Nonlinear expectation and application to VaR

Shige Peng

(Joint work with Shuzhen Yang & Jianfeng Yao)

Feb. 25, 2021

Shandong University

Uncertainty of Probabilities $\{P_\theta\}_{\theta\in\Theta}$ Nonlinearities of Expectation: $\hat{\mathbb{E}}[X]$

$$\hat{\mathbb{E}}[X] := \sup_{\theta\in\Theta} E_{P_\theta}[X]$$

- Super-expectation

$$\mathbb{E}: \text{ A linear space of random variables } \mapsto \mathbb{R}$$

- $\mathbb{E}$ is a nonlinear functional
  - **Monotonicity:** $\mathbb{E}[X] \geq \mathbb{E}[Y]$ if $X \geq Y$
  - **Constant preserving:** $\mathbb{E}[c] = c$;
  - **Sublinearity:** $\mathbb{E}[X+Y] \leq \mathbb{E}[X] + \mathbb{E}[Y]$ and

$$\mathbb{E}[\lambda X] = \lambda \mathbb{E}[X], \quad \forall \lambda \geq 0$$

Expectation nonlinearity v.s. Probability & distribution uncertainty

$$\mathbb{E}[X] = \max_{\theta\in\Theta} E_\theta[X] = \max_{\theta\in\Theta} \int_\Omega X dP_\theta,$$

$\{P_\theta\}_{\theta\in\Theta}$: probability model uncertainty

The corresponding distribution uncertainty is

$$F_\theta(x) := P_\theta(X \leq x)$$

It's also a sublinear expectation

$$\mathbb{F}[\varphi] = \sup_{\theta\in\Theta} F_\theta[\varphi] = \sup_{\theta\in\Theta} \int_\Omega \varphi(X) dP_\theta$$

Nonlinear version of i.i.d. of two random variables X and Y

Definition

- X and Y have the same distribution uncertainty

$$X \overset{d}{=} Y \iff \mathbb{E}[\varphi(X)] = \mathbb{E}[\varphi(Y)],$$
$$X \overset{d}{\geq} Y \iff \mathbb{E}[\varphi(X)] \geq \mathbb{E}[\varphi(Y)],$$

- Y is indenpendent of X if

$$\mathbb{E}[\varphi(X, Y)] = \mathbb{E}[\mathbb{E}[\varphi(x, Y)]_{x=X}].$$

A general random generator of nonlinear i.i.d. sequence $\{\xi_i\}_{i=1}^{\infty}$

- The ‘black box’ is a random generator producing random vectors

$$\xi_i(\omega), \quad i = 1, 2, 3, \cdots$$

such that

$$\mathcal{L}(\xi_i) \in \{F_\theta\}_{\theta \in \Theta}$$

A general random generator of nonlinear i.i.d. sequence $\{\xi_i\}_{i=1}^{\infty}$

- The 'black box' is a random generator producing random vectors

$$\xi_i(\omega), \quad i = 1, 2, 3, \cdots$$

such that

$$\mathcal{L}(\xi_i) \in \{F_\theta\}_{\theta \in \Theta}$$

- It is easy to check that

$$\xi_{i+1} \overset{d}{=} \xi_i, \qquad \text{and} \ \ \xi_{i+1} \text{ is independent of } (\xi_1, \xi_2, \cdots, \xi_i)$$

Universality of i.i.d.

- Given a random sequence of data $\{X_i\}_{i=1}^{\infty}$, one can always find a strong enough super expectation $\mathbb{E}$ such that $\{X_i\}_{i=1}^{\infty}$ is an i.i.d. with respect $\mathbb{E}$.
- In principle, one can use a more robust $\hat{\mathbb{E}}[\cdot]$ to ensure that $\{X_i\}_{i=1}^{\infty}$ is i.i.d. under $\mathbb{E}[\cdot]$.

Nonlinear Law of large numbers

Theorem (P. 2007)

Let $\{Y_i\}_{i=1}^{\infty}$ *be IID sequence. Assume* $\lim_{c\to\infty}\mathbb{E}[(|Y_1|-c)^+]=0$.

Nonlinear Law of large numbers

Theorem (P. 2007)

Let $\{Y_i\}_{i=1}^{\infty}$ *be IID sequence. Assume* $\lim_{c\to\infty}\mathbb{E}[(|Y_1|-c)^+]=0.$.
Then, for each $\varphi\in C_b(\mathbb{R})$,

$$\lim_{n\to\infty}\mathbb{E}[\varphi(\frac{Y_1+\cdots+Y_n}{n})]=\mathbb{E}[(\varphi(Y)]=\max_{v\in[\underline{\mu},\overline{\mu}]}\varphi(v).$$

where $\overline{\mu}=\mathbb{E}[Y_1]$, $\underline{\mu}=-\mathbb{E}[-Y_1]$.

