sample point \mathbf{x} , let $\hat{\theta}(\mathbf{x})$ be a parameter value at which $L(\theta|\mathbf{x})$ attains its maximum as a function of θ , with \mathbf{x} held fixed. Then $\hat{\theta}(\mathbf{x})$ is called a maximum likelihood estimator of the parameter θ based on a sample \mathbf{X} .

Assume we have i.i.d. data $\mathbf{x}_1, \dots, \mathbf{x}_n \in \mathbb{R}^d$ and want to fit a skewed t distribution. Summarize the parameters by $\boldsymbol{\theta} = (\nu, \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ and the problem is to maximize the log likelihood

$$\log L(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n) = \sum_{i=1}^n \log f_X(\mathbf{x}_i; \boldsymbol{\theta}), \quad (2.19)$$

where $f_X(\cdot; \boldsymbol{\theta})$ denotes the skewed t density function.

The problem looks formidable at first glance due to the number of parameters and the necessity of maximizing over covariance matrices Σ . However, if the latent mixing variables W_1, \dots, W_n were observable, the optimization would be much easier. The joint density of any pair \mathbf{X}_i and W_i is given by

$$f_{\mathbf{X},W}(\mathbf{x}, w; \boldsymbol{\theta}) = f_{\mathbf{X}|W}(\mathbf{x}|w; \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})h_W(w; \nu), \quad (2.20)$$

where $h_W(\cdot; \nu)$ is the density of $InverseGamma(\nu/2, \nu/2)$. We could then construct the augmented log-likelihood

$$\begin{aligned} \log \tilde{L}(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n, w_1, \dots, w_n) &= \sum_{i=1}^n \log f_X(\mathbf{x}_i, w_i; \boldsymbol{\theta}) \\ &= \sum_{i=1}^n \log f_{\mathbf{X}|W}(\mathbf{x}_i|w_i; \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}) + \sum_{i=1}^n \log h_W(w_i; \nu) \\ &= L_1(\boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}; \mathbf{x}_1, \dots, \mathbf{x}_n | w_1, \dots, w_n) + L_2(\nu; w_1, \dots, w_n), \end{aligned} \quad (2.21)$$

where $f_{\mathbf{X}|W}(\cdot|w_i; \boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma})$ is the density of the condition normal $N(\boldsymbol{\mu} + w\boldsymbol{\gamma}, w\Sigma)$. L_1 and L_2 could be maximized separately if the latent mixing variables were observable.

To overcome such latency, we maximize the *expected value* of the augmented-likelihood $\log \tilde{L}$ conditional on the observed data and a guess for the parameters $\boldsymbol{\theta}$. Such conditioning is necessary because the distribution of W depends on the parameters. Maximizing the expectation of $\log \tilde{L}$ produces an updated guess for $\boldsymbol{\theta}$, which we then use to repeat the procedure until convergence.

This can be summarized as an iterated two-step process consisting of an E-step and an M-step.

E-step: Compute an objective function

$$Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{[k]}) = E\left(\log \tilde{L}(\boldsymbol{\theta}; \mathbf{x}_1, \dots, \mathbf{x}_n, W_1, \dots, W_n) \middle| \mathbf{x}_1, \dots, \mathbf{x}_n; \boldsymbol{\theta}^{[k]}\right), \quad (2.22)$$

where $\boldsymbol{\theta}^{[k]}$ denotes the parameter estimate after the k th step.

M-step: Maximize Q with respect to $\boldsymbol{\theta}$ to get the updated estimate $\boldsymbol{\theta}^{[k+1]}$. Repeat.

Now we will derive the formulas necessary for the implementation of the EM algorithm.

Similar to (2.11), the density of the conditional normal distribution can be written as

$$f_{\mathbf{X}|W}(x|w) = \frac{1}{(2\pi)^{\frac{d}{2}} |\Sigma|^{\frac{1}{2}} w^{\frac{d}{2}}} e^{(\mathbf{x}-\boldsymbol{\mu})' \Sigma^{-1} \gamma - \frac{\rho}{2w} - \frac{w}{2} \gamma' \Sigma^{-1} \gamma}, \quad (2.23)$$

where ρ is the quadratic form defined in (2.16).

Therefore,

$$\begin{aligned} L_1(\boldsymbol{\mu}, \Sigma, \boldsymbol{\gamma}; \mathbf{x}_1, \dots, \mathbf{x}_n | w_1, \dots, w_n) = & \\ & - \frac{n}{2} \log |\Sigma| - \frac{d}{2} \sum_{i=1}^n \log w_i + \sum_{i=1}^n (\mathbf{x}_i - \boldsymbol{\mu})' \Sigma^{-1} \boldsymbol{\gamma} \\ & - \frac{1}{2} \sum_{i=1}^n \frac{\rho_i}{w_i} - \frac{1}{2} \boldsymbol{\gamma}' \Sigma^{-1} \boldsymbol{\gamma} \sum_{i=1}^n w_i. \end{aligned} \quad (2.24)$$

From (2.6) and (2.7),

$$\begin{aligned} L_2(\nu; w_1, \dots, w_n) = & \\ & - \left(\frac{\nu}{2} - 1\right) \sum_{i=1}^n \log w_i - \frac{\nu}{2} \sum_{i=1}^n w_i^{-1} - n \log \Gamma\left(\frac{\nu}{2}\right) + \frac{n\nu}{2} \log \frac{\nu}{2} \end{aligned} \quad (2.25)$$

From (2.24) and (2.25), it can be seen that computing the objective function $Q(\boldsymbol{\theta}; \boldsymbol{\theta}^{[k]})$ requires formulas for three types of quantities:

$$E(W_i | \mathbf{x}_i; \boldsymbol{\theta}^{[k]}), E(W_i^{-1} | \mathbf{x}_i; \boldsymbol{\theta}^{[k]}) \text{ and } E(\log W_i | \mathbf{x}_i; \boldsymbol{\theta}^{[k]}).$$

To calculate these conditional expectations, we compute the following conditional density function of W

$$f_{W|\mathbf{X}}(w|\mathbf{x}; \boldsymbol{\theta}) = \frac{f_{\mathbf{X}|W}(\mathbf{x}|w; \boldsymbol{\theta}) h_W(w)}{f_{\mathbf{X}}(\mathbf{x}; \boldsymbol{\theta})}. \quad (2.26)$$

By some algebra and (2.7), we can get

$$W_i | \mathbf{X}_i \sim N^{-}\left(-\frac{d+\nu}{2}, \rho_i + \nu, \boldsymbol{\gamma}' \Sigma^{-1} \boldsymbol{\gamma}\right). \quad (2.27)$$

For convenience, we'll use a standard notation of Liu and Rubin (1994), Protasov (2004) and McNeil et al. (2005).

$$\delta_i^{[1]} = E(W_i^{-1} | \mathbf{x}_i; \boldsymbol{\theta}^{[1]}), \quad \eta_i^{[1]} = E(W_i | \mathbf{x}_i; \boldsymbol{\theta}^{[1]}), \quad \xi_i^{[1]} = E(\log W_i | \mathbf{x}_i; \boldsymbol{\theta}^{[1]}), \quad (2.28)$$

and

$$\bar{\delta} = \frac{1}{n} \sum_{i=1}^n \delta_i, \quad \bar{\eta} = \frac{1}{n} \sum_{i=1}^n \eta_i, \quad \bar{\xi} = \frac{1}{n} \sum_{i=1}^n \xi_i \quad (2.29)$$

Using (2.8) and (2.9), we get

$$\delta_i^{[k]} = \left(\frac{A^{[k]}}{B^{[k]}} \right)^{-\frac{1}{2}} \frac{K_{\frac{\nu+d+2}{2}} \left(\sqrt{A^{[k]} B^{[k]}} \right)}{K_{\frac{\nu+d}{2}} \left(\sqrt{A^{[k]} B^{[k]}} \right)}, \quad (2.30)$$

$$\eta_i^{[k]} = \left(\frac{A^{[k]}}{B^{[k]}} \right)^{\frac{1}{2}} \frac{K_{\frac{\nu+d-2}{2}} \left(\sqrt{A^{[k]} B^{[k]}} \right)}{K_{\frac{\nu+d}{2}} \left(\sqrt{A^{[k]} B^{[k]}} \right)}, \quad (2.31)$$

$$\xi_i^{[k]} = \frac{1}{2} \log \left(\frac{A^{[k]}}{B^{[k]}} \right) \frac{\frac{\partial K_{-\frac{\nu+d}{2}+\alpha} \left(\sqrt{A^{[k]} B^{[k]}} \right)}{\partial \alpha} \Big|_{\alpha=0}}{K_{\frac{\nu+d}{2}} \left(\sqrt{A^{[k]} B^{[k]}} \right)}, \quad (2.32)$$

where

$$A^{[k]} = \rho_i^{[k]} + \nu^{[k]} \quad \text{and} \quad B^{[k]} = \gamma^{[k]'} \Sigma^{[k]-1} \gamma^{[k]}.$$

In the M-step, L_1 can be maximized by taking its partial derivative with respect to $\boldsymbol{\mu}$, Σ and $\boldsymbol{\gamma}$, and set

$$\frac{\partial L}{\partial \boldsymbol{\mu}} = \mathbf{0}, \quad \frac{\partial L}{\partial \Sigma} = \mathbf{0} \quad \text{and} \quad \frac{\partial L}{\partial \boldsymbol{\gamma}} = \mathbf{0}. \quad (2.33)$$

Solving the above equation array, we get the following estimates:

$$\boldsymbol{\gamma} = \frac{n^{-1} \sum_{i=1}^n w_i^{-1} (\bar{\boldsymbol{x}} - \boldsymbol{x}_i)}{n^{-2} (\sum_{i=1}^n w_i) (\sum_{i=1}^n w_i^{-1}) - 1}, \quad (2.34)$$

$$\boldsymbol{\mu} = \frac{n^{-1} \sum_{i=1}^n w_i^{-1} \boldsymbol{x}_i - \boldsymbol{\gamma}}{n^{-1} \sum_{i=1}^n w_i^{-1}}, \quad (2.35)$$

$$\Sigma = \frac{1}{n} \sum_{i=1}^n w_i^{-1} (\boldsymbol{x}_i - \boldsymbol{\mu})(\boldsymbol{x}_i - \boldsymbol{\mu})' - \frac{1}{n} \sum_{i=1}^n w_i \boldsymbol{\gamma} \boldsymbol{\gamma}'. \quad (2.36)$$

Setting $\frac{\partial L_2}{\partial \nu} = 0$, we get the following equation:

$$-\frac{\Gamma' \left(\frac{\nu}{2} \right)}{\Gamma \left(\frac{\nu}{2} \right)} + \log \left(\frac{\nu}{2} \right) + 1 - \frac{1}{n} \sum_{i=1}^n w_i^{-1} - \frac{1}{n} \sum_{i=1}^n \log w_i = 0, \quad (2.37)$$

which can be solved numerically to get the estimate of ν .

Applying the δ, η, ξ notation to (2.34), (2.35), (2.36) and (2.37), we get the detailed algorithm.

Algorithm 2.11. EM Algorithm for Skewed t Distributions.

1. Set the iteration counter $k=1$ and select starting values for $\theta^{[1]}$. Reasonable starting values for $\boldsymbol{\mu}$, Σ and $\boldsymbol{\gamma}$ are the sample mean, the sample covariance matrix and the zero vector respectively.
2. Compute $\delta_i^{[k]}$, $\eta_i^{[k]}$ and $\xi_i^{[k]}$ and their averages $\bar{\delta}^{[k]}$, $\bar{\eta}^{[k]}$ and $\bar{\xi}^{[k]}$ using (2.30), (2.31) and (2.32).
3. Update $\boldsymbol{\gamma}$, $\boldsymbol{\mu}$ and Σ according to

$$\boldsymbol{\gamma}^{[k+1]} = \frac{n^{-1} \sum_{i=1}^n \delta_i^{[k]} (\bar{\boldsymbol{x}} - \boldsymbol{x}_i)}{\bar{\delta}^{[k]} \bar{\eta}^{[k]} - 1}, \quad (2.38)$$

$$\boldsymbol{\mu}^{[k+1]} = \frac{n^{-1} \sum_{i=1}^n \delta_i^{[k]} \boldsymbol{x}_i - \boldsymbol{\gamma}^{[k+1]}}{\bar{\delta}^{[k]}}, \quad (2.39)$$

$$\Sigma^{[k+1]} = \frac{1}{n} \sum_{i=1}^n \delta_i^{[k]} (\boldsymbol{x}_i - \boldsymbol{\mu}^{[k+1]})(\boldsymbol{x}_i - \boldsymbol{\mu}^{[k+1]})' - \bar{\eta}^{[k]} \boldsymbol{\gamma}^{[k+1]} \boldsymbol{\gamma}^{[k+1]}' . \quad (2.40)$$

4. Compute $\nu^{[k+1]}$ by numerically solving the equation

$$-\frac{\Gamma'(\frac{\nu}{2})}{\Gamma(\frac{\nu}{2})} + \log\left(\frac{\nu}{2}\right) + 1 - \bar{\xi}^{[k]} - \bar{\delta}^{[k]} = 0. \quad (2.41)$$

5. Set counter $k=k+1$ and go back to step 2 unless the relative increment of log likelihood is below a given tolerance.

3 Risk Forecasts on a Fixed Timescale

In this section we describe the GARCH method for filtering time series data and for forecasting risk in the simplest case of a fixed timescale, such as is discussed in [Hu and Kercheval \(2010, 2007\)](#). The method allows us to forecast the full pdf of the returns distribution, but for simplicity and concreteness we focus here on forecasting VaR; other kinds of risk forecasts will be similar.

3.1 Value at Risk

Definition 3.1. Value at Risk (VaR). Given $\alpha \in (0, 1)$, the value-at-risk at confidence level α for loss L of a security or a portfolio is defined as

$$VaR_\alpha(L) = \inf\{l \in \mathbb{R} : F_L(l) \geq \alpha\},$$

where F_L is the cumulative distribution function of L .

In probabilistic terms, VaR is a quantile of the loss distribution. Typical values for α are between 0.95 and 0.995. VaR can also be based on returns instead of losses, in which case α are small values like 0.05 or 0.01.

For example, intuitively, a 95% value at risk, $VaR_{0.95}$, is a level L such that a loss exceeding L has only a 5% chance of occurring.

3.2 Data and Stylized Facts

Given a set of daily closing prices for some index, we first convert them into negative log returns and then would like to calibrate a skewed t distribution with the EM algorithm. However, there's another complication: financial data are not i.i.d. and the maximum likelihood method is not yet applicable.

Financial time series, such as log-returns on equities, indexes, exchange rates and commodity prices, are studied extensively and a collection of deeply entrenched empirical observations and inferences have been established. Some of these so called "stylized facts" include:

- (1) Returns series are not i.i.d. though they show little serial correlation.
- (2) Series of absolute or squared returns show profound serial correlation.
- (3) Conditional expected returns are close to zero.
- (4) Volatility appears to vary over time and is clustered.
- (5) Returns series are leptokurtic and skewed.

These facts persist on all time intervals ranging from intraday returns to weekly or monthly returns.

In this paper our primary data set consists of daily returns of the S&P500 index (based on adjusted daily closing prices) from January 1, 1991 to December 31, 2009, about 4750 observations.

Fact (1) and (2) can be illustrated through Figure 1.

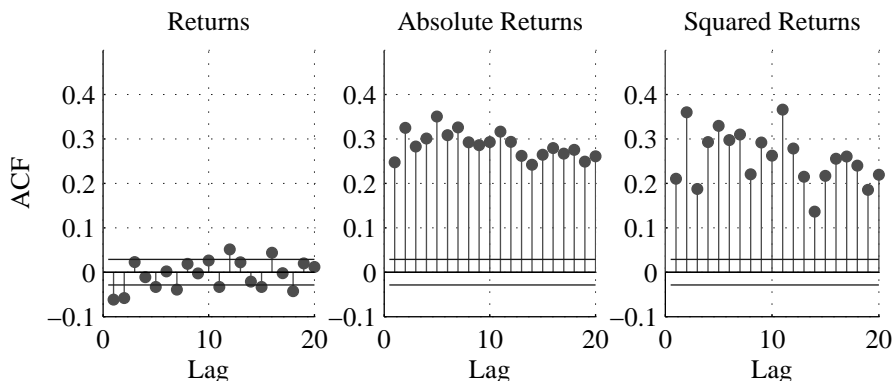


Figure 1: ACF for Unfiltered S&P500 Daily Returns Data, 1991-2009.

3.3 GARCH Filter

A GARCH (generalized autoregressive conditionally heteroskedastic) model can be used to filter the negative log returns into an approximately i.i.d. series. We provide some of the essentials of the classical time series analysis that are most relevant to the GARCH model. A more comprehensive summary based on standard texts like [Brockwell and Davis \(2002\)](#) can be found in [McNeil et al. \(2005\)](#). In the following, \mathbb{Z} denotes either the positive integers or the non-negative integers.

Definition 3.2. Covariance Stationarity. A sequence of r.v.'s $(X_t)_{t \in \mathbb{Z}}$ is covariance stationary if the first two moments exist and satisfy

$$E(X_t) = \mu, \quad t \in \mathbb{Z},$$

$$E(X_t X_s) = E(X_{t+k} X_{s+k}), \quad t, s, k \in \mathbb{Z}.$$

Definition 3.3. Strict White Noise. $(X_t)_{t \in \mathbb{Z}}$ is a strict white noise process if it is a sequence of i.i.d. r.v.'s with finite variance. A strict white noise process with mean 0 and variance σ^2 is denoted as $SWN(0, \sigma^2)$.

Definition 3.4. Martingale Difference. $(X_t)_{t \in \mathbb{Z}}$ is a martingale difference sequence with respect to the filtration $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$ if $E|X_t| < \infty$, X_t is \mathcal{F}_t -measurable and

$$E(X_t | \mathcal{F}_{t-1}) = 0, \forall t \in \mathbb{Z}.$$

The unconditional mean of such a process is also zero:

$$E(X_t) = E(E(X_t | \mathcal{F}_{t-1})) = 0.$$

Moreover, if $E(X_t^2) < \infty$, then autocovariances satisfy

$$E(X_t X_s) = \begin{cases} E(E(X_t X_s | \mathcal{F}_{s-1})) = E(X_t E(X_s | \mathcal{F}_{s-1})) = 0, & t < s, \\ E(E(X_t X_s | \mathcal{F}_{t-1})) = E(X_s E(X_t | \mathcal{F}_{t-1})) = 0, & t > s. \end{cases} \quad (3.1)$$

Thus a finite-variance martingale-difference process has zero mean and zero covariance. If the variance is constant for all t , the process is covariance stationary.

Definition 3.5. GARCH(p,q) Process. Let $(Z_t)_{t \in \mathbb{Z}}$ be $SWN(0, 1)$. The process $(X_t)_{t \in \mathbb{Z}}$ is a $GARCH(p, q)$ process if it satisfies the following equations:

$$X_t = \sigma_t Z_t, \quad (3.2)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^p \alpha_i X_{t-i}^2 + \sum_{j=1}^q \beta_j \sigma_{t-j}^2, \quad (3.3)$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \dots, p$ and $\beta_j \geq 0$, $j = 1, \dots, q$, are constants, and Z_t is independent of $(X_s)_{s < t}$ for each t . The r.v.'s Z_t are called the innovations.

Let $\mathcal{F}_t = \sigma\{X_s : s \leq t\}$ denote the sigma algebra representing the history of the process up to time t so that $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$ is the natural filtration. It can be easily verified that the GARCH process has the martingale-difference property with respect to $\{\mathcal{F}_t\}_{t \in \mathbb{Z}}$. Zero covariance implies zero autocorrelation, which suits stylized fact (1). We'll further show that it has constant variance and is therefore covariance stationary.

In practice, low-order GARCH models are most widely used and their mathematical analysis is relatively straightforward. We'll concentrate on the $GARCH(1,1)$ process.

It follows from (3.2) and (3.3) that for a $GARCH(1,1)$ process we have

$$\sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta \sigma_{t-1}^2 = \alpha_0 + (\alpha_1 Z_{t-1}^2 + \beta) \sigma_{t-1}^2. \quad (3.4)$$

This is a stochastic recurrence equation (SRE) of the form $Y_t = A_t Y_{t-1} + B_t$, where $(A_t)_{t \in \mathbb{Z}}$ and $(B_t)_{t \in \mathbb{Z}}$ are i.i.d. r.v.'s. As shown by [Brandt \(1986\)](#), sufficient conditions for a solution are that

$$E(\max\{0, \log |B_t|\}) < \infty \text{ and } E(\log |A_t|) < 0. \quad (3.5)$$

The unique solution is given by

$$Y_t = B_t + \sum_{i=1}^{\infty} B_{t-i} \prod_{j=0}^{i-1} A_{t-j}, \quad (3.6)$$

where the sum converges absolutely, almost surely.

We'll use these facts about SREs to derive the sufficient and necessary condition for the covariance stationarity of the GARCH(1,1) process.

Proposition 3.6. *The GARCH(1,1) process is a covariance-stationary process if and only if $\alpha_1 + \beta < 1$. The variance of the covariance-stationary process is given by $\alpha_0 / (1 - \alpha_1 - \beta)$.*

Proof. Assuming covariance stationarity, it follows from [\(3.4\)](#) and $E(Z_t^2) = 1$ that

$$\sigma^2 = E(\sigma_t^2) = \alpha_0 + (\alpha_1 E(Z_t^2) + \beta) E(\sigma_{t-1}^2) = \alpha_0 + (\alpha_1 + \beta) \sigma^2.$$

As a result, $\sigma^2 = \alpha_0 / (1 - \alpha_1 - \beta)$ and we must have $\alpha_1 + \beta < 1$ since $\alpha_0 > 0$.

Since Z_t is independent of $(X_s)_{s < t}$, σ_t and Z_t are independent immediately from [\(3.4\)](#). By [\(3.2\)](#) and such independence,

$$\text{var}(X_t) = E(X_t^2) = E(\sigma_t^2) E(Z_t^2) = \sigma^2 \cdot 1 = \alpha_0 / (1 - \alpha_1 - \beta).$$

Conversely, if $\alpha_1 + \beta < 1$, by Jensen's inequality

$$E(\log((\alpha_1 Z_{t-1}^2 + \beta) \sigma_{t-1}^2)) < \log(E((\alpha_1 Z_{t-1}^2 + \beta) \sigma_{t-1}^2)) = \log(\alpha_1 + \beta) < 0.$$

On the other hand, since α_0 is a constant,

$$E(\max\{0, \log |\alpha_0|\}) < \infty.$$

Now that both two sufficient conditions in (3.5) are satisfied, by (3.6), the solution to (3.4) is

$$\sigma_t^2 = \alpha_0 + \alpha_0 \sum_{i=1}^{\infty} \prod_{j=0}^{i-1} (\alpha_1 Z_{t-j}^2 + \beta). \quad (3.7)$$

Take expectation, then by $E(Z_t^2) = 1$,

$$\begin{aligned} E(\sigma_t^2) &= \alpha_0 + \alpha_0 \sum_{i=1}^{\infty} \prod_{j=0}^{i-1} (\alpha_1 E(Z_{t-j}^2) + \beta) \\ &= \alpha_0 + \alpha_0 \sum_{i=1}^{\infty} \prod_{j=0}^{i-1} (\alpha_1 + \beta) \\ &= \alpha_0 \sum_{i=0}^{\infty} (\alpha_1 + \beta)^i = \alpha_0 / (1 - \alpha_1 - \beta). \end{aligned}$$

□

While the GARCH(1,1) process has constant variance, its conditional variance is given by

$$\text{var}(X_t | \mathcal{F}_{t-1}) = E(\sigma_t^2 Z_t^2 | \mathcal{F}_{t-1}) = \sigma_t^2 E(Z_t^2) = \sigma_t^2 = \alpha_0 + \alpha_1 X_{t-1}^2 + \beta \sigma_{t-1}^2.$$

In other words, its conditional standard deviation σ_t , or volatility, is a continually changing function of both $|X_{t-1}|$ and σ_{t-1} . If one or both are particularly large, then X_t is effectively drawn from a distribution with large variance and more likely to be large itself. In this way, the model generates volatility clusters as described by stylized fact (4).

Assuming a GARCH(1,1) process has a fourth moment, we can calculate the kurtosis of X_t . Square both sides of (3.7), take expectations and by stationarity,

$$E(\sigma_t^4) = \alpha_0^2 + (\alpha_1^2 \kappa_Z + \beta^2 + 2\alpha_1 \beta) E(\sigma_t^4) + 2\alpha_0 (\alpha_1 + \beta) E(\sigma_t^2),$$

where $\kappa_Z = E(Z_t^4) / (E(Z_t^2))^2 = E(Z_t^4)$ denotes the kurtosis of the innovations. Solve for $E(\sigma_t^4)$ and using $E(\sigma_t^2) = E(X_t^2) = \alpha_0 / (1 - \alpha_1 - \beta)$, we'll get

$$E(X_t^4) = \kappa_Z E(\sigma_t^4) = \frac{\alpha_0^2 \kappa_Z (1 - (\alpha_1 + \beta)^2)}{(1 - \alpha_1 - \beta)^2 (1 - \alpha_1^2 \kappa_Z - \beta^2 - 2\alpha_1 \beta)},$$

from which it follows that

$$\kappa_X = \frac{\kappa_Z(1 - (\alpha_1 + \beta)^2)}{1 - \alpha_1^2\kappa_Z - \beta^2 - 2\alpha_1\beta}.$$

It can be seen that whenever $\kappa_Z > 1$, the kurtosis of X_t is inflated in comparison with that of Z_t . Therefore, the stationary distribution of the GARCH process is leptokurtic (i.e., $\kappa_X > 3$) for Gaussian or Student t innovations, capturing stylized fact (5).

Higher-order GARCH models have the same general behavior as GARCH(1,1). The necessary and sufficient condition for covariance stationarity is $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$ and the constant variance is $\alpha_0 / (1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j)$. For more details, see [Bougerol and Picard \(1992\)](#).

There are many variants on and extensions of the basic GARCH model, such as the following:

Definition 3.7. ARMA Process with GARCH Errors. *Let $(Z_t)_{t \in \mathbb{Z}}$ be $SWN(0, 1)$. The process $(X_t)_{t \in \mathbb{Z}}$ is an $ARMA(p_1, q_1)$ process with $GARCH(p_2, q_2)$ errors if it is covariance stationary and satisfies the equations:*

$$X_t = \mu_t + \sigma_t Z_t, \quad (3.8)$$

$$\mu_t = \mu_0 + \sum_{i=1}^{p_1} \phi_i (X_{t-i} - \mu) + \sum_{j=1}^{q_1} \theta_j (X_{t-j} - \mu_{t-j}), \quad (3.9)$$

$$\sigma_t^2 = \alpha_0 + \sum_{i=1}^{p_2} \alpha_i (X_{t-i} - \mu_{t-i})^2 + \sum_{j=1}^{q_2} \beta_j \sigma_{t-j}^2, \quad (3.10)$$

where $\sum_{i=1}^{p_2} \alpha_i + \sum_{j=1}^{q_2} \beta_j < 1$, $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \dots, p_2$ and $\beta_j \geq 0$, $j = 1, \dots, q_2$, and the innovation Z_t is independent of $(X_s)_{s < t}$ for all t .

Since return series show little serial correlation, the ARMA component is unnecessary, i.e., we choose $p_1 = q_1 = 0$. On the other hand, stylized fact (3) states that conditional expected returns are close but not equal to zero. Therefore, we retain the constant μ_0 in (3.9). We'll use such a standard GARCH(1,1) model with a constant non-zero mean to filter the negative log return process and get an approximately i.i.d. series. [Hu and Kercheval \(2007\)](#) demonstrated in this context that the traditional normal distribution is a poor candidate for the distribution of the filtered returns due to its thin tails. We adopt the more heavy-tailed skewed

t distribution, which has shown good results in comparison to common heavy-tailed alternatives. For the same reason, we choose Student t innovations when calibrating GARCH. Correlograms of the filtered series (Figure 2) are included to show the performance of the GARCH filter.

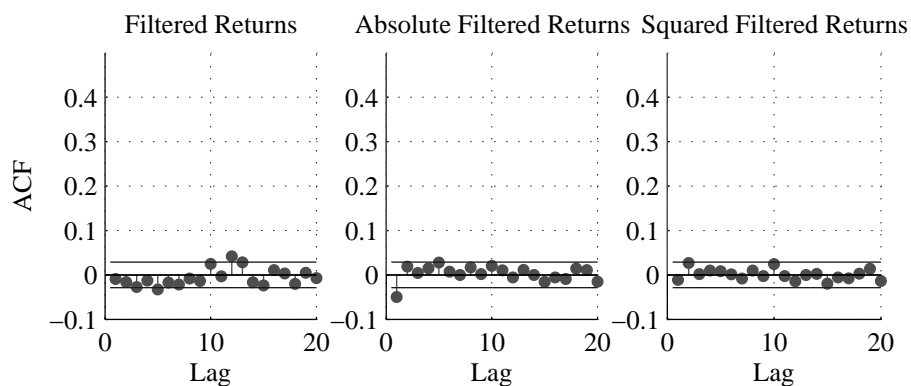


Figure 2: ACF for Filtered S&P500 Daily Returns Data, 1991 - 2009

In Figure 3, we plot $\{\sigma_t\}$ of the fitted GARCH process to illustrate stylized fact (4).

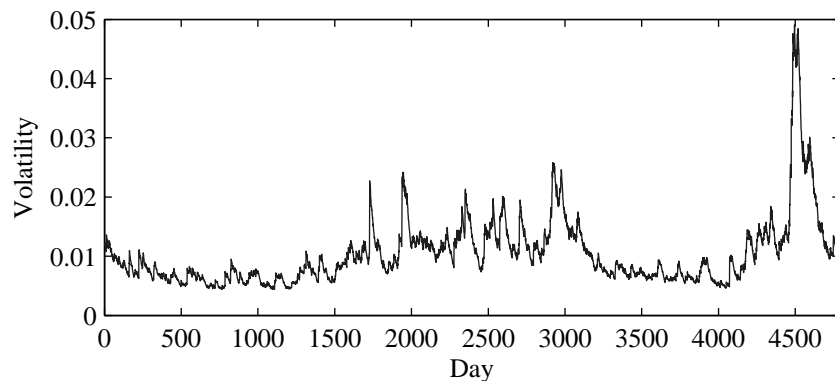


Figure 3: GARCH Volatility for S&P500 Daily Returns Data, 1991-2009

3.4 VaR Forecasting

Once the distribution of the filtered data is calibrated with the EM algorithm, VaR forecasts can be immediately derived via de-filtering. At time T , since σ_{T+1} is \mathcal{F}_T -measurable and Z_{T+1} independent of $\{X_t\}_{t \leq T+1}$ and thereby \mathcal{F}_T , the conditional VaR for time $T + 1$ will be

$$VaR_\alpha(X_{T+1}|\mathcal{F}_T) = VaR_\alpha(\sigma_{T+1}Z_{T+1}+\mu_0|\mathcal{F}_T) = \sigma_{T+1}VaR_\alpha(Z_{T+1})+\mu_0 = \sigma_{T+1}q_\alpha+\mu_0,$$

where q_α is the α -quantile of the calibrated skewed t distribution.

3.5 Drawbacks of the Fixed Frequency Approach

The EM algorithm uses a maximum likelihood method, which requires a sufficient amount of input data for accuracy. Therefore, as the time horizon grows, data availability can become an issue of major concern.

To investigate to what extent the EM algorithm depends on the sample size, we generate skewed t random variables with $\nu = 6.4$, $\mu = -0.14$, $\gamma = 0.12$ and $\sigma = 0.65$ ¹, and apply the EM algorithm to samples of different sizes. The calibration of each sample size is repeated 20 times and the performance is recorded in Table 1.A and 1.B.

Table 1.A: Dependence of EM Algorithm on Sample Size Part I

Sample Size	ν		μ		γ		σ	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
100	25.80	23.58	-0.536	0.977	0.491	0.947	0.650	0.118
250	10.45	7.64	-0.210	0.218	0.191	0.198	0.662	0.108
500	7.07	2.08	-0.134	0.147	0.116	0.111	0.641	0.064
750	7.14	2.01	-0.161	0.137	0.132	0.104	0.668	0.050
1000	7.15	1.44	-0.171	0.085	0.148	0.077	0.656	0.040
5000	6.55	0.50	-0.161	0.046	0.136	0.034	0.650	0.018

¹These values are the long-term averages from calibrating historical data.