Nonlinear Law of large numbers

Theorem (P. 2007)

Let $\{Y_i\}_{i=1}^{\infty}$ *be IID sequence. Assume* $\lim_{c\to\infty}\mathbb{E}[(|Y_1|-c)^+]=0.$.
Then, for each $\varphi\in C_b(\mathbb{R})$,

$$\lim_{n\to\infty}\mathbb{E}[\varphi(\frac{Y_1+\cdots+Y_n}{n})]=\mathbb{E}[(\varphi(Y)]=\max_{v\in[\underline{\mu},\overline{\mu}]}\varphi(v).$$

where $\overline{\mu}=\mathbb{E}[Y_1]$, $\underline{\mu}=-\mathbb{E}[-Y_1]$.

$$Y\overset{d}{=}M_{[\overline{\mu},\underline{\mu}]}:\ \textit{Maximal distribution.}$$

Nonlinear Law of large numbers

Theorem (P. 2007)

Let $\{Y_i\}_{i=1}^{\infty}$ *be IID sequence. Assume* $\lim_{c\to\infty}\mathbb{E}[(|Y_1|-c)^+]=0.$.
Then, for each $\varphi\in C_b(\mathbb{R})$,

$$\lim_{n\to\infty}\mathbb{E}[\varphi(\frac{Y_1+\cdots+Y_n}{n})]=\mathbb{E}[(\varphi(Y)]=\max_{v\in[\underline{\mu},\overline{\mu}]}\varphi(v).$$

where $\overline{\mu}=\mathbb{E}[Y_1]$, $\underline{\mu}=-\mathbb{E}[-Y_1]$.

$$Y\overset{d}{=}M_{[\overline{\mu},\underline{\mu}]}:\ \textit{Maximal distribution.}$$

Nonlinear Central limit theorem

Theorem (P. 2008-2010)

Let $\{X_i\}_{i=1}^{\infty}$ be IID sequence. We assume furthermore that

$$\mathbb{E}[|X_1|^3] < \infty \quad \mathbb{E}[X_1] = \mathbb{E}[-X_1] = 0$$

Then, for each $\varphi \in C_b(\mathbb{R})$,

$$\lim_{n\to\infty} \mathbb{E}[\varphi(\frac{X_1 + \cdots + X_n}{\sqrt{n}})] = \mathbb{E}[\varphi(X)].$$

where X is $N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$-distributed.

Nonlinear Central limit theorem

Theorem (P. 2008-2010)

Let $\{X_i\}_{i=1}^{\infty}$ be IID sequence. We assume furthermore that

$$\mathbb{E}[|X_1|^3] < \infty \quad \mathbb{E}[X_1] = \mathbb{E}[-X_1] = 0$$

Then, for each $\varphi \in C_b(\mathbb{R})$,

$$\lim_{n\to\infty} \mathbb{E}[\varphi(\frac{X_1 + \cdots + X_n}{\sqrt{n}})] = \mathbb{E}[\varphi(X)].$$

where X is $N(0, [\underline{\sigma}^2, \overline{\sigma}^2])$-distributed.

The distributions of $X \overset{d}{=} M_{[\underline{\mu},\bar{\mu}]}$ and $Y \overset{d}{=} N(0, [\underline{\sigma}^2, \bar{\sigma}^2])$

Valuation of $\mathbb{E}[\varphi(X)]$

$\mathbb{E}[\varphi(x + tX)]$=u(t,x) solves the PDE

$$\partial_t u = g(D_x u), \quad u(0, x) = \varphi(x)$$

where $g(p) = \overline{\mu} p^+ - \underline{\mu} p^-$, $\quad \overline{\mu} := \mathbb{E}[X_1]$, $\underline{\mu} := -\mathbb{E}[-X_1]$

Valuation of $\mathbb{E}[\varphi(Y)]$

$\mathbb{E}[\varphi(x + \sqrt{t}Y)] := v(t, x)$ solves the PDE

$$\partial_t v = G(D_x^2 v), \ \ v(0, x) = \varphi(x),$$

where $G(a) = \frac{\bar{\sigma}^2}{2} a^+ - \frac{\underline{\sigma}^2}{2} a^-$, $\quad \bar{\sigma}^2 := \mathbb{E}[Y_1^2]$, $\underline{\sigma}^2 := -\mathbb{E}[-Y_1^2]$.

Optimality of the estimate

The optimality of the above estimate is based on the following quite simple, but very fundamental result:

Theorem (Jin-P. 2016)

Let $Y^1, \cdots, Y^m$ be i.i.d. and maximally distributed:

$$Y^i \overset{d}{=} M_{[\underline{\mu},\overline{\mu}]}, \quad i = 1, \cdots, m,$$

where $\underline{\mu} \leq \overline{\mu}$ is two unknown parameters. Then

Optimality of the estimate

The optimality of the above estimate is based on the following quite simple, but very fundamental result:

Theorem (Jin-P. 2016)

Let $Y^1, \cdots, Y^m$ be i.i.d. and maximally distributed:

$$Y^i \overset{d}{=} M_{[\underline{\mu},\overline{\mu}]}, \qquad i = 1, \cdots, m,$$

where $\underline{\mu} \leq \overline{\mu}$ is two unknown parameters. Then

$$\underline{\mu} \leq \min\{Y^1(\omega), \cdots, Y^n(\omega)\} \leq \max\{Y^1(\omega), \cdots, Y^n(\omega)\} \leq \overline{\mu}.$$

Moreover

$$\widehat{\overline{\mu}}_n = \max\{Y^1, \cdots, Y^n\},$$

is the maximum unbiased estimate of $\overline{\mu}$.