Table 1.B: Dependence of EM Algorithm on Sample Size Part II

Sample Size	Performance		Number of Iterations	
	Blowup	Success	Mean	Std
100	10	10	65	157.85
250	4	16	82	120.41
500	1	19	76	44.76
750	1	19	73	45.65
1000	0	20	73	37.76
5000	0	20	62	16.42

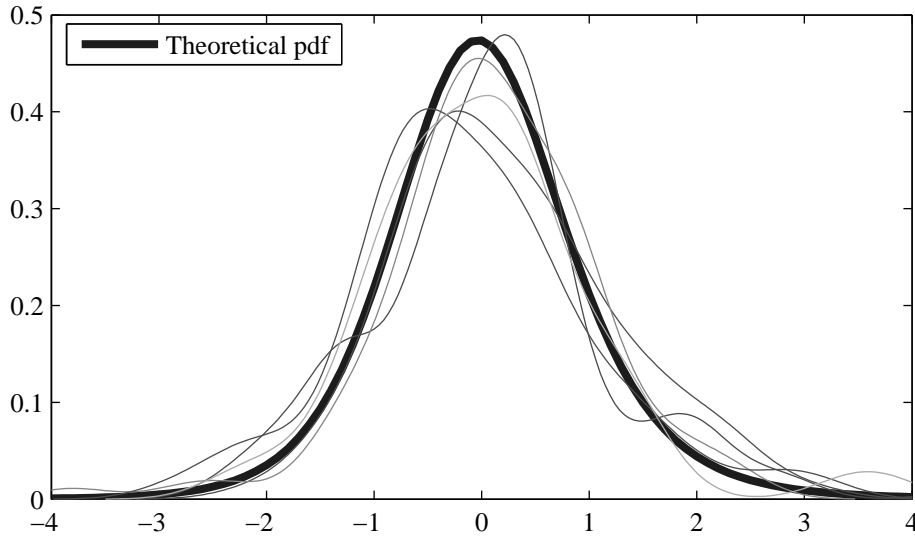


Figure 4.A: Theoretical pdf vs Sample pdfs, Sample Size=100

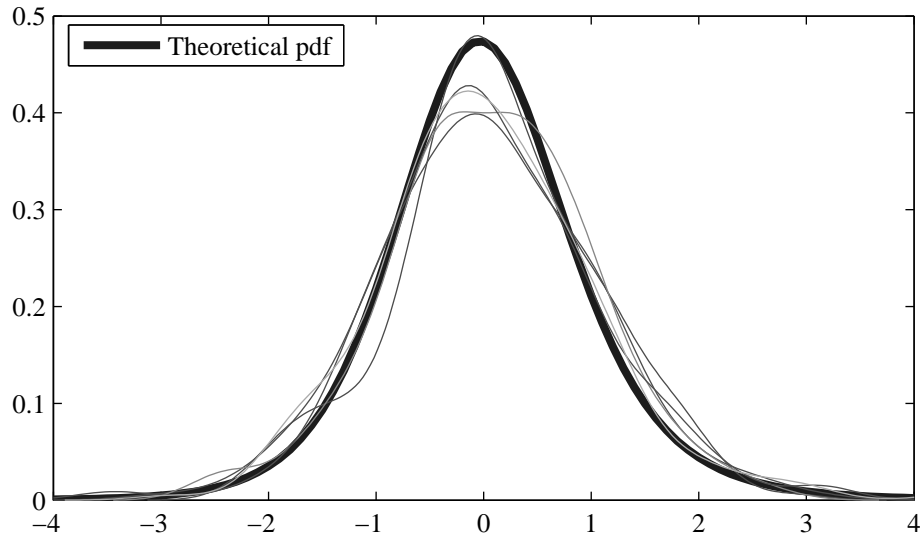


Figure 4.B: Theoretical pdf vs Sample pdfs, Sample Size=250

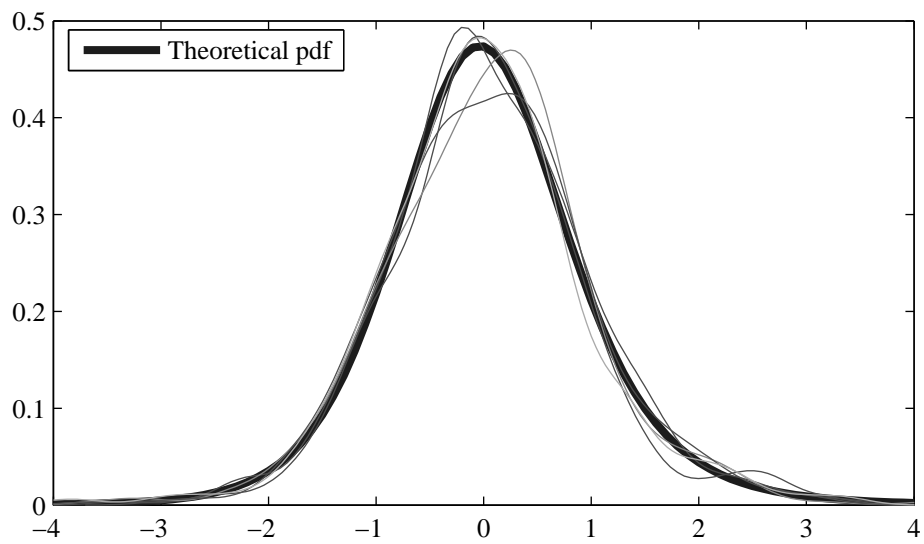


Figure 4.C: Theoretical pdf vs Sample pdfs, Sample Size=500

When the sample size is small, the means of calibrated parameters differ greatly from the exact values and standard deviations are huge. In other words, calibration results are highly unstable between different groups of simulated random variables. An overly small sample size implies a high probability of encountering intractable samples whose empirical distribution is far from representative of the underlying skewed t density (Figure 4.A, 4.B and 4.C are included to illustrate such situations). The algorithm struggles to handle such samples, as reflected by high ratios of blowup² and a large standard deviation of the number of iterations. As the sample size increases, notably after after reaching 750 or 1000, all these anomalies disappear and calibration results are much improved.

Calibrating GARCH parameters³ to non-i.i.d. data uses the same idea of maximum likelihood estimation. To ensure effective filtering, we also need a reasonably large sample.

We simulate a GARCH(1,1) process with $\mu_0 = 8.71 \times 10^{-4}$, $\alpha_0 = 4.26 \times 10^{-6}$, $\alpha_1 = 0.0486$ and $\beta_1 = 0.9161$ ⁴, then filter the simulated processes of different lengths with GARCH repeatedly (100 times each). Similar to our earlier experiment, Table 2 indicates that we need at least 750-1000 samples to approximately retrieve exact parameters.

A minimum of 750 observations is equivalent to about $750/52 \approx 14$ years of weekly data. This may already exceed the entire history for certain stocks (e.g., emerging markets). When it comes to monthly forecasts, a commonly used time horizon for portfolio management, the figure will further grow to $750/12 \approx 62$ years, making it all but impossible to implement the entire framework.⁵

Even if enough data are documented, changes at all market levels (e.g., the integration of global markets, the increasing use of financial derivatives, an industry changed fundamentally by new technology, a firm after merger and acquisition) can still render historical data outdated or irrelevant. As a result, the validity of forecasts will be dubious.

²When the maximum iteration number (usually between 250 and 300) is reached while the relative increment is still above the tolerance (usually set as 10^{-5} or 10^{-6}).

³We use the GARCH Toolbox in Matlab.

⁴Again, to make results more relevant to the application we have in mind, these parameters come from historical data

⁵Daily data of major indexes like DOW and SP500 date back only to the 1950s.

Table 2: Dependence of GARCH Calibration on Sample Size

Sample Size	μ_0		α_0		α_1		β_1	
	Mean	Std	Mean	Std	Mean	Std	Mean	Std
100	8.05E-04	9.31E-04	7.31E-05	3.03E-04	0.095	0.134	0.539	0.383
250	8.61E-04	5.49E-04	1.79E-05	2.25E-05	0.066	0.068	0.747	0.297
500	8.78E-04	3.83E-04	1.03E-05	1.48E-05	0.054	0.033	0.853	0.155
750	8.65E-04	2.95E-04	9.09E-06	1.63E-05	0.054	0.026	0.863	0.186
1000	8.81E-04	2.79E-04	7.07E-06	9.74E-06	0.051	0.021	0.881	0.139
5000	8.66E-04	1.21E-04	4.26E-06	9.63E-07	0.048	0.007	0.917	0.013

4 Multiple Timescale Forecasts

To overcome the aforementioned drawbacks, we'll introduce a high-low frequency approach. Given an n -day time horizon, instead of basing calibration and forecasts on the low-frequency n -day return data, we'll use high-frequency data (e.g., daily returns) to calibrate the return distribution, then switch back to the lower frequency and make forecasts on the n -day timescale. If this approach can be implemented successfully, the restrictions imposed by the scarcity of weekly and monthly historical data will be substantially reduced.

Now we'll describe the procedure for this high-low frequency approach.

Algorithm 4.1. High-Low Frequency VaR Forecast.

1. Choose a conversion factor K based on the amount of available historical data and/or the time range considered most relevant. When the time horizon is n days, K can be any factor of n except 1. For simplicity's sake, we'll assume $K = n$, i.e., n -day is the low frequency, daily is the high frequency.
2. Filter daily data with $GARCH(1,1)$ to get an approximately *i.i.d.* sequence.
3. Calibrate the skewed t distribution to the filtered data. Denote the corresponding parameter set as $(\nu, \mu, \gamma, \sigma)$.

4. At time T , forecast n days into the future using the GARCH(1,1) mechanism⁶:

For $t = T + 1, \dots, T + n$

$$\begin{aligned} X_t &= \mu_0 + \sigma_t Z_t, \\ \sigma_t^2 &= \alpha_0 + \alpha_1 (X_{t-1} - \mu_0)^2 + \beta_1 \sigma_{t-1}^2, \end{aligned}$$

where innovations $Z_t \sim \text{SkewT}(\nu, \mu, \gamma, \sigma)$ and $\mu_0, \alpha_0, \alpha_1, \beta_1$ are GARCH parameters from Step.2⁷.

5. Make n -day VaR forecasts based on the distributions of X_{T+1}, \dots, X_{T+n} .

4.1 Simulation of the GARCH Sum

When this high-low frequency approach is used, VaR estimates must be based on a sequence of n forecasts X_{T+1}, \dots, X_{T+n} . Further, since VaR is not even subadditive, we must consider their sum. Expanding this sum according to (3.8), we get

$$\sum_{i=1}^n X_{T+i} = n\mu_0 + \sigma_{T+1}Z_{T+1} + \dots + \sigma_{T+n}Z_{T+n}. \quad (4.1)$$

Except for σ_{T+1} ⁸, all volatility coefficients are themselves r.v.'s. And by the very nature of GARCH, they are serially correlated.

Since it's extremely difficult to analytically determine the distribution of this GARCH sum, we'll compute VaR estimates by Monte Carlo simulation of the GARCH sum distribution:

- For each of M independent simulations, generate n i.i.d. $\text{SkewT}(\nu, \mu, \gamma, \sigma)$ random variables by Algorithm 2.8. Compute X_{T+1}, \dots, X_{T+n} one by one according to GARCH, then add them up to get the n -day negative log returns L_1, \dots, L_M (i.e., losses).

⁶We'll justify using multiple steps of GARCH later in Section 5

⁷Since $Z_t \sim \text{SkewT}(\nu, \mu, \gamma, \sigma)$ instead of $\text{SWN}(0,1)$, strictly speaking, these forecasts are no longer a GARCH process.

⁸ σ_{T+1} is solely determined by observed data, thus constant.

- Let $\hat{F}_{L,M}(x)$ denote the empirical distribution of losses based on M simulations,

$$\hat{F}_{L,M}(x) = \frac{1}{M} \sum_{i=1}^M 1_{\{L_i \leq x\}}.$$

Estimate VaR_α by the empirical quantile \hat{x}_α

$$\widehat{VaR}_\alpha \left(\sum_{i=1}^n X_{T+i} | \mathcal{F}_T \right) = \hat{x}_\alpha = \hat{F}_{L,M}^{-1}(\alpha),$$

where the inverse of the piecewise constant function $\hat{F}_{L,M}$ is defined as

$$\hat{F}_{L,M}^{-1}(u) = \inf \{ u : \hat{F}_{L,M}(x) \geq u \}.$$

4.2 Confidence Intervals for the Forecasts

The empirical quantile \hat{x}_α converges to the true quantile x_α with probability 1 as $M \rightarrow \infty$. Assuming that L has a strictly positive density f in the neighborhood of x_α , a central limit theorem, as shown in [Serfling \(1980\)](#), states that

$$\sqrt{M}(\hat{x}_\alpha - x_\alpha) \sim \frac{\sqrt{\alpha(1-\alpha)}}{f(x_\alpha)} N(0, 1).$$

This provides the basis for a large-sample $1 - p$ confidence interval⁹ for x_α of the form

$$\hat{x}_\alpha \pm z_{p/2} \frac{\sqrt{\alpha(1-\alpha)}}{f(x_\alpha)\sqrt{M}},$$

where $z_{p/2}$ is the $p/2$ -quantile of standard normal distribution. Since the density f is unknown, we can divide the sample of M simulations into batches, compute an estimate \hat{x}_α for each batch, and form a confidence interval based on the sample standard deviation of the estimates across batches.

[Glasserman \(2003\)](#) proposed an alternative way without computing the sample standard deviation. Using the fact that the number of samples exceeding x_α has binomial distribution with parameters M and α , we can form the confidence interval without relying on a central limit theorem. Let

$$L_{(1)} \leq L_{(2)} \leq \dots \leq L_{(M)}$$

⁹Later we'll use confidence intervals to see if the difference between high-low frequency and fixed frequency forecasts is statistically significant.

denote the order statistics of the losses. An interval of the form $[L_{(r)}, L_{(s)}], r < s$, covers x_α with probability

$$P(L_{(r)} \leq x_\alpha < L_{(s)}) = \sum_{i=r}^{s-1} \binom{M}{i} (1-\alpha)^i \alpha^{M-i}.$$

Choose r, s so that this probability is close to the desired confidence level.

5 Backtesting

GARCH plays dual roles in the high-low frequency approach. First, it removes serial dependence in historical data so that the maximum likelihood method is applicable for distribution fitting. Later, it's used to mimic the return process n days into the future to enable the switch from the high frequency back to the low frequency. Being an effective filter¹⁰ does not necessarily ensure being a good simulation mechanism. We must justify its validity through backtests.

5.1 Independence of Violation Indicators

Define indicator variables

$$I_{t+1}^\alpha = 1_{\{\sum_{i=1}^n X_{t+i} > VaR_\alpha(\sum_{i=1}^n X_{t+i} | \mathcal{F}_t)\}},$$

which indicates whether the actual losses over the next n days exceeds the conditional VaR forecasted at time t .

Recall that,

$$\sum_{i=1}^n X_{t+i} = n\mu_0 + \sigma_{t+1}Z_{t+1} + \dots + \sigma_{t+n}Z_{t+n}.$$

Since $\sigma_{t+2}, \dots, \sigma_{t+n}$ are actually determined by $\sigma_{t+1}, Z_{t+1}, \dots, Z_{t+n-1}$, we can summarize this sum as

$$\sum_{i=1}^n X_{t+i} = f(\mu_0, \sigma_{t+1}, Z_{t+1}, \dots, Z_{t+n}).$$

¹⁰As illustrated by Figure 1 and 2. [Hu and Kercheval \(2007\)](#) also provided favorable evidence.

Similarly, the conditional VaR can be expressed as

$$VaR_\alpha\left(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t\right) = g(\mu_0, \sigma_{t+1}, Z^{(1)}, \dots, Z^{(n)}).^{11}$$

Noticing that μ_0 and the distribution of $Z^{(1)}, \dots, Z^{(n)}$ hold constant throughout the GARCH process, we can further reduce the above notation to

$$VaR_\alpha\left(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t\right) = g(\sigma_{t+1}).$$

Now consider two indicator variables

$$I_{t+1}^\alpha = 1_{\{f(\mu_0, \sigma_{t+1}, Z_{t+1}, \dots, Z_{t+n}) > g(\sigma_{t+1})\}}$$

and

$$I_{t+j+1}^\alpha = 1_{\{f(\mu_0, \sigma_{t+j+1}, Z_{t+j+1}, \dots, Z_{t+j+n}) > g(\sigma_{t+j+1})\}}.$$

Among the variables, μ_0 is a constant. Both σ_{t+1} and σ_{t+j+1} are \mathcal{F}_{t+j} -measurable, so they are deterministic at time $t + j$. Therefore, as long as Z_{t+1}, \dots, Z_{t+n} and $Z_{t+j+1}, \dots, Z_{t+j+n}$ don't overlap, i.e., $j \geq n$, these i.i.d. innovations will ensure that I_{t+1}^α and I_{t+j+1}^α are independent.