Application to VaR

The following two papers are about the application of nonlinear expectation of VaR

- P.,-Yang-Yao. Improving Value-at-Risk prediction under model uncertainty, arXiv:1805.03890,1-42, Journal of Financial Econometrics, 1-30, June 2020 (Accept).
- P.- Yang. Autoregressive models of the time series under volatility uncertainty and application to VaR model, arXiv:2011.09226, 1-29.

In the following, we report the details of the first paper **P.-Yang-Yao** (2020a).

Outline

- Background of VaR and G-expectation
- G-normal distribution and G-VaR
- Implementation of G-VaR
- Empirical results of G-VaR
- Conclusion

Background-History of VaR

- Since its birth at J.P. Morgan in the 1990s, value-at-risk (VaR) has become one of the most used instruments for assessing downside risk in **financial markets**.
- The regulatory authorities also incorporate VaR measures into their recommendations to the banking industry (**Basel Accords I-III**), which has accelerated the spread of VaR.

Background-History of VaR

- For example, one of the best performers is obtained by applying extreme value theory (EVT) to the residuals of AR-GARCH fit using skewed-t innovations (AR-GARCH St-EVT).
- **Kuester et al. (2006)** conclude that, at least for the **NASDAQ Composite Index**, "conditionally heteroskedastic models yield acceptable forecasts" and that the conditional skewed-t (**AR-GARCH-St**) together with the conditional skewed-t coupled with EVT (**AR-GARCH-St-EVT**) perform best in general.

Benchmark predictors for VaR

- **AR-GARCH filtering (AR-GARCH)**: returns are assumed to follow a mean-variance decomposition of type

$$r_t = \mu_t + \sigma_t z_t, \tag{1}$$

where the mean process (μ_t) follows an AR model (or, more generally, an ARMA model) and the residual is modeled by a GARCH process (Bollerslev, 1986).

Background-Sublinear Expectation

- The point of view of SLE theory is radically different: instead of assuming the existence of one unique model P_0, it views returns as originating from a large number of different models, say $\{P_\theta\}_{\theta\in\Theta}$, and this family of potential models is of infinitely dimension.

G-normal distribution

- We consider a simple case, where mean μ is constant (independent of time) and variance σ^2 is time-varying within some interval $[\underline{\sigma}^2, \overline{\sigma}^2]$.
- The interval $[\underline{\sigma}^2, \overline{\sigma}^2]$ is used to characterize the unknown family of distributions $\{F_\theta\}_{\theta \in \Theta}$.

G-VaR

- Consider X under the model uncertainty represented by a family of distributions $\{F_\theta(x)\}_{\theta\in\Theta}$. The VaR of X under each F_θ is

$$\mathrm{VaR}_\alpha^{F_\theta}(X) = -\inf\{x \ : \ F_\theta(x) > \alpha\}.$$

- The worst-case VaR of X is here defined as

$$\mathrm{VaR}_\alpha^*(X) := \sup_{\theta\in\Theta} \mathrm{VaR}_\alpha^\theta(X). \tag{2}$$

The representative distribution for G-VaR

Remark

The G-VaR-distributions is a given family of distributions F_θ

$$VaR^*_\alpha(X) := \sup_{\theta\in\Theta} VaR^{F_\theta}_\alpha(X) = VaR^{\hat{F}}_\alpha(X).$$

G-VaR

- In fact, function $\hat{F}$ has the following closed-form expressions;

$$\hat{F}(x) = \frac{2\overline{\sigma}}{\overline{\sigma}+\underline{\sigma}}\Phi(\frac{x}{\overline{\sigma}})\ I(x \leq 0) + \left\{1 - \frac{2\underline{\sigma}}{\overline{\sigma}+\underline{\sigma}}\Phi(-\frac{x}{\underline{\sigma}})\right\}\ I(x > 0), \tag{3}$$

where Φ denotes the distribution function of the standard normal.

- This G-normal distribution has a negative mean $\sqrt{\frac{2}{\pi}}(\underline{\sigma}-\overline{\sigma})$, and a negative skew. As an example, the G-normal density function with parameters $(\underline{\sigma},\overline{\sigma})=(0.5,1)$ is compared to the standard normal density in Figure 1.