5.2 Backtest Algorithm

Given a data set of N daily negative log returns $\{X_1, \dots, X_N\}$, suppose our target time horizon is n days and the sample size for skewed t calibration is C . Use the following algorithm for backtest.

Algorithm 5.1. Backtest. For $t = C, C + n, C + 2n, \dots, C + kn, \dots$

1. Apply GARCH filter to $\{X_{t-C+1}, \dots, X_t\}$ and calibrate the skewed t distribution.
2. Compute $\widehat{VaR}_\alpha(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t)$, the estimate for conditional VaR_α , for $\alpha = 0.95, 0.975, 0.99$ and 0.995 with Monte Carlo simulation.

¹¹VaR is based on the distribution of $\sum_{i=1}^n X_{t+i}$, not the actual values of Z_{t+1}, \dots, Z_{t+n} . Distinguishing Z 's with a time index is no longer necessary. It suffices to use $Z^{(1)}, \dots, Z^{(n)}$, i.e., n i.i.d. skewed t r.v.'s.

3. Compare the sum of next n observations with $\widehat{\text{VaR}}_\alpha$. Violation is counted by $I_{t+1}^\alpha = 1_{\{\sum_{i=1}^n X_{t+i} > \widehat{\text{VaR}}_\alpha(\sum_{i=1}^n X_{t+i} | \mathcal{F}_t)\}}$.

5.3 Statistical Tests

Definition 5.2. Bernoulli(p) Distribution. A random variable X has a Bernoulli(p) distribution if

$$X = \begin{cases} 1 & \text{with probability } p \\ 0 & \text{with probability } 1 - p, \end{cases} \quad 0 \leq p \leq 1.$$

If the estimates of conditional VaR are successful, the violation indicators should behave like i.i.d. Bernoulli($1 - \alpha$) trials.

Proposition 5.3. Bernoulli MLE. Let X_1, \dots, X_n be i.i.d. Bernoulli(p). The likelihood function is

$$L(p|\mathbf{x}) = \prod_{i=1}^n p^{x_i} (1-p)^{1-x_i} = p^y (1-p)^{n-y},$$

where $y = \sum_{i=1}^n x_i$. The MLE of p is $\sum_{i=1}^n X_i/n$.

Proof. $L(p|\mathbf{x}) = p^y (1-p)^{n-y}$ follows immediately from the definition of likelihood functions and the pmf of Bernoulli(p). $L(p|\mathbf{x}) = p^y (1-p)^{n-y}$ is not too hard to differentiate, but it is easier to differentiate the log likelihood

$$\log L(p|\mathbf{x}) = y \log p + (n-y) \log(1-p).$$

If $0 < y < n$, differentiating $\log L(p|\mathbf{x})$ and setting the result equal to 0 yields the solution $\hat{p} = y/n$. It can be easily verified that y/n is the global maximum. If $y = 0$ or $y = n$, it is again straightforward to verify that $\hat{p} = y/n$. Thus, $\sum_{i=1}^n X_i/n$ is the MLE of p . \square

Definition 5.4. Likelihood Ratio Test Statistic. Let Θ denote the entire parameter space, and Θ_0 some parameter subset. The likelihood ratio test (LRT) statistic for testing $H_0 : \theta \in \Theta_0$ versus $H_1 : \theta \in \Theta_0^c$ is

$$\lambda(\mathbf{x}) = \frac{\sup_{\Theta_0} L(\theta|\mathbf{x})}{\sup_{\Theta} L(\theta|\mathbf{x})}.$$

To understand the rationale behind the LRT, consider a sample of discrete random variables, in which case the likelihood function is the product of pmf's (i.e., discrete probabilities). The numerator of $\lambda(\mathbf{x})$ is the maximum probability, computed over parameters in the null hypothesis, of the observed sample. The denominator is the maximum probability over all possible parameters. The ratio of these two maxima is small if there are parameter points in the alternative hypothesis for which the observed sample is much more likely than for any parameter point in the null hypothesis. In that case, the LRT criterion concludes H_0 should be rejected.

Proposition 5.5. Bernoulli LRT. *Let X_1, \dots, X_n be a random sample from a Bernoulli(p) population. Consider testing $H_0 : p = p_0$ versus $H_1 : p \neq p_0$. The LRT statistic is*

$$\lambda(\mathbf{x}) = \frac{p_0^y (1 - p_0)^{n-y}}{\left(\frac{y}{n}\right)^y \left(1 - \frac{y}{n}\right)^{n-y}}.$$

Proof. Since there is only one value specified by H_0 , the numerator of $\lambda(\mathbf{x})$ is $L(p_0|\mathbf{x})$. The MLE of p is y/n by Proposition 5.3, so the denominator is $L(\frac{y}{n}|\mathbf{x})$. Together, the LRT statistic is

$$\lambda(\mathbf{x}) = \frac{L(p_0|\mathbf{x})}{L(\frac{y}{n}|\mathbf{x})} = \frac{p_0^y (1 - p_0)^{n-y}}{\left(\frac{y}{n}\right)^y \left(1 - \frac{y}{n}\right)^{n-y}}.$$

□

Theorem 5.6. Asymptotic Distribution of the LRT. *For testing $H_0 : \theta = \theta_0$ versus $H_1 : \theta \neq \theta_0$, suppose X_1, \dots, X_n are i.i.d. with pdf or pmf $f(x|\theta)$, $\hat{\theta}$ is the MLE of θ , and $f(x|\theta)$ satisfies certain regularity conditions. Then under H_0 , as $n \rightarrow \infty$,*

$$-2 \log \lambda(\mathbf{x}) \rightarrow \chi^2(1) \text{ in distribution .}$$

Proof. See [Casella and Berger \(2002\)](#). □

Recall that we did a sequence of backtests by counting VaR violations. Suppose there are altogether Y tests and let y denote the number of total violations. The actual violation frequency is y/Y , which is the MLE of a Bernoulli distribution, while the expected violation probability, q , should be 0.05, 0.025, 0.01 and 0.005.

We then evaluate the backtest results with the likelihood ratio test¹² based on Theorem 5.6:

$$\begin{cases} H_0 : & \text{The expected violation ratio} = q \\ H_1 : & \text{The expected violation ratio} \neq q \end{cases}$$

Under the null hypothesis, the test statistic is

$$-2[(Y - y) \log(1 - q) + y \log(q)] + 2[(Y - y) \log(1 - y/Y) + y \log(y/Y)],$$

which is asymptotically $\chi^2(1)$ distributed.

The stepsize of the backtest algorithm is set as n to ensure the independence of violation indicators. As a result, the number of tests that can be done on a fixed amount of daily data will shrink substantially when the time horizon increases. To extract more information about the violations, we can implement the backtest algorithm n times, each with a different starting point in the time index (i.e., $t = C, C+1, \dots, C+n-1$). Each of the n backtests will contain the same total number of tests Y , but a different number of violations: y_1, \dots, y_n .

5.4 n -Day Horizon

We list the actual violation ratios and the corresponding p -values of the LRT in Table 3.A through 3.J. All VaR backtesting is based on S&P500 daily close prices from 1991/1/1 to 2009/12/31. 1000 samples are used to calibrate each skewed t distribution. Depending on the length of the time horizon, the total number of backtests ranges between 500 to 1900. For an n -day horizon, we have n groups of results representing different starting points in the time index.

From a 2-day to a 10-day horizon, altogether $(2+10) \times 9 \times 4/2 = 216$ likelihood ratio tests are done, among which the five p -values lower than 0.05¹³ are sparsely distributed. Setting the confidence level of the LRT to be 95% implies that even if the model is perfect, we'll still have a 5% chance of observing LRT failures (i.e., Type I errors). Since the failure ratio of the backtests $5/216 \approx 2.3\%$ is much lower than 5%, we consider our model performance's satisfactory.

Still, the high-low frequency method does have its limits. As shown in Table 3.J, when the time horizon is extended to 15-day, 9 failures are observed in 60 LRT. This is not surprising. As the time horizon stretches we are ignoring more and more new information, leading to deteriorating GARCH performance.

¹²Since the sum of i.i.d. Bernoulli random variable is a binomial random variable, another alternative is a standard two-sided binomial test, as described by Casella and Berger (2002).

¹³We reject the null hypothesis when the p -value is less than 0.05.

Table 3.A: Backtest Results: 2-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.059	0.026	0.010	0.005	0.076	0.809	0.827	0.864
Group 2	0.050	0.027	0.010	0.007	0.975	0.596	0.827	0.277

Table 3.B: Backtest Results: 3-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.057	0.027	0.012	0.008	0.260	0.663	0.513	0.175
Group 2	0.054	0.032	0.013	0.009	0.532	0.143	0.239	0.091
Group 3	0.052	0.025	0.012	0.007	0.807	0.921	0.513	0.313

Table 3.C: Backtest Results: 4-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.055	0.025	0.012	0.004	0.490	0.942	0.624	0.730
Group 2	0.041	0.030	0.013	0.006	0.202	0.378	0.426	0.574
Group 3	0.052	0.023	0.013	0.005	0.801	0.728	0.426	0.902
Group 4	0.053	0.030	0.020	0.008	0.690	0.378	0.006	0.170

Table 3.D: Backtest Results: 5-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.045	0.024	0.007	0.005	0.514	0.828	0.317	0.913
Group 2	0.053	0.026	0.012	0.008	0.722	0.804	0.612	0.293
Group 3	0.046	0.029	0.011	0.008	0.630	0.485	0.876	0.293
Group 4	0.052	0.026	0.016	0.011	0.849	0.804	0.136	0.059
Group 5	0.057	0.037	0.012	0.005	0.400	0.048	0.612	0.913

Table 3.E: Backtest Results: 6-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.052	0.019	0.011	0.006	0.785	0.318	0.783	0.645
Group 2	0.049	0.022	0.013	0.008	0.927	0.649	0.514	0.336
Group 3	0.056	0.025	0.011	0.008	0.529	0.949	0.783	0.336
Group 4	0.048	0.021	0.014	0.011	0.782	0.470	0.310	0.061
Group 5	0.057	0.030	0.011	0.006	0.421	0.422	0.783	0.645
Group 6	0.051	0.032	0.011	0.006	0.927	0.298	0.783	0.645

Table 3.F: Backtest Results: 7-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.043	0.024	0.011	0.007	0.418	0.890	0.799	0.459
Group 2	0.044	0.017	0.013	0.006	0.546	0.187	0.508	0.857
Group 3	0.041	0.033	0.013	0.004	0.308	0.238	0.508	0.654
Group 4	0.044	0.024	0.013	0.006	0.546	0.890	0.508	0.857
Group 5	0.039	0.024	0.019	0.013	0.218	0.890	0.075	0.029
Group 6	0.054	0.020	0.011	0.007	0.696	0.477	0.799	0.459
Group 7	0.059	0.030	0.006	0.006	0.337	0.503	0.257	0.857

Table 3.G: Backtest Results: 8-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.034	0.021	0.013	0.006	0.089	0.586	0.570	0.689
Group 2	0.055	0.023	0.015	0.006	0.618	0.811	0.325	0.689
Group 3	0.047	0.028	0.008	0.004	0.733	0.728	0.732	0.809
Group 4	0.038	0.017	0.015	0.006	0.218	0.235	0.325	0.689
Group 5	0.051	0.030	0.008	0.004	0.933	0.528	0.732	0.809
Group 6	0.053	0.017	0.008	0.006	0.770	0.235	0.732	0.689
Group 7	0.044	0.023	0.015	0.008	0.576	0.811	0.325	0.331
Group 8	0.049	0.023	0.006	0.004	0.899	0.811	0.394	0.809

Table 3.H: Backtest Results: 9-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.036	0.014	0.007	0.005	0.158	0.127	0.535	0.944
Group 2	0.045	0.024	0.012	0.005	0.649	0.875	0.703	0.944
Group 3	0.043	0.029	0.010	0.005	0.492	0.647	0.921	0.944
Group 4	0.033	0.021	0.010	0.002	0.096	0.631	0.921	0.396
Group 5	0.048	0.024	0.007	0.005	0.822	0.875	0.535	0.944
Group 6	0.057	0.026	0.014	0.010	0.511	0.877	0.407	0.243
Group 7	0.052	0.031	0.017	0.014	0.824	0.451	0.210	0.028
Group 8	0.052	0.024	0.014	0.010	0.824	0.875	0.407	0.243
Group 9	0.040	0.017	0.010	0.002	0.355	0.245	0.921	0.396

Table 3.I: Backtest Results: 10-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.040	0.016	0.011	0.008	0.340	0.224	0.910	0.456
Group 2	0.048	0.026	0.016	0.008	0.831	0.858	0.290	0.456
Group 3	0.048	0.024	0.021	0.016	0.831	0.881	0.058	0.017
Group 4	0.040	0.019	0.013	0.008	0.340	0.398	0.548	0.456
Group 5	0.034	0.016	0.005	0.005	0.141	0.224	0.312	0.937
Group 6	0.032	0.016	0.013	0.008	0.082	0.224	0.548	0.456
Group 7	0.040	0.016	0.008	0.008	0.340	0.224	0.676	0.456
Group 8	0.048	0.026	0.011	0.005	0.831	0.858	0.910	0.937
Group 9	0.042	0.026	0.016	0.008	0.483	0.858	0.290	0.456
Group 10	0.053	0.032	0.011	0.005	0.797	0.420	0.910	0.937

Table 3.J: Backtest Results: 15-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.046	0.036	0.015	0.010	0.645	0.118	0.230	0.117
Group 2	0.049	0.034	0.014	0.010	0.940	0.184	0.405	0.117
Group 3	0.044	0.029	0.020	0.010	0.512	0.553	0.026	0.117
Group 4	0.058	0.031	0.012	0.007	0.395	0.399	0.652	0.557
Group 5	0.054	0.027	0.015	0.012	0.627	0.735	0.230	0.044
Group 6	0.054	0.031	0.017	0.012	0.627	0.399	0.121	0.044
Group 7	0.054	0.041	0.020	0.007	0.627	0.024	0.026	0.557
Group 8	0.060	0.034	0.019	0.007	0.303	0.184	0.058	0.557
Group 9	0.056	0.037	0.017	0.012	0.504	0.072	0.121	0.044
Group 10	0.056	0.029	0.015	0.012	0.504	0.553	0.230	0.044
Group 11	0.061	0.041	0.014	0.009	0.227	0.024	0.405	0.274
Group 12	0.065	0.032	0.014	0.007	0.119	0.277	0.405	0.557
Group 13	0.056	0.032	0.014	0.009	0.504	0.277	0.405	0.274
Group 14	0.058	0.031	0.014	0.012	0.395	0.399	0.405	0.044
Group 15	0.058	0.029	0.015	0.010	0.395	0.553	0.230	0.117

5.5 A Variant Backtest: Forecasting a One Day Return, n Days Ahead

The backtest above evaluates the accuracy of the simulated sum $\sum_{i=1}^n X_{t+i}$ as a whole. To further examine each term inside that sum, we can slightly modify the previous backtest algorithm:

Algorithm 5.7. Alternative Backtest. For $t = C, C + n, C + 2n, \dots, C + kn, \dots$

1. Apply GARCH filter to $\{X_{t-C+1}, \dots, X_t\}$ and calibrate the skewed t distribution.
2. Compute $\widehat{VaR}_\alpha(X_{t+n}|\mathcal{F}_t)$, i.e., VaR for Day $t+n$, for $\alpha = 0.95, 0.975, 0.99$ and 0.995 with Monte Carlo simulation.
3. Compare the observation on Day $t + n$ with $\widehat{VaR}_\alpha(X_{t+n}|\mathcal{F}_t)$. Violation is counted if $X_{t+n} > \widehat{VaR}_\alpha(X_{t+n}|\mathcal{F}_t)$.