G-VaR

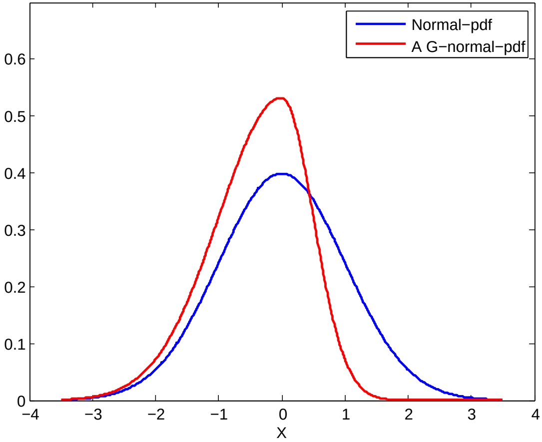

Figure: 1. Density of a G-normal distribution with variance parameters $(\underline{\sigma}, \overline{\sigma}) = (0.5, 1)$ in comparison with standard normal density.

Implementation of G-VaR

- Let $\{X_t\}_{0\leq t\leq T}$ be a return time-series from a risk asset.
  We assume that

$$X_s \overset{d}{=} N(0, [\underline{\sigma}^2, \bar{\sigma}^2]), \qquad s = t - W, \cdots, t, t+1$$

$$X_{t+1} \text{ is independent from } X_0, \cdots, X_t$$

- The sample size W is fixed
- At each time t, we use the data $X_{t-W}, \cdots, X_t$ to estimate $\bar{\sigma}, \underline{\sigma}$ of X_{t+1} and use them to forecast the VaR of X_{t+1}.

The estimators of $\underline{\sigma}^2$ and $\bar{\sigma}^2$

- We have, by nonlinear LLN

$$\frac{1}{m}(X_1^2 + \cdots + X_m^2) \overset{m\to\infty}{\Longrightarrow} M_{[\underline{\sigma}^2, \bar{\sigma}^2]}$$

Implementation of G-VaR

- If $(Y_1, \cdots, Y_n)$ is an i.i.d. sample from a maximal distribution $M_{[\underline{\mu}, \overline{\mu}]}$, then
$$\underline{\mu} \le \min\{Y_1, \cdots, Y_n\} \le \max\{Y_1, \cdots, Y_n\} \le \overline{\mu},$$
- $\hat{\overline{\mu}} = \max\{Y_1, \cdots, Y_n\}$ is the best unbiased estimator of $\bar{\mu}$,
- $\hat{\underline{\mu}} = \min\{Y_1, \cdots, Y_n\}$ is the best unbiased estimator of $\underline{\mu}$ (see Theorem 24 in Jin and P. (2016) shows that)

φ-max-mean algorithm of X from i.i.d. sample $\{X_i\}_{i=1}^{mn}$ Nonlinear Monté-Carlo method

$$\max\left\{\underbrace{\frac{X_1^2+\cdots+X_n^2}{n}}_{Y_n^0},\cdots,\underbrace{\frac{\cdots}{n}}_{Y_n^k},\cdots,\underbrace{\frac{X_{(m-1)n+1}^2+\cdots+X_{mn}^2}{n}}_{Y_n^{m-1}}\right\}$$

Implementation of G-VaR

- Let $W_0 \leq W$ be the window width. The following moving window approach is then employed. For each time s, let

$$\hat{\sigma}^2_{s,W_0} = \frac{1}{W_0} \sum_{j=1}^{W_0} X^2_{s-j+1}$$

be the sample variance from the sample $(X_{s-W_0+1}, \ldots, X_s)$, that is, the history of length W_0 before time s.

Implementation of G-VaR

- Let $k = \lfloor \frac{W}{W_0} \rfloor$ be the largest integer satisfying $kW_0 \leq W$.
- By nonlinear law of large numbers, $\hat{\sigma}^2_{\bar{t}-s,W_0} \sim Max_{[\underline{\sigma}^2,\bar{\sigma}^2]}$
- Clearly: $\hat{\sigma}^2_{\bar{t}-s,W_0}, s = 0, W_0, 2W_0, \cdots, (k-1)W_0$ are independent We denote

$$\hat{\bar{\sigma}}^2_{\bar{t},k} = \max\{\hat{\sigma}^2_{\bar{t}-s,W_0} : \ s = 0, W_0, 2W_0, \cdots, (k-1)W_0\},$$
$$\hat{\underline{\sigma}}^2_{\bar{t},k} = \min\{\hat{\sigma}^2_{\bar{t}-s,W_0} : \ s = 0, W_0, 2W_0, \cdots, (k-1)W_0\},$$
$$\hat{\bar{\sigma}}^2_{\bar{t},W_0} = \max\{\hat{\sigma}^2_{\bar{t}-s,W_0} : \ 0 \leq s \leq W - W_0\},$$
$$\hat{\underline{\sigma}}^2_{\bar{t},W_0} = \min\{\hat{\sigma}^2_{\bar{t}-s,W_0} : \ 0 \leq s \leq W - W_0\}.$$