Repeat this alternative version with stepsize from 2 to n days and we'll get a day-by-day picture of the performance of GARCH simulation. We list the results in Table 4.A through 4.I.

Table 4.A: Alternative Backtest Results: 2-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.058	0.030	0.011	0.006	0.165	0.229	0.611	0.595
Group 2	0.048	0.027	0.007	0.004	0.721	0.682	0.276	0.569

Not a single LRT failure is observed up to a 7-day horizon. Successfully forecasting day-by-day VaR lends further confidence in the high-low frequency method.

Table 4.B: Alternative Backtest Results: 3-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.055	0.030	0.010	0.003	0.475	0.326	1.000	0.333
Group 2	0.052	0.027	0.007	0.005	0.773	0.689	0.314	1.000
Group 3	0.054	0.027	0.011	0.006	0.566	0.689	0.754	0.664

Table 4.C: Alternative Backtest Results: 4-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.063	0.033	0.011	0.004	0.125	0.164	0.856	0.687
Group 2	0.043	0.020	0.011	0.003	0.345	0.364	0.856	0.320
Group 3	0.055	0.021	0.008	0.007	0.563	0.510	0.568	0.538
Group 4	0.061	0.033	0.009	0.005	0.168	0.164	0.853	0.898

Table 4.D: Alternative Backtest Results: 5-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.050	0.022	0.005	0.003	1.000	0.593	0.173	0.538
Group 2	0.063	0.028	0.010	0.007	0.149	0.608	1.000	0.582
Group 3	0.057	0.030	0.010	0.007	0.463	0.447	1.000	0.582
Group 4	0.043	0.030	0.012	0.005	0.444	0.447	0.689	1.000
Group 5	0.053	0.028	0.010	0.005	0.711	0.608	1.000	1.000

Table 4.E: Alternative Backtest Results: 6-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.056	0.034	0.010	0.006	0.546	0.221	1.000	0.759
Group 2	0.050	0.018	0.004	0.004	1.000	0.292	0.125	0.743
Group 3	0.056	0.022	0.008	0.006	0.546	0.661	0.641	0.759
Group 4	0.048	0.030	0.010	0.006	0.836	0.487	1.000	0.759
Group 5	0.060	0.032	0.008	0.006	0.319	0.336	0.641	0.759
Group 6	0.054	0.032	0.012	0.002	0.685	0.336	0.663	0.279

Table 4.F: Alternative Backtest Results: 7-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.056	0.028	0.014	0.009	0.571	0.693	0.431	0.256
Group 2	0.070	0.037	0.007	0.005	0.071	0.126	0.511	0.923
Group 3	0.061	0.023	0.009	0.005	0.323	0.827	0.891	0.923
Group 4	0.051	0.023	0.007	0.005	0.895	0.827	0.511	0.923
Group 5	0.042	0.023	0.009	0.005	0.439	0.827	0.891	0.923
Group 6	0.056	0.026	0.009	0.005	0.571	0.926	0.891	0.923
Group 7	0.051	0.030	0.012	0.009	0.895	0.491	0.733	0.256

Table 4.G: Alternative Backtest Results: 8-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.051	0.035	0.008	0.005	0.953	0.257	0.687	0.928
Group 2	0.040	0.024	0.008	0.008	0.358	0.901	0.687	0.449
Group 3	0.059	0.029	0.008	0.005	0.453	0.601	0.687	0.928
Group 4	0.069	0.040	0.011	0.008	0.104	0.087	0.898	0.449
Group 5	0.061	0.035	0.011	0.008	0.330	0.257	0.898	0.449
Group 6	0.051	0.021	0.005	0.000	0.953	0.641	0.319	NaN*
Group 7	0.053	0.021	0.008	0.008	0.769	0.641	0.687	0.449
Group 8	0.056	0.027	0.008	0.005	0.601	0.838	0.687	0.928

*NaN (not a number) arises from $\log 0$ in the likelihood ratio statistic when violation count is 0.

Table 4.H: Alternative Backtest Results: 9-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.051	0.024	0.009	0.003	0.930	0.909	0.853	0.577
Group 2	0.054	0.021	0.006	0.000	0.737	0.633	0.429	NaN
Group 3	0.075	0.030	0.015	0.006	0.050	0.569	0.392	0.801
Group 4	0.060	0.030	0.009	0.006	0.414	0.569	0.853	0.801
Group 5	0.069	0.036	0.009	0.006	0.130	0.226	0.853	0.801
Group 6	0.045	0.018	0.009	0.009	0.673	0.391	0.853	0.351
Group 7	0.066	0.042	0.018	0.012	0.199	0.069	0.186	0.125
Group 8	0.042	0.024	0.012	0.006	0.494	0.909	0.721	0.801
Group 9	0.039	0.024	0.009	0.003	0.341	0.909	0.853	0.577

Table 4.I: Alternative Backtest Results: 10-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.047	0.017	0.003	0.000	0.789	0.326	0.178	NaN
Group 2	0.050	0.027	0.003	0.000	1.000	0.855	0.178	NaN
Group 3	0.070	0.030	0.013	0.010	0.133	0.591	0.581	0.280
Group 4	0.043	0.030	0.007	0.007	0.588	0.591	0.537	0.697
Group 5	0.067	0.030	0.007	0.000	0.207	0.591	0.537	NaN
Group 6	0.053	0.030	0.013	0.010	0.793	0.591	0.581	0.280
Group 7	0.070	0.027	0.017	0.010	0.133	0.855	0.290	0.280
Group 8	0.053	0.027	0.010	0.007	0.793	0.855	1.000	0.697
Group 9	0.043	0.030	0.010	0.007	0.588	0.591	1.000	0.697
Group 10	0.070	0.023	0.010	0.007	0.133	0.852	1.000	0.697

5.6 A Note on the Monthly Forecast Horizon

As shown above, when the target horizon is monthly, a daily frequency is no longer appropriate for the high frequency level. In the high-low frequency approach, timescales on both levels are evenly spaced. However, physical months are of varying lengths (in terms of trading days). We must first find some substitute for the monthly horizon.

Not considering occasional national holidays, each year typically has 260 trading days and thereby $260/12=21.7$ days each month on average. We'll use 20 or 21 days as the monthly horizon, as they are divisible by reasonable conversion factors.¹⁴ LRT results are presented in Table 5.A through 5.D¹⁵. Again, the high-low frequency model passes the backtests.

¹⁴Reasonable conversion factors are those that can effectively reduce reliance on data availability without going beyond the limits of GARCH forecasting. For example, the conversion factors for 22-day horizon can be 2, 11 or 22. Using 11 and 22 risks extrapolating much too far into the future while choosing 2 means you still need 30 years of data.

¹⁵To get enough backtests, the data set is expanded to 1971-2009.

Table 5.A: Backtest Results: 7-Day to 21-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.054	0.021	0.012	0.012	0.747	0.627	0.725	0.126
Group 2	0.051	0.030	0.018	0.006	0.940	0.575	0.188	0.804
Group 3	0.045	0.018	0.012	0.012	0.664	0.386	0.725	0.126

Table 5.B: Backtest Results: 5-Day to 20-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.060	0.032	0.018	0.008	0.340	0.351	0.110	0.389
Group 2	0.052	0.028	0.012	0.008	0.871	0.695	0.677	0.389
Group 3	0.044	0.026	0.012	0.006	0.504	0.910	0.677	0.769
Group 4	0.054	0.028	0.012	0.008	0.716	0.695	0.677	0.389

Table 5.C: Backtest Results: 4-Day to 20-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.045	0.030	0.018	0.008	0.570	0.460	0.100	0.371
Group 2	0.045	0.026	0.012	0.010	0.570	0.853	0.643	0.157
Group 3	0.051	0.028	0.008	0.006	0.951	0.641	0.660	0.744
Group 4	0.053	0.026	0.016	0.010	0.790	0.853	0.204	0.157
Group 5	0.061	0.030	0.014	0.006	0.289	0.460	0.381	0.744

Table 5.D: Backtest Results: 3-Day to 21-Day

q	Violation Ratio				p -value			
	0.05	0.025	0.01	0.005	0.05	0.025	0.01	0.005
Group 1	0.041	0.021	0.017	0.013	0.335	0.606	0.161	0.045
Group 2	0.043	0.017	0.011	0.011	0.460	0.246	0.883	0.130
Group 3	0.058	0.021	0.013	0.009	0.456	0.606	0.557	0.323
Group 4	0.058	0.032	0.013	0.006	0.456	0.349	0.557	0.679
Group 5	0.064	0.036	0.017	0.017	0.179	0.141	0.161	0.004
Group 6	0.049	0.032	0.011	0.006	0.932	0.349	0.883	0.679
Group 7	0.038	0.024	0.011	0.009	0.233	0.834	0.883	0.323

5.7 Stability of the Likelihood Ratio Test and Algorithm Speed

VaR estimates based on Monte Carlo simulation are inherently random. If we run a specific backtest more than once, the number of VaR violations will not be a fixed number. Will such uncertainty affect the conclusion of likelihood ratio tests?

To investigate this problem, we first backtest a daily VaR forecast using the fixed frequency method. Since VaR is computed using numerical integration and root finding, the resulting violation numbers are constant and can be used as a benchmark. We then replace the VaR forecast step with Monte Carlo simulation as used in the high-low frequency method, repeat the same backtest again and again and examine the discrepancy between violation numbers and the benchmark.

As shown in Table 6.A through 6.D, all 20 trials, with either 10K or 25K simulations, lead to the same conclusion for the likelihood ratio tests. As a matter of fact, over the entire course, the error in violation count never exceeds 6, which is a tiny fraction of the 8843 tests¹⁶ and also minimal when compared with the acceptable band¹⁷ of likelihood ratio tests. Therefore, we can conclude that when 25K¹⁸ simulations are used, the randomness of the violation counts is already

¹⁶Using 1971-2009 daily returns.

¹⁷Kupiec (1995) provided an extensive survey of the performance of LRT in the context of risk management.

¹⁸Although using 25K simulations reduces both relative error and absolute error in VaR by about one third as compared to 10K, in terms of average discrepancy and maximum discrepancy,

Table 6.A: Stability of Violation Counts MC=10K

α	95%	97.50%	99%	99.50%
Benchmark	451	238	91	46
Trial 1	455	244	90	50
Trial 2	445	242	89	47
Trial 3	452	243	92	48
Trial 4	447	244	93	50
Trial 5	450	240	93	49
Trial 6	455	244	90	50
Trial 7	445	242	89	47
Trial 8	452	243	92	48
Trial 9	447	244	93	50
Trial 10	450	240	93	49
Avg Error	3.2	4.6	1.6	2.8
Max Error	6	6	2	4

Table 6.B: Stability of Violation Counts MC=25K

α	95%	97.50%	99%	99.50%
Benchmark	451	238	91	46
Trial 1	451	239	89	48
Trial 2	445	239	88	47
Trial 3	446	244	88	48
Trial 4	449	238	90	45
Trial 5	446	241	88	51
Trial 6	453	243	94	48
Trial 7	451	242	90	49
Trial 8	451	235	91	50
Trial 9	451	237	92	49
Trial 10	446	241	91	47
Avg Error	2.5	2.7	1.7	2.4
Max Error	6	6	3	5

Table 6.C: VaR Error

α	Mean Absolute Error*				Mean Relative Error			
	95%	97.50%	99%	99.50%	95%	97.50%	99%	99.50%
10K	0.00020	0.00029	0.00049	0.00074	1.30%	1.49%	1.95%	2.51%
25K	0.00013	0.00019	0.00031	0.00047	0.83%	0.95%	1.25%	1.62%

*As compared with the forecast made by the fixed frequency method, with absolute value taken.

Table 6.D: Acceptable* Band of LRT

α	95%	97.50%	99%	99.50%
Expected Violation	442	221	89	42
Min Acceptable	403	193	71	32
Max Acceptable	482	250	107	57

*So that $p\text{-value} > 0.05$.

immaterial and likelihood ratio test results are stable.

In our experience, filtering a sample of 1000 data with GARCH can be finished within a second or two.¹⁹ It then takes the EM algorithm about 2 minutes²⁰ to calibrate the skewed t distribution. In the forecast step, depending on the conversion factor, every 10K simulation costs about 1.5-3 seconds. Since the target horizon for the high-low frequency approach is at least 2-day, the model is quick enough to produce a single VaR forecast to any desired accuracy level (by increasing the number of simulations). When 25K simulations are used, the mean absolute error and relative error for $VaR_{95\%}$ are 0.00013 and 0.83%.

5.8 A Quick Look at Intraday Data

As mentioned earlier, financial data behave similarly on all time intervals. The entire approach can be readily applied to intraday data. However, as the time interval is shortened to minutes or even seconds, data typically get more noisy and the order of GARCH filter must be increased to get satisfactory results. To illustrate this, we examined a time series of 1-minute returns of a US long term bond index futures contract over the week of 3/16/2009. Auto-correlation of GARCH filtering at different orders are shown in Figure 5.A through 5.C.

VaR is not traditionally used for time horizons less than one day, but to illu-

there is not much difference. Still, to be cautious, we performed all backtests listed in tables prior to Table 6.A using 30K simulations.

¹⁹We use a laptop with 2.8GHz CPU and 2GB memory. The software is MATLAB R2007b.

²⁰In backtest, the time spent on calibration can be significantly reduced. Since most data are overlapping, we can use the parameters from calibration results of the previous period as the initial values and start from scratch only infrequently, say every 200 days, to avoid overestimation.

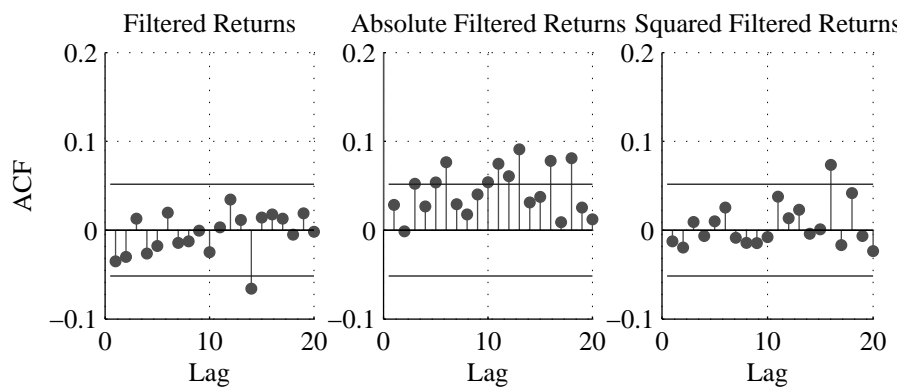


Figure 5.A: ACF for Minute Data using GARCH(1,1)

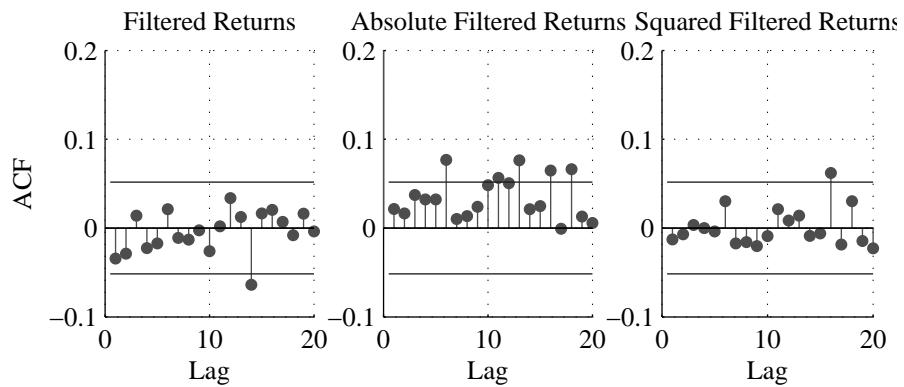


Figure 5.B: ACF for Minute Data using GARCH(2,2)

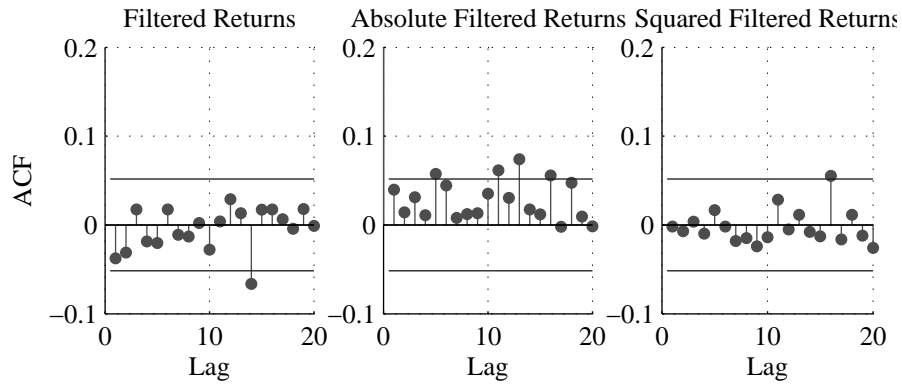


Figure 5.C: ACF for Minute Data using GARCH(3,3)

minimize the possibilities for high-frequency applications, we show multiscale VaR forecast backtest results for two through five minute horizons, based on the same intraday data, in Tables 6.A through 6.C below.