Implementation of G-VaR

By Jin and P. (2016),

- $\hat{\overline{\sigma}}^2_{\bar{t},k}$ is the largest unbiased estimator for the upper mean $\overline{\sigma}^2_{\bar{t}}$;
- $\hat{\underline{\sigma}}^2_{\bar{t},k}$ is the smallest unbiased estimator for the lower mean $\underline{\sigma}^2_{\bar{t}}$.
- Also note that

$$\underline{\sigma}^2_{\bar{t}} \lesssim \hat{\underline{\sigma}}^2_{\bar{t},W_0} \leq \hat{\underline{\sigma}}^2_{\bar{t},k} \leq \hat{\overline{\sigma}}^2_{\bar{t},k} \leq \hat{\overline{\sigma}}^2_{\bar{t},W_0} \lesssim \overline{\sigma}^2_{\bar{t}},$$

  which shows that, under the given length of W historical data and W_0,
- $\hat{\overline{\sigma}}^2_{\bar{t},W_0}$ is a best estimator for the upper volatility $\overline{\sigma}^2_{\bar{t}}$,
- $\hat{\underline{\sigma}}^2_{\bar{t},W_0}$ a best estimator for the lower volatility $\underline{\sigma}^2_{\bar{t}}$.

Implementation of G-VaR

Consequently, the final G-VaR estimate for the VaR of $X_{\bar{t}+1}$ at level α is

$$\text{G-VaR}_{\alpha,\bar{t}}^{W_0}(X_{\bar{t}+1}) = -\left\{\hat{F}_{\bar{t}}^{W_0}\right\}^{-1}(\alpha). \tag{4}$$

where

$$\begin{aligned}\hat{F}_{\bar{t}}^{W_0}(x) = \; & \frac{2\hat{\overline{\sigma}}_{\bar{t},W_0}}{\hat{\overline{\sigma}}_{\bar{t},W_0}+\hat{\underline{\sigma}}_{\bar{t},W_0}}\Phi\Big(\frac{x}{\hat{\overline{\sigma}}_{\bar{t},W_0}}\Big)\; I(x \leq 0) \\ & + \left\{1 - \frac{2\hat{\underline{\sigma}}_{\bar{t},W_0}}{\hat{\overline{\sigma}}_{\bar{t},W_0}+\hat{\underline{\sigma}}_{\bar{t},W_0}}\Phi\Big(-\frac{x}{\hat{\underline{\sigma}}_{\bar{t},W_0}}\Big)\right\}\; I(x > 0).\end{aligned} \tag{5}$$

Implementation of G-VaR: How to choose W_0?

We say that W_0 is *adaptive* if there exists a positive constant α_{W_0} such that, for sufficiently large n,

$$\frac{1}{n-W}\sum_{\bar{t}=W}^{n} I(X_{\bar{t}+1} < -\text{G-VaR}_{\alpha,\bar{t}}^{W_0}(X_{\bar{t}+1})) \sim \alpha_{W_0}.$$

Namely, the expected violation rate value is the mean value of violation rate.

Remark (6)

See $\{X_t\}_{0\le t\le T}$, window W, and risk level α, a robust adaptive W_0 with $\alpha_{W_0} = \alpha$ can be always approximately found.

Implementation of G-VaR

- Remark 6 guarantees coherence between the empirical percentages of violations and G-VaR measure $\text{G-VaR}_{\alpha,\bar{t}}^{W_0}$ in (4) for the entire dataset $\{X_t\}_{1\le t\le T}$ with historical window size W and estimation window size W_0.

Empirical results of G-VaR

The G-VaR forecasts are evaluated for the NASDAQ Composite Index and S&P 500 Index.

- **Step 1 - data preparation:** NASDAQ Composite Index is denoted by $\{Z_{1,t}\}$, running from February 8,1971 to June 22, 2001, with a total of $N = 7675$ observations of percentages.
- S&P 500 Index is denoted by $\{Z_{2,t}\}$, running from January 3, 2000 to February 7, 2018, with a total of $N = 4550$ observations. Their daily log-returns are

$$r_{i,t} = 100(\ln Z_{i,t} - \ln Z_{i,t-1}), \quad i = 1, 2.$$

Empirical results of G-VaR

- Kuester et al. (2006) found that when using a historical window $W = 1000$, the best VaR predictions for the NASDAQ index are obtained by AR-GARCH filtered modeling such as there commended AR-GARCH-Skewed-t or AR-GARCH-Skewed- t-EVT models.
- The G-VaR predictor proposed in this paper is compared with these two benchmarks, as well as with a more traditional AR-GARCH-Normal predictor using standard normals for the filtered residuals.

Empirical results of G-VaR

- **Step 2 - AR(1) filtering:** To carry out the G-VaR prediction, we first filter the data with the following AR(1) process; that is, the series $r_{1,t}$ and $r_{2,t}$ satisfy the model equations

$$\begin{aligned} r_{1,t} &= a_1 r_{1,t-1} + \epsilon_1 \\ r_{2,t} &= a_2 r_{2,t-1} + \epsilon_2, \end{aligned} \tag{6}$$

where the ϵ_i follow G-normal distributions $N(0, [\underline{\sigma}_i^2, \overline{\sigma}_i^2])$, $i = 1, 2$, respectively.