We see some rejections of the backtest null hypothesis for GARCH(1,1), but GARCH(2,2) and GARCH(3,3) behave quite reasonably. Therefore, at least for this high frequency returns series, intraday VaR forecasts are a viable option with the multiscale GARCH method. Practitioners should test their own data for suitability of this approach for their own use.

Table 6.A: Backtest Results (p -values) Using GARCH(1,1)

2-Minute Horizon				
α	95%	97.5%	99%	99.5%
Group 1	0.349	0.344	0.438	0.050
Group 2	0.059	0.510	0.047	0.114

3-Minute Horizon				
α	95%	97.5%	99%	99.5%
Group 1	0.317	0.513	0.526	0.392
Group 2	0.264	0.406	0.225	0.196
Group 3	0.177	0.631	0.129	0.022

4-Minute Horizon				
α	95%	97.5%	99%	99.5%
Group 1	0.552	0.571	0.113	0.094
Group 2	0.096	0.867	0.049	0.020
Group 3	0.465	0.441	0.704	0.263
Group 4	0.317	0.441	0.379	0.871

5-Minute Horizon				
α	95%	97.5%	99%	99.5%
Group 1	0.354	0.024	0.027	NaN
Group 2	0.036	0.453	0.182	0.055
Group 3	0.036	0.453	0.838	0.508
Group 4	0.881	0.081	0.573	0.885
Group 5	0.530	0.744	0.838	0.508

Table 6.B: Backtest Results (p -values) Using GARCH(2,2)

2-Minute Horizon				
α	95%	97.5%	99%	99.5%
Group 1	0.157	0.652	0.583	0.380
Group 2	0.089	0.967	0.917	0.050

3-Minute Horizon				
α	95%	97.5%	99%	99.5%
Group 1	0.317	0.631	0.526	0.655
Group 2	0.444	0.840	0.526	0.196
Group 3	0.217	0.593	0.225	0.078

4-Minute Horizon				
α	95%	97.5%	99%	99.5%
Group 1	0.851	0.871	0.113	0.095
Group 2	0.053	0.542	0.113	0.095
Group 3	0.961	0.675	0.585	0.536
Group 4	0.460	0.542	0.222	0.095

5-Minute Horizon				
α	95%	97.5%	99%	99.5%
Group 1	0.744	0.584	0.079	NaN
Group 2	0.888	0.921	0.349	0.509
Group 3	0.744	0.438	0.349	0.509
Group 4	0.864	0.212	0.183	0.055
Group 5	0.631	0.738	0.840	0.216

Table 6.C: Backtest Results (p -values) Using GARCH(3,3)

2-Minute Horizon				
α	95%	97.5%	99%	99.5%
Group 1	0.259	0.752	0.583	0.584
Group 2	0.108	0.323	0.438	0.050

3-Minute Horizon				
α	95%	97.5%	99%	99.5%
Group 1	0.518	0.513	0.932	0.655
Group 2	0.114	0.593	0.661	0.196
Group 3	0.377	0.712	0.526	0.196

4-Minute Horizon				
α	95%	97.5%	99%	99.5%
Group 1	0.202	0.871	0.380	0.264
Group 2	0.053	0.973	0.222	0.020
Group 3	0.744	0.675	0.821	0.536
Group 4	0.053	0.542	0.113	0.095

5-Minute Horizon				
α	95%	97.5%	99%	99.5%
Group 1	0.350	0.584	0.183	NaN
Group 2	0.764	0.584	0.840	0.055
Group 3	0.864	0.585	0.575	0.738
Group 4	0.631	0.212	0.575	0.216
Group 5	0.094	0.921	0.575	0.216

6 Further Analysis: Long-Term GARCH and Comparisons using Simulated Data

6.1 Long-Term Behavior of GARCH

Results in Section 5.4 suggest that the conversion factor of the high-low frequency approach cannot be set too high. In this section we discuss the long-horizon behavior of GARCH forecasts, in particular, the conditional variances $\text{var}(X_{t+n}|\mathcal{F}_t)$ and $\text{var}(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t)$ as n grows, to find some explanation for such restrictions. We see below that GARCH forecasts begin to degenerate at long horizons, so that they are bound to lose their usefulness once n grows too large.

When $n > 1$ and since $E(Z_t^2) = 1$, we have, by measurability of σ_t w.r.t. \mathcal{F}_{t-1} and independence of Z_t from $\{\mathcal{F}_s\}_{s < t}$,

$$\begin{aligned}\text{var}(X_{t+n}|\mathcal{F}_t) &= E(\sigma_{t+n}^2 Z_{t+n}^2|\mathcal{F}_t) = E[E(\sigma_{t+n}^2 Z_{t+n}^2|\mathcal{F}_{t+n-1})|\mathcal{F}_t] \\ &= E[\sigma_{t+n}^2 E(Z_{t+n}^2|\mathcal{F}_{t+n-1})|\mathcal{F}_t] = E(\sigma_{t+n}^2|\mathcal{F}_t).\end{aligned}$$

Using similar arguments and (3.4),

$$\begin{aligned}E(\sigma_{t+n}^2|\mathcal{F}_t) &= E[E(\sigma_{t+n}^2|\mathcal{F}_{t+n-2})|\mathcal{F}_t] \\ &= E[E(\alpha_0 + \alpha_1 \sigma_{t+n-1}^2 Z_{t+n-1}^2 + \beta \sigma_{t+n-1}^2|\mathcal{F}_{t+n-2})|\mathcal{F}_t] \\ &= E[\alpha_0 + \alpha_1 \sigma_{t+n-1}^2 E(Z_{t+n-1}^2|\mathcal{F}_{t+n-2}) + \beta \sigma_{t+n-1}^2|\mathcal{F}_t] \\ &= E[\alpha_0 + (\alpha_1 + \beta) \sigma_{t+n-1}^2|\mathcal{F}_t] = \alpha_0 + (\alpha_1 + \beta) E(\sigma_{t+n-1}^2|\mathcal{F}_t).\end{aligned}$$

$E(\sigma_{t+n}^2|\mathcal{F}_t) = \alpha_0 + (\alpha_1 + \beta) E(\sigma_{t+n-1}^2|\mathcal{F}_t)$ can be transformed into

$$E(\sigma_{t+n}^2|\mathcal{F}_t) - \frac{\alpha_0}{1 - \alpha_1 - \beta} = (\alpha_1 + \beta) \left[E(\sigma_{t+n-1}^2|\mathcal{F}_t) - \frac{\alpha_0}{1 - \alpha_1 - \beta} \right]. \quad (6.1)$$

Since $\text{var}(X_{t+n}|\mathcal{F}_t) = E(\sigma_{t+n}^2|\mathcal{F}_t)$, repeating (6.1) and rearranging terms produces

$$\text{var}(X_{t+n}|\mathcal{F}_t) = \frac{\alpha_0}{1 - \alpha_1 - \beta} + (\alpha_1 + \beta)^{n-1} \left[E(\sigma_{t+1}^2|\mathcal{F}_t) - \frac{\alpha_0}{1 - \alpha_1 - \beta} \right], \quad (6.2)$$

where

$$E(\sigma_{t+1}^2|\mathcal{F}_t) = \sigma_{t+1}^2 = \alpha_0 + \alpha_1 \sigma_t^2 Z_t^2 + \beta \sigma_t^2$$

as σ_t and Z_t are already observed at time t .

This shows that the conditional variance will converge to the unconditional variance $\frac{\alpha_0}{1-\alpha_1-\beta}$ as the second term in (6.2) decays geometrically (we have previously assumed $\alpha_1 + \beta < 1$ for the sake of covariance stationarity). That is, the variance forecast is eventually completely insensitive to the current information at time t .

We next compute the conditional variance $\text{var}(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t)$. By similar arguments as in (3.1), it can be shown that $\text{cov}(X_{t+i}X_{t+j}|\mathcal{F}_t) = 0$ when $i \neq j$. Therefore,

$$\text{var}\left(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t\right) = \sum_{i=1}^n \text{var}(X_{t+i}|\mathcal{F}_t). \quad (6.3)$$

Further algebra based on (6.2) yields

$$\text{var}\left(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t\right) = \frac{n\alpha_0}{1-\alpha_1-\beta} + \frac{1-(\alpha_1+\beta)^n}{1-\alpha_1-\beta} \left[E(\sigma_{t+1}^2|\mathcal{F}_t) - \frac{\alpha_0}{1-\alpha_1-\beta} \right], \quad (6.4)$$

which implies that when n is large so that $(\alpha_1 + \beta)^n$ is close to 0, the conditional variance will grow linearly in n .

Consider $\text{var}(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t^{[1]})$ and $\text{var}(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t^{[2]})$, a pair of conditional variances with the same GARCH parameter α_0 , α_1 and β but different initial volatility $E(\sigma_{t+1}^2|\mathcal{F}_t^{[1]}) = \sigma_{t+1}^{[1]}$ and $E(\sigma_{t+1}^2|\mathcal{F}_t^{[2]}) = \sigma_{t+1}^{[2]}$. Their difference is given by

$$\frac{1-(\alpha_1+\beta)^n}{1-\alpha_1-\beta} \left[\sigma_{t+1}^{[1]} - \sigma_{t+1}^{[2]} \right], \quad (6.5)$$

which grows/decays nonlinearly at first but ultimately becomes almost a constant. A constant gap between $\text{var}(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t^{[1]})$ and $\text{var}(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t^{[2]})$ means the power of the difference in the conditioning information $\sigma_{t+1}^{[1]}$ and $\sigma_{t+1}^{[2]}$ to influence the variance of the sum $\sum_{i=1}^n X_{t+i}$ has reached its limit.

We expect these results about the variance of sums of high frequency returns to carry over to similar results about VaR. Since the analysis for VaR is more difficult, we instead illustrate the results for VaR via simulation. That is, we fix initial values of σ_t and X_t , set $n = 250$, and compute via MC simulation two sequences of length 250: the conditional variances analyzed above

$$\left\{ \text{var}\left(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t\right) \right\}_{n=1}^{250},$$

and the conditional values-at-risk

$$\{VaR_{95\%}(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t)\}_{n=1}^{250}.$$

Not surprisingly, we found that they are closely correlated with correlation $\rho \approx 0.98$ for a variety of choices of initial conditions. This supports our expectation that the conditional VaR will behave similarly to the conditional variance.

The analog of equation (6.5) for VaR is illustrated by the initially widening and then almost constant gap between the curves in Figure 6, which plots simulated conditional VaR as a function of n for two different initial values of σ_t and Z_t ²¹. The resulting conditional VaR of n -day returns as an asymptotically linear function of the time horizon n is obviously unrealistic and therefore should fail backtests with real data for large enough n .

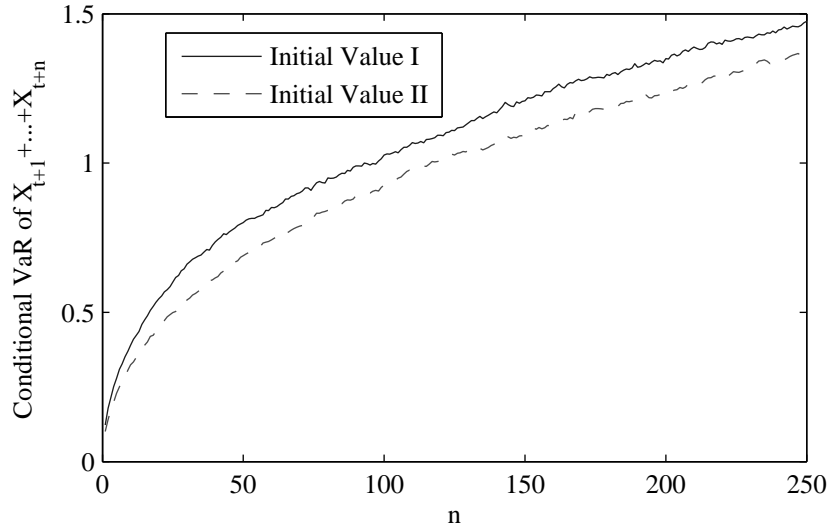


Figure 6: Comparison of $\{VaR_{95\%}(\sum_{i=1}^n X_{t+i}|\mathcal{F}_t)\}_{n=1}^{250}$ via simulation, for two different initial GARCH volatilities

²¹ $\mu_0 = 0, \alpha_0 = 0.00005, \alpha_1 = 0.04, \beta = 0.95; \sigma_t^{[1]} = 0.08, Z_t^{[1]} = 0.375; \sigma_t^{[2]} = 0.05, Z_t^{[2]} = 0.25.$

Results for general GARCH(p,q) are similar. By similar arguments, the iterative equation for the conditional variance of the GARCH(p,q) process is

$$\begin{aligned} E(\sigma_{t+n}^2 | \mathcal{F}_t) &= \frac{\alpha_0}{1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j} \\ &= \sum_{k=1}^{\max(p,q)} (\alpha_k + \beta_k) \left[E(\sigma_{t+n-k}^2 | \mathcal{F}_t) - \frac{\alpha_0}{1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j} \right], \end{aligned} \quad (6.6)$$

where $\alpha_k = 0$ if $k > p$ and $\beta_k = 0$ if $k > q$.

Denote $Y_k = E(\sigma_{t+k}^2 | \mathcal{F}_t) - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j$ and $d = \max(p, q)$, then an array of iterative equations in the form of (6.6) can be rewritten in matrices:

$$\begin{pmatrix} Y_n \\ \vdots \\ Y_{n-d+1} \end{pmatrix} = A \begin{pmatrix} Y_{n-1} \\ \vdots \\ Y_{n-d} \end{pmatrix} = \dots = A^{n-d} \begin{pmatrix} Y_d \\ \vdots \\ Y_1 \end{pmatrix}, \quad (6.7)$$

where

$$A_{ij} = \begin{cases} \alpha_i + \beta_i & \text{if } i = j, \\ 1 & \text{if } i = j + 1, \\ 0 & \text{otherwise.} \end{cases}$$

The lower diagonal matrix A can be diagonalized and then A^{n-d} will be straightforward to compute. Each term in the initial vector can be calculated through the known observations $\{\sigma_{t-i}, Z_{t-i}\}_{i=0}^{d-1}$. Again, thanks to the covariance stationarity condition $\sum_{i=1}^p \alpha_i + \sum_{j=1}^q \beta_j < 1$, the conditional variance will converge to $\alpha_0 / (1 - \sum_{i=1}^p \alpha_i - \sum_{j=1}^q \beta_j)$, the unconditional variance.

6.2 High-Low vs Fixed Frequency with Abundant Data

So far, the high-low frequency approach has been used to reduce reliance on data availability. What if data is in plentiful supply? Will it still be a worthy alternative?

Within a fixed time range, the high-low frequency approach utilizes at least twice as much data as the fixed-frequency approach. Intuitively, this may reveal information not previously captured.

Suppose we have a sequence of n -day negative log returns $X_1^{(L)}, \dots, X_T^{(L)}$ ²² and the calibrated distribution of the filtered data (i.e., innovations $Z_t^{(L)}$) is skewed t

²²L stands for low frequency and H stands for high frequency.

with parameters $(\nu, \mu, \gamma, \sigma)$. Since $X_{T+1}^{(L)} = \mu_0 + \sigma_{T+1} Z_{T+1}^{(L)}$ and σ_{T+1} is a constant at time T , by Proposition 2.5:

$$X_{T+1}^{(L)} \sim SkewT(\nu, \mu_0 + \sigma_{T+1}\mu, \sigma_{T+1}\sigma, \sigma_{T+1}\gamma).$$

On the other hand, if we use the high-low frequency approach based on daily data on the same period $X_1^{(H)}, \dots, X_{nT}^{(H)}$, the distribution of the next n -day return can be determined by simulating the sum $X_{nT+1}^{(H)} + \dots + X_{nT+n}^{(H)}$.