Empirical results of G-VaR

- **Step 3 - selection of historical and estimation window lengths W and W_0:** The implementation of the G-VaR requires the values of the two window lengths W and W_0 for a given risk level α. Note that, in our G-VaR model, W_0 is dependent on α and W.

- Similarly to Kuester et al. (2006), we consider three historical windows, W=1000, 500, and 250. The corresponding values of W_0 are selected empirically to ensure that the adaptivity (Remark 6).

NASDAQ Composite Index

- We first compare the G-VaR model with the AR(1)-GARCH(1,1)-Normal, AR(1)-GARCH(1,1)-Skewed-t, and AR(1)-GARCH(1,1)-Skewed-t-EVT VaR models. For given windows W=1000, 500, and 250, we show how the adaptivity works for W_0.

- For example, for a given $W = 1000, \alpha = 0.01$, and time point $\bar{t}$, we calculate the G-VaR of $r_{1,\bar{t}}$ (NASDAQ return) with different $W_0 \leq W$.

NASDAQ Composite Index

- Then, for sufficient large n, we choose W_0 s.t.

$$\frac{1}{n-W}\sum_{\bar{t}=W}^{n} I(r_{1,\bar{t}+1} < -\text{G-VaR}_{\alpha,\bar{t}}^{W_0}(r_{1,\bar{t}+1})) \sim 0.01.$$

  Note that the above percentage is the violation rate of $r_{1,t}$ under the G-VaR model from time $W+1$ to $n+1$, hereafter denoted as %Viol(n).

NASDAQ Composite Index

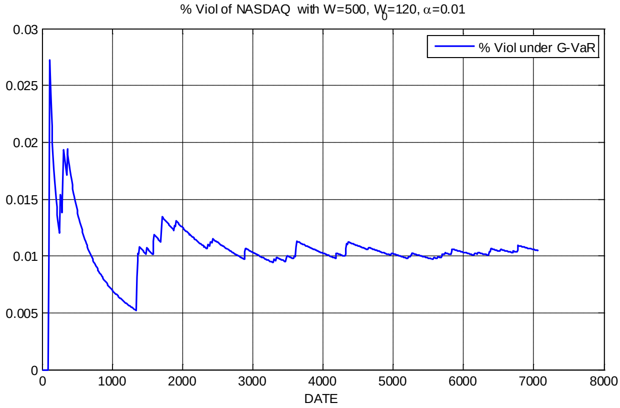

Figure: 2, NASDAQ Composite Index: convergence of the violation rate for W = 500, and $\alpha = 0.01$. The adaptive window sizes are $W_0 = 120$.

NASDAQ Composite Index

- Figure 2 plots the evolution of %Viol(n) as time n varies. Here, $\alpha = 0.01$ is used, but the findings are similar for other values of α. For all window sizes $W = 500$, we find that when $3000 \leq n - W$, the violation rate becomes close to the target $\alpha = 0.01$. All three cases use a well-calibrated value of $W_0 = 120$.

NASDAQ Composite Index

- To assess the predictive performance of the models under consideration, we follow the test of unconditional coverage, or the binomial test (Kuester et al., 2006).

- More precisely, let $\hat{\alpha} = m_1/(m_0 + m_1)$ be the sample violation rate %Viol, where m_1 is the sample number of violations, and the total number of observations is $m_0 + m_1 = T - W$. Using the well-known asymptotic $\chi^2(1)$ distribution, the p-value of the test is

$$\mathrm{LR}_{uc} = P\left(\chi^2(1) > 2m_1 \frac{\hat{\alpha}}{\alpha} + 2m_0 \frac{1-\hat{\alpha}}{1-\alpha}\right).$$

NASDAQ Composite Index

- For the AR(1)-GARCH(1,1)-Normal, AR(1)-GARCH(1,1)-Skewed-t, and AR(1)-GARCH(1,1)-Skewed-t-EVT models, we spend about 4 hour to calculate one thousand VaR.

- For G-VaR, we spend about 10 second to calculate one thousand VaR.

NASDAQ Composite Index

Table 2: NASDAQ Composite Index: Empirical statistics of G-VaR forecast compared with forecasts of three benchmark predictors reported in Kuester et al. (2006) with W=1000. The bottom plot shows the LR_{uc}.