Since both $X_{T+1}^{(L)}$ and $X_{nT+1}^{(H)} + \dots + X_{nT+n}^{(H)}$ denote the negative log return in the same period, we can compare the fixed and high-low frequency approach by plotting the skewed t density of $X_{T+1}^{(L)}$ and the density function²³ of the simulated $X_{nT+1}^{(H)} + \dots + X_{nT+n}^{(H)}$ for various values of T in our standard S&P500 data time series. See Figures 7.A, 7.B, 7.C and 7.D.

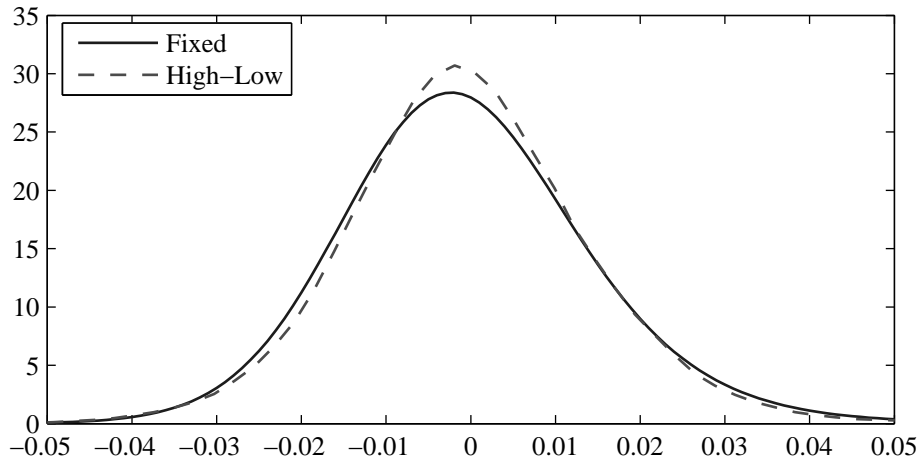


Figure 7.A: Forecast pdf for a 2-Day Negative Log Return on Day 4740 (S&P500 Daily Data)

The difference between fixed-freq density and high-low freq density does not shown any predictable pattern. The high-low freq density can have higher peaks and thinner tails (Figure 7.A), but the reverse can also be true (Figure 7.B). Further, such reversal can be quite volatile. Figure 7.B happens only 10 days (2 weeks) after Figure 7.A, a tiny lag given the 3750 daily (750 weekly) returns used for calibration.

²³Using a kernel smoothing method.

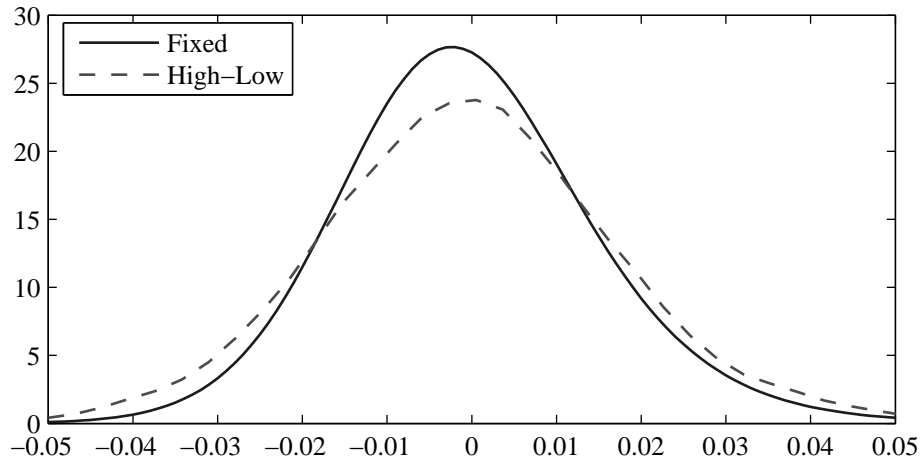


Figure 7.B: Forecast pdf for a 2-Day Negative Log Return on Day 4750 (S&P500 Daily Data)

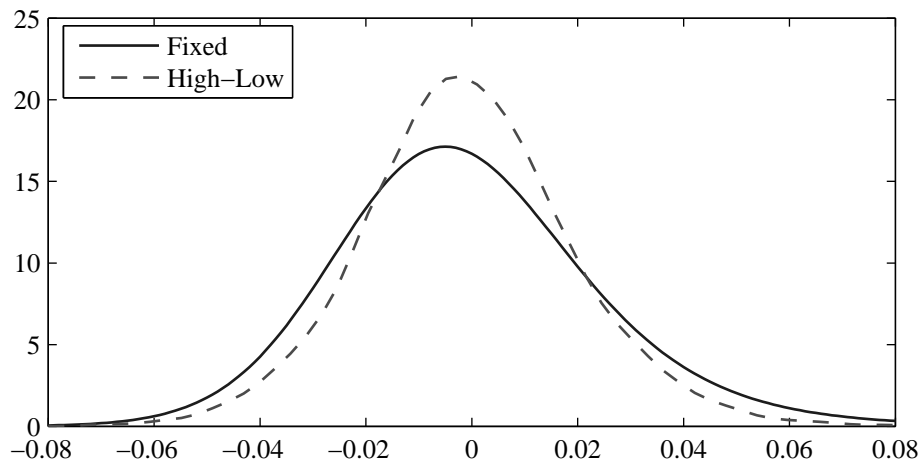


Figure 7.C: Forecast pdf for a 4-Day Negative Log Return on Day 4740 (S&P500 Daily Data)

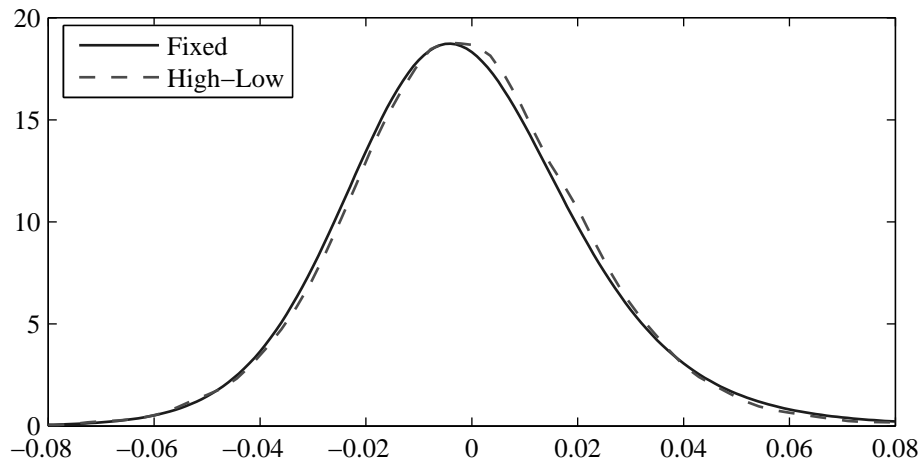


Figure 7.D: Forecast pdf for a 5-Day Negative Log Return on Day 4740 (S&P500 Daily Data)

As time horizon increases, the difference in densities does not widen steadily. For example, there can be greater deviation under 4-day horizon (Figure 7.C) than 2-day (Figure 7.A). However, 5-day densities (Figure 7.D) can be hardly distinguishable.

To further explore their relationship in the long run, we re-examined our S&P500 daily returns series for the years 1971-2009²⁴. After a lead-in period of 750 weeks for fitting the parameters, we made weekly VaR forecasts for the next 1200 weeks, using both the fixed frequency method with a weekly horizon and the high-low frequency method based on daily data.

We have the following observations from Table 6 and Figure 8.

- Daily-based forecasts and weekly-based forecasts are closely correlated, with $\rho \approx 0.92$. As a result, their long-term trends are roughly the same.
- Less than 10% of the weekly-based forecasts fall within the 95% confidence interval of daily-based forecast. So the difference between the two methods are statistically significant.

²⁴VaR forecasts in this study have to be for non-overlapping weeks to make violation indicators independent. Therefore, we again expand our data set.

Table 6: VaR_{Fixed} vs VaR_{HL} : Weekly Forecasts, 1971-2009

α	95%	97.5%	99%	99.5%
Violations of VaR_{Fixed}	62	34	15	8
Violations of VaR_{HL}	60	32	17	10
Simultaneous Violations	50	25	12	5
VaR_{Fixed} outside 95% confidence interval*	1108	1103	1112	1099
VaR_{Fixed} inside 95% confidence interval	92	97	88	101
Correlation of VaR_{Fixed} and VaR_{HL}	0.918	0.917	0.916	0.913
Variance of VaR_{Fixed}	0.0003	0.0004	0.0007	0.001
Variance of VaR_{HL}	0.0005	0.0008	0.0012	0.0016

*Of VaR_{HL} .

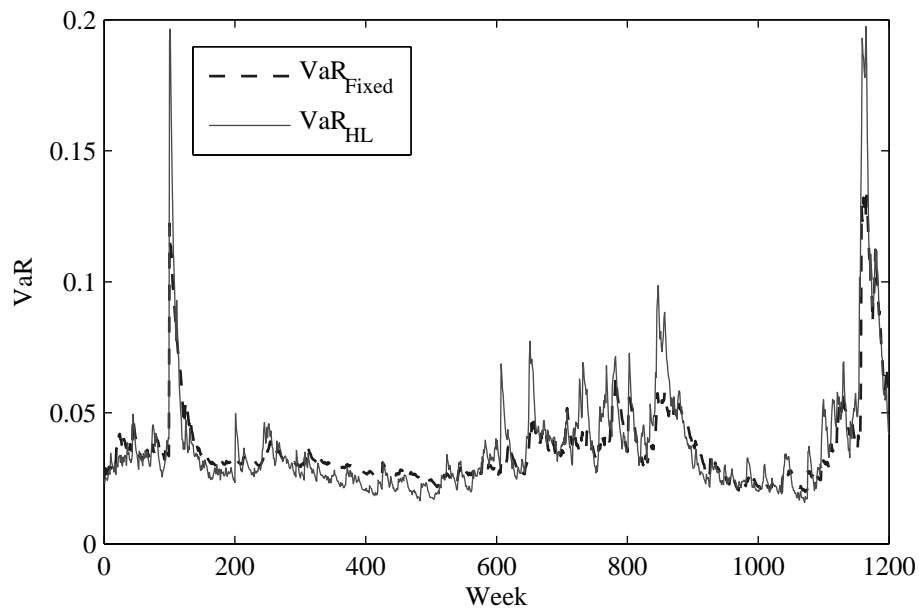


Figure 8: VaR_{Fixed} vs VaR_{HL} : Weekly Forecasts for S&P500 returns, 1971-2009. Week 0 in the figure corresponds to the 750th week after Jan. 1, 1971.

- Both models have passed the backtest, so unsurprisingly they record approximately the same number of VaR violations. However, the timings of these violations are different. More than 85% of $VaR_{95\%}$ violations arrived simultaneously while the rate declines to 60% at the 99.5% level.
- Daily-based forecasts are more volatile, magnifying the fluctuations in weekly-based forecasts.

When we take the fixed-frequency approach, GARCH is applied on the weekly scale. Correspondingly, what's most directly responsible for the VaR forecasts is the return and volatility of the previous week.²⁵ Similarly, when GARCH is applied on the daily scale, as in the high-low frequency method, VaR forecasts are most susceptible to the previous day, not week.

Daily return/volatility and weekly return/volatility are bound to be correlated and exhibit similar patterns when viewed globally. However, having less time to smooth out or recover from extreme events, daily data are more volatile and will give rise to steeper ups and downs as well as forecasts different from the weekly-based ones.

This last feature makes high-low frequency method a quicker responder to volatility changes. For example, if volatility starts to rise only in the last week of a particular month, its severity may be dampened by the first three good or uneventful weeks. By paying more attention to the most recent week, the high-low frequency model will have a better chance to detect this signal which is otherwise masked in a fixed monthly horizon.

Admittedly, the opposite situation can also happen, especially when the conversion factor between high and low frequencies is too large. For example, a daily return/volatility may hardly reflect what's actually happening in an entire month. Despite the possibility of false alarms, high-low frequency approach can at least offer an alternative view of risk. If a risk manager intends to be more cautious, he may implement both models and pick the VaR forecast that is higher.

²⁵Since the innovation terms in a GARCH process are $SWN(0,1)$, the filtered returns will approximately have mean 0 and variance 1. Even though the calibration of skewed t distributions produces different parameters as time evolves, the scale of these distributions will not differ greatly. As a result, σ_{t+1} in the de-filtering/forecasting equation $X_{t+1} = \mu_0 + \sigma_{t+1}Z_{t+1}$ is the most decisive factor for the scale (i.e., variance) of X_{t+1} and consequently the size of VaR.

6.3 A Simulated Weekly Returns Scenario

To illustrate the potential advantage of the high-low frequency approach, we'll compare the two methods using simulated data.²⁶ A weekly returns time series is simulated using GARCH, the parameters of which are derived from historical S&P500 weekly returns. The total length of the simulated series is 1950 months. The first 750 months are used as history to make the first monthly VaR forecast starting at month 751. We then make a series of 1200 monthly VaR forecasts using both the fixed frequency (monthly horizon) and high-low frequency (weekly-monthly) methods.²⁷ Additionally, now that the weekly return process is strictly GARCH, we can nail down the true values of each monthly VaR, again using MC simulation.

As pointed out earlier, the size of VaR is closely related to return volatilities in previous periods, which we'll measure in four ways:

- Standard deviation of weekly returns in the previous month. (*MonthlyStd*)
- Sum of absolute values of weekly returns in the previous month. (*AbsMonthly*)
- Average of weekly volatilities in the previous month. (*VolAvg*)
- Standard deviation of weekly volatilities in the previous month. (*VolStd*)

Table 7.A presents correlations between VaR_{True} , the true monthly VaR; VaR_{HL} , the VaR computed using the high-low frequency method; VaR_{Fixed} , the VaR computed using the fixed frequency method; and the four volatility measures. It shows:

- Among the proposed volatility measures, *VolAvg* is most closely related to VaR_{True} .
- VaR_{HL} are almost perfectly correlated to both VaR_{True} and *VolAvg*.
- The correlations between VaR_{Fixed} and all four volatility measures are weaker than those of VaR_{HL} .

Table 7.B, 7.C and Figure 9.A study errors and show:

²⁶Direct comparison of model performance in terms of error size is impossible when historical data is used, as we do not know the "true" VaR.

²⁷We'll only examine $VaR_{95\%}$ in this section.

Table 7.A: Comparison using Simulated *Weekly* Data: Correlations Part I

	VaR_{True}	VaR_{HL}	VaR_{Fixed}	$VolAvg$	$VolStd$	$AbsMonthly$	$MonthlyStd$
VaR_{True}	1	0.9937	0.7887	0.9886	0.4626	0.6098	0.5960
VaR_{HL}		1	0.7867	0.9860	0.4506	0.5968	0.5807
VaR_{Fixed}			1	0.7791	0.3683	0.4568	0.4443
$VolAvg$				1	0.5195	0.6565	0.6386
$VolStd$					1	0.6424	0.6577
$AbsMonthly$						1	0.9262
$MonthlyStd$							1

Table 7.B: Comparison using Simulated *Weekly* Data: Correlations Part II

	$VolAvg$	$VolStd$	$AbsMonthly$	$MonthlyStd$
Error* of VaR_{HL}	0.1804	0.0358	0.0364	0.0187
Error of VaR_{Fixed}	-0.4452	-0.1996	-0.3127	-0.3042

*Error = (Estimate-True) / True.

Table 7.C: Comparison using Simulated *Weekly* Data: Error

	Mean of Absolute Error*	Std
Error of VaR_{HL}	0.0022	0.0025
Error of VaR_{Fixed}	0.0091	0.0128

*With absolute value taken.