Model		100α	%Viol.	LR_{uc}	$100\overline{\mathrm{VaR}}$
AR(1)-GARCH(1,1)-N:		1	2.23	0.00	2.05
		2.5	3.92	0.00	1.72
		5	6.21	0.00	1.43
AR(1)-GARCH(1,1)-St:		1	1.2	0.12	2.57
		2.5	2.72	0.25	2.01
		5	5.12	0.65	1.59
AR(1)-GARCH(1,1)-St-EVT:		1	0.97	0.82	2.70
		2.5	2.47	0.87	2.07
		5	5.06	0.82	1.61
G-VaR:	W_0= 350	1	0.99	0.93	2.78
	650	2.5	2.51	0.96	2.06
	900	5	5.03	0.90	1.52

NASDAQ Composite Index

- The results in Table 2 show that, for a given $W = 1000, \alpha = 0.01$, 0.025, 0.05, the adaptive W_0 are $350, 650, 900$,
- The %Viol of G-VaR is better than that of the AR(1)-GARCH(1,1) model with Normal, Skewed-t, Skewed-t-EVT innovations. See the plot of the p-value LR_{uc} at the bottom of the table. In addition, the values of $100\overline{\mathrm{VaR}}$ in the four models are very close to one another.

NASDAQ Composite Index

- Kuesteretal.(2006) concluded that the AR-GARCH-Skewed-t and AR-GARCH-Skewed-t-EVT VaR models achieve better performance with lager windows, e.g., $W = 1000$, than with smaller windows, e.g., $W = 500, 250$.
- For G-VaR, however, as suggested by Figure 2, the %Viol statistics are much more stable for $W = 500, 250$, which suggests better performance with smaller windows.

NASDAQ Composite Index

Table 3: NASDAQ Composite Index with W=500, 250

Model			100α	%Viol.	LR_{uc}	$100\overline{\text{VaR}}$
W=500:	W_0=	50	0.3	0.30	0.96	4.71
		70	0.5	0.49	0.95	4.00
		120	1	1.05	0.70	3.11
		270	2.5	2.54	0.81	2.14
		420	5	5.00	0.99	1.60
W=250:	W_0=	35	0.3	0.30	1.00	4.29
		50	0.5	0.52	0.82	3.67
		75	1	1.02	0.84	2.98
		150	2.5	2.49	0.93	2.12
		210	5	5.05	0.84	1.61

Adaptive window size W_0

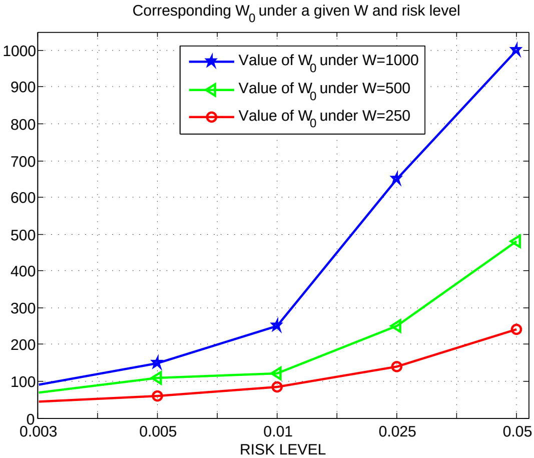

Figure: 3. S&P 500 Index data: variation of adaptive window W_0 for different risk levels α and historical windows W.

Adaptive window size W_0

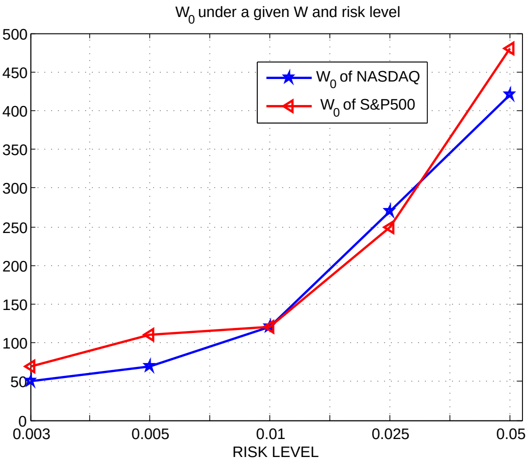

Figure: 4. For a given $W = 500, \alpha = 0.003, 0.005, 0.01, 0.025, 0.05$, W_0 of NASDAQ and S&P500 indexes.

Adaptive window size W_0

- The information carried by these experimental values of W_0 can be pushed further. In Figure 4, we compare the values found for both the NASDAQ Composite Index and S&P500 Index under different risk levels α, with the historical window size fixed at $W = 500$.
- Because under our worst-case scenario, smaller window sizes W_0 correspond to higher volatility (and thus higher risk in the index), we can assume that the S&P500 Index return is riskier than the NASDAQ Composite Index return at risk level $\alpha = 0.025$. Their degree of risk is comparable at risk level $\alpha = 0.01$, and the S&P500 Index is probably less risky at risk levels $\alpha = 0.003, 0.005$, and 0.05.

Discussion

- This paper introduces a new VaR predictor, G-VaR, for financial return series. Our methodology is based on the model-uncertainty principle that the volatility of returns cannot be adequately characterized by a single statistical distribution or model.
- Rather, an infinite-dimensional family of distributions is necessary for full characterization. Considering the worst-case volatility scenario among these numerous potential distributions, and using the our new theory of NLE, we identity G-VaR through a new mathematical object called G-normal distribution.