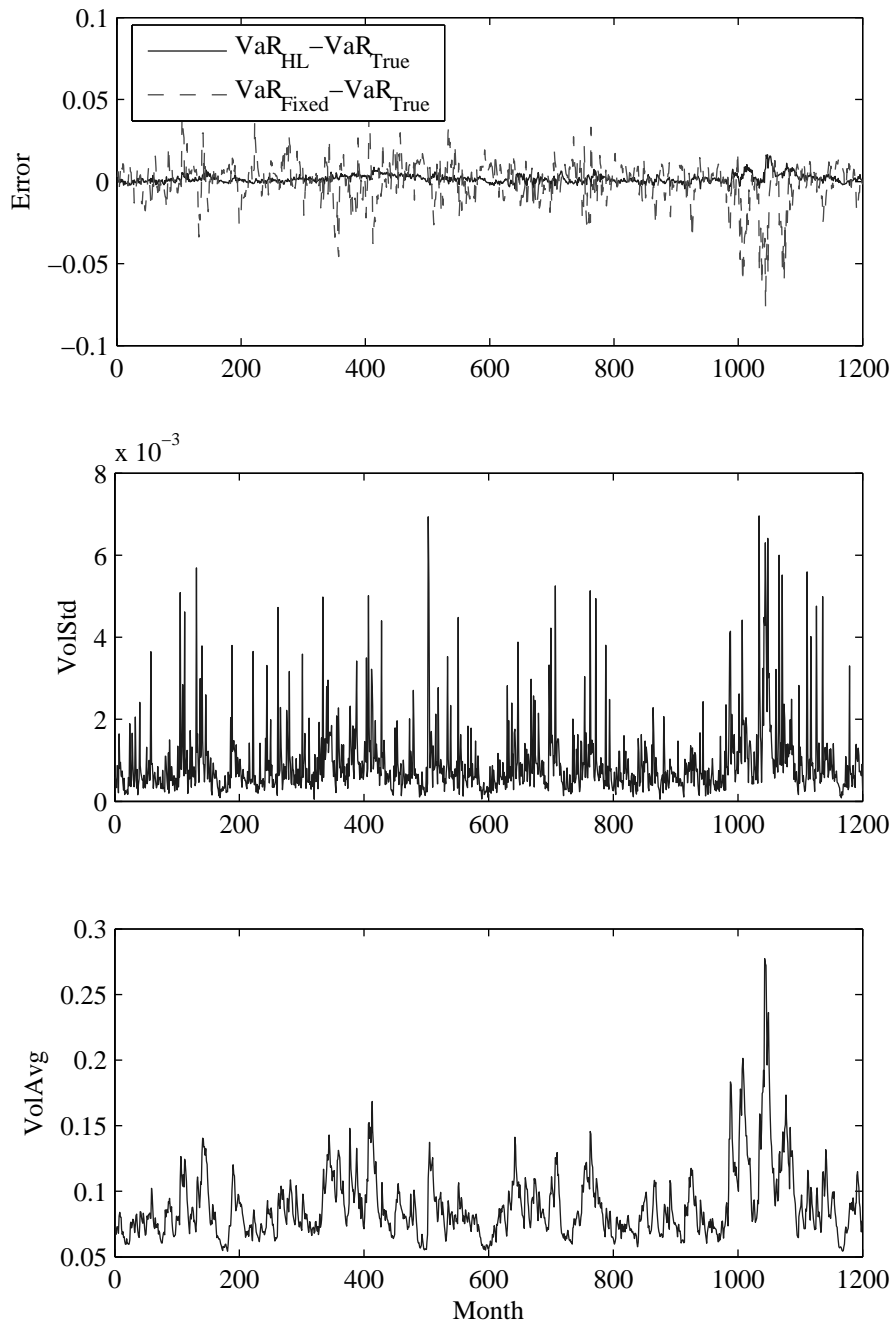


Figure 9.A: Error Comparison of VaR_{Fixed} and VaR_{HL}

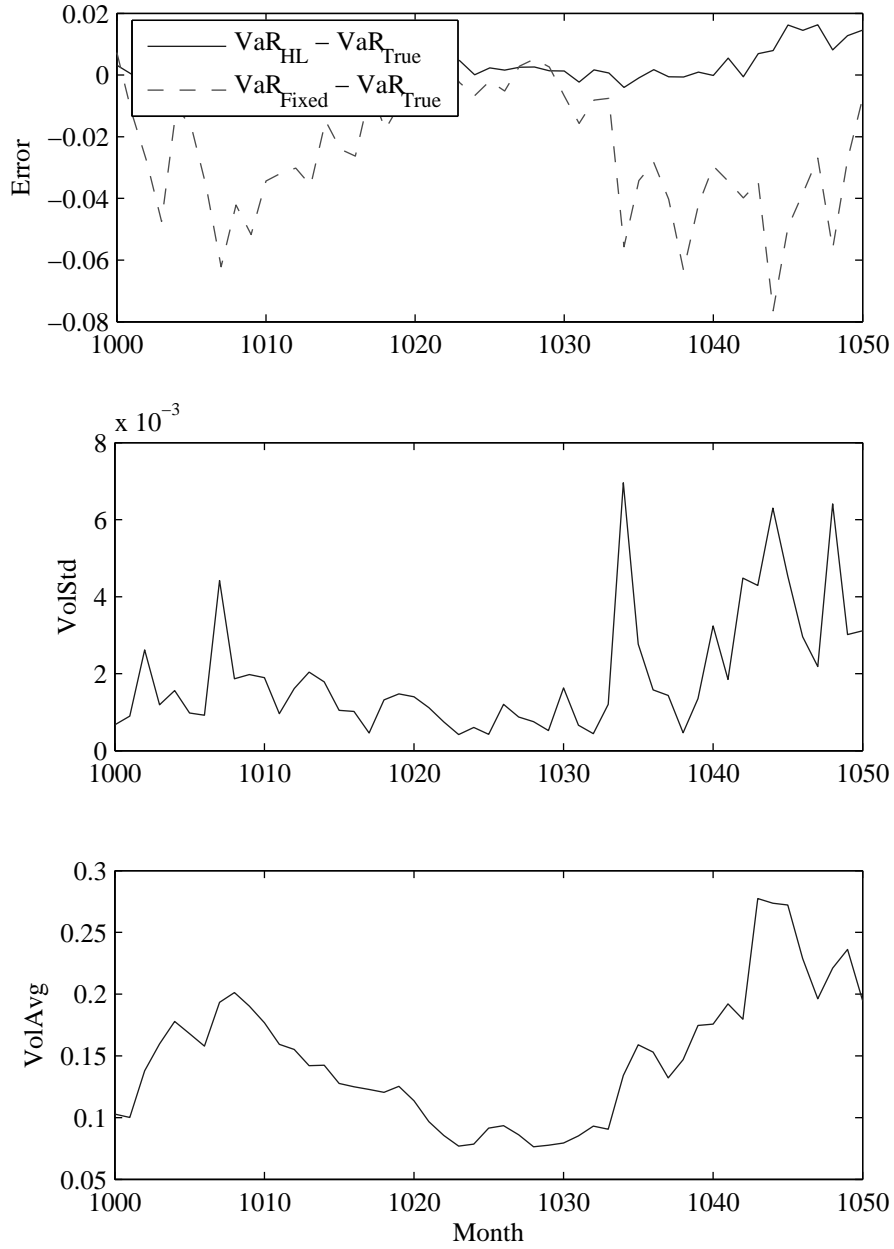


Figure 9.B: Error Comparison of VaR_{Fixed} and VaR_{HL} (Detail, centered at month 1025)

- Relative errors of VaR_{Fixed} are negatively correlated to all four volatility measures. Graphically (Figure 9.A), errors tend to be negative when $VolStd$ rises and vice versa. In other words, the fixed frequency approach underestimates VaR when volatility surges and overestimates VaR when volatility subsides.
- The correlation between relative errors of VaR_{HL} and $VolStd$, $AbsMonthly$ is minimal, which shows that the accuracy of the high-low frequency approach is relatively immune to volatility changes.
- Absolute errors of VaR_{HL} have significantly lower mean and standard deviation compared to absolute errors of VaR_{Fixed} .

At the 1044th month (see Figure 9.B), VaR_{true} for the next month is 22.43%, almost the global maximum in our entire 1200 simulated months (the global maximum VaR_{true} is 22.65%). If we zoom in on the time leading to that month (Figure 9.B), we can see that both $VolAvg$ and $VolStd$ have been increasing since the 1035th month and approaching their respective global maximum as well. During this “chaotic” period, the gap between $VaR_{fixed} - VaR_{true}$ and $VaR_{HL} - VaR_{true}$ is widening steadily. As a result, VaR_{fixed} is merely 14.78%, seriously underestimated. In contrast, VaR_{HL} , at 23.22%, is much closer to VaR_{true} .

6.4 A Simulated Daily Returns Scenario

It can be argued that the simulated weekly GARCH process has an unfair advantage in the above comparison, since the weekly simulation frequency matches the weekly high frequency horizon in our high-low frequency approach. To address²⁸ such bias, we’ll simulate a daily GARCH process, aggregate daily returns into weekly returns and then repeat our earlier experiment using the same weekly/monthly horizons.

As shown by Table 8.A through 8.C, the high-low frequency method still beats the fixed frequency method by displaying stronger correlation with volatility measures and lower errors. However, now that the GARCH mechanism operates on the daily level, filtering the weekly returns with GARCH won’t perfectly capture the varying volatilities. As a result, correlations between errors of VaR_{HL} and

²⁸Completely eliminating this bias is impossible since you cannot simulate a monthly GARCH process and then split it into weekly returns for the high-low frequency method to work on.

Table 8.A: Comparison using Simulated *Daily* Data: Correlations Part I

	VaR_{True}	VaR_{HL}	VaR_{Fixed}	$VolAvg$	$VolStd$	$AbsMonthly$	$MonthlyStd$
VaR_{True}	1	0.8647	0.6339	0.9475	0.5810	0.7308	0.6848
VaR_{HL}		1	0.7659	0.8746	0.4639	0.8233	0.7752
VaR_{Fixed}			1	0.6697	0.3470	0.5419	0.3892
$VolAvg$				1	0.5318	0.6881	0.6378
$VolStd$					1	0.3739	0.3590
$AbsMonthly$						1	0.8938
$MonthlyStd$							1

Table 8.B: Comparison using Simulated *Daily* Data: Correlations Part II

	$VolAvg$	$VolStd$	$AbsMonthly$	$MonthlyStd$
Error of VaR_{HL}	-0.4029	-0.3368	-0.0750	-0.0659
Error of VaR_{Fixed}	-0.6181	-0.4151	-0.4516	-0.5116

Table 8.C: Comparison using Simulated *Daily* Data: Error

	Mean of Absolute Error	Std
Error of VaR_{HL}	0.0074	0.01
Error of VaR_{Fixed}	0.0114	0.0155

volatility measures turn negative. Still, the magnitude of those negative correlation coefficients are smaller than what the fixed frequency method can offer.

Since the stylized facts pervade all scales of financial data, the high-low frequency approach will always uncover volatility changes ignored by the coarser fixed frequency method and thus be a quicker responder. These simulated data studies refer to “weeks” and “months”, but the timescales are in fact arbitrary so long as they are adequately modeled by a GARCH process. The practitioner need only take care that the conversion factor between the two time scales not be too large.

7 Conclusion

Financial returns are fat-tailed, heteroskedastic, and exhibit serial dependence. To make a risk forecast based on historical returns, we need to remove the serial dependence and calibrate a fat-tailed distribution to the filtered series. This is accomplished using GARCH as a filter, and GH distributions, notably the skewed t distribution, for calibration. This approach is successful provided we have sufficient data for the calibration.

The focus of the paper has been to examine a way to use higher frequency data to form a lower frequency risk forecast by using the calibrated GARCH process to forecast the intermediate time steps. When the ratio between the long horizon and the short horizon is no more than about ten, this works well in our studies based on S&P500 daily returns, and for simulated data.

Risk managers making fixed-horizon risk forecasts should consider this multiscale method because

- by reducing the required lead-in period by at least a factor of two, it makes forecasts possible for a security with a history otherwise too short to train the GARCH filter and the skewed t distribution,
- it enables the removal of outdated and thus irrelevant data in earlier periods,
- the greater abundance of higher frequency data makes statistical estimation more robust and more stable numerically, and
- risk forecasts will be quicker to react to changing conditions, for example by better reflecting a change in the volatility regime that might occur late in the previous period.

All the results reported here pertain to the one-dimensional case of returns of a single index; naturally it will be important to carry out a similar investigation in the multi-dimensional case of portfolios of N assets. In that case, the forecast will be a forecast of the N -dimensional distribution of the N -vector of all the asset returns, from which the portfolio return is obtained, as usual, as an inner product with the vector of portfolio holdings. The full N -dimensional fitted distribution of returns will be required for typical portfolio optimization applications. This is the subject of subsequent research.

References

- Aas, K. and I. Hobaek Haff (2006). The generalized hyperbolic skew student's t -distribution. *Journal of Financial Econometrics* 4(2).
- Artzner, P., F. Delbaen, J. Eber, and D. Heath (1999). Coherent measures of risk. *Mathematical Finance* 9, 203–228.
- Barndorff-Nielsen, O. E. (1977). Exponentially decreasing distributions for the logarithm of the particle size. *Proceedings of the Royal Society, Series A: Mathematical and Physical Sciences* 350, 401–419.
- Barndorff-Nielsen, O. E. (1978). Hyperbolic distribution and distribution on hyperbolae. *Scandinavian Journal of Statistics* 5, 151–157.
- Barndorff-Nielsen, O. E. and P. Blæsild (1981). Hyperbolic distributions and ramifications: contribution to theory and applications. In C. Taillie, G. Patil, and B. Baldessari (Eds.), *Statistical Distributions in Scientific Work*, Volume 4, pp. 19–44. Dordrecht: Reidel.
- Bougerol, P. and N. Picard (1992). Stationarity of GARCH processes and of some non-negative time series. *Journal of Econometrics* 52, 115–127.
- Brandt, A. (1986). The stochastic equation $y_{n+1} = a_n y_n + b_n$ with stationary coefficients. *Advances in Applied Probability* 18, 211–220.
- Brockwell, P. J. and R. A. Davis (2002). *Introduction to Time Series and Forecasting* (2nd ed.). Springer.
- Casella, G. and R. L. Berger (2002). *Statistical Inference*. Pacific Grove, CA: Duxbury.

- Glasserman, P. (2003). *Monte Carlo Methods in Financial Engineering*. Springer.
- Hu, W. (2005). *Calibration of Multivariate Generalized Hyperbolic Distributions using the EM Algorithm, with Applications in Risk Management, Portfolio Optimization and Portfolio Credit Risk*. Phd dissertation, Florida State University, Tallahassee, FL.
- Hu, W. and A. N. Kercheval (2007). Risk management with generalized hyperbolic distributions. In *Proceedings of the Fourth IASTED International Conference on Financial Engineering and Applications*, pp. 19–24. ACTA Press, Calgary.
- Hu, W. and A. N. Kercheval (2008). The skewed t distribution for portfolio credit risk. *Advances in Econometrics* 22, 55–83.
- Hu, W. and A. N. Kercheval (2010). Portfolio optimization for t and skewed t returns. *Quantitative Finance* 10(1), 91–105.
- Jørgensen, B. (1982). *Statistical Properties of the Generalized Inverse Gaussian Distribution*. Lecture notes in statistics. Springer, Heidelberg.
- Keel, S., F. H. and H. Geering (2006). Optimal portfolios with skewed and heavy-tailed distributions. *Proceedings of the Third IASTED International Conference on Financial Engineering and Applications*, 42–48.
- Kupiec, P. (1995). Techniques for verifying the accuracy of risk measurement models. *Journal of Derivatives Winter*, 73–84.
- Liu, C. and D. B. Rubin (1994). The ECME algorithm: a simple extension of EM and ECM with faster monotone convergence. *Biometrika* 81, 633–648.
- Markowitz, H. (1952). Portfolio selection. *Journal of Finance* 7(1), 77–91.
- McNeil, A., R. Frey, and P. Embrechts (2005). *Quantitative Risk Management: Concepts, Techniques and Tools*. Princeton University Press.
- Protasov, R. S. (2004). EM-based maximum likelihood parameter estimation of multivariate generalized hyperbolic distributions with fixed λ . *Statistics of Computing* 14, 67–77.
- Rockafellar, R. and S. Uryasev (2002). Conditional value-at-risk for general loss distributions. *Journal of Banking and Finance* 26, 1443–1471.

Serfling, R. J. (1980). *Approximation Theorems of Mathematical Statistics*. Wiley, New York.

Sharpe, W. F. (1964). Capital asset prices - a theory of market equilibrium under conditions of risk. *Journal of Finance* 19(3), 425–42.