Discussion

- Extensive empirical analysis using the NASDAQ Composite Index and S&P500 Index shows the G-VaR predictor to outperform many of the existing benchmark predictors of VaR. Its superiority is particularly significant for low risk levels, such as $\alpha = 1\%$ or 0.5%.
- However, a number of unanswered question remain to be investigated in future. In particular, the implementation of G-VaR depends on an adaptive window W_0.

Autoregressive models

The model developed in **P. -Yang-Yao** (2020a) is a static model, in which we use the daily data to estimate the parameters and then calculate the risk for the following period.
We now investigate a dynamic counterpart of the regressive model for forecasting one-period parameters in **P. -Yang** (2020b). Now, we report the details of the paper **P. -Yang** (2020b).

- P.-Yang. Autoregressive models of the time series under volatility uncertainty and application to VaR model, arXiv:2011.09226, 1-29.

Data Z_t obeys the G-normal distribution $N(r_t, [\underline{\sigma}_t^2, \overline{\sigma}_t^2])$. The maximum volatility $\overline{\sigma}_t$, minimum volatility $\underline{\sigma}_t$, and return r_t satisfy

$$\overline{\sigma}_t^2 = \alpha_0 + \alpha_1 \overline{\sigma}_{t-1}^2; \tag{7}$$

$$\underline{\sigma}_t^2 = \beta_0 + \beta_1 \underline{\sigma}_{t-1}^2; \tag{8}$$

$$r_t = \gamma_0 + \gamma_1 r_{t-1}. \tag{9}$$

We denote the 100 times log-returns daily data of S&P500 Index as Z, and one observation sequence $\{X_t\}_{t=1}^n$ is from January 4, 2010 to July 17, 2020, with a total of $n = 2653$ daily data.

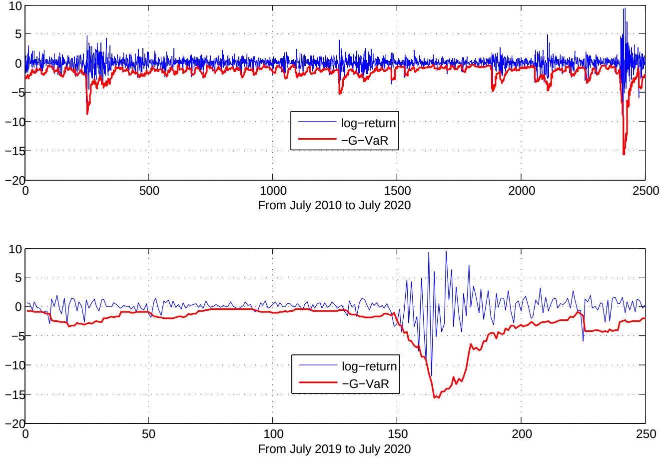

Figure: 5, G-VaR value for log-return of S&P500 Index and the dynamic value of log-returns of S&P500 Index with risk level $\alpha = 0.05$.

In the first picture of Figure 5, the blue line denotes the log-returns of S&P500 Index, and the red line denotes the VaR under $-\text{G-VaR}_{\alpha}(Z_{t+1})$ model, where

$$\text{G-VaR}_{\alpha}(Z_{t+1}) = -\tilde{r}_{t+1} - \tilde{\tilde{\sigma}}_{t+1}\Phi^{-1}\left(\frac{1+\tilde{\kappa}_{t+1}}{2}\alpha\right),$$

and the time $t+1$ is from July 2010 to July 2020. Moreover, as per the second picture of Figure 5, the measure $\text{G-VaR}_{\alpha}(Z_{t+1})$ can capture the local changes of the log-returns of S&P500 Index.

Similarly, we take $K = 6$, $L = 5$, $N = 100$, and conclude the G-VaR forecasting results for log-returns of S&P500 Index in Figure 6.

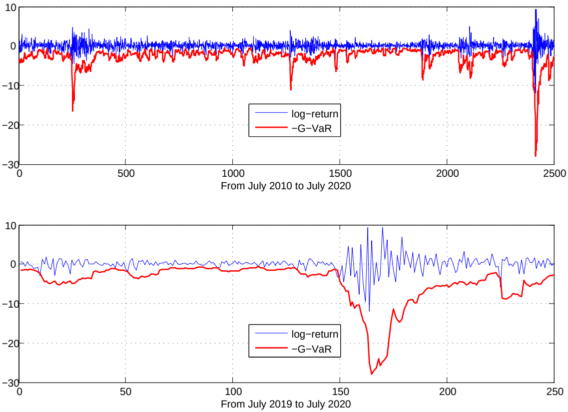

Figure: 6, G-VaR value for log-return of S&P500 Index and the dynamic value of log-returns of S&P500 Index with risk level $\alpha = 0.01$.

Discussion

- To capture the inherent uncertainty of time series, the parameters of maximum volatility, minimum volatility, and the volatility uncertainty index are introduced, and then autoregressive models are developed, indicating the maximum and minimum volatilities to forecast the one-period value.
- In the autoregressive models, we assume the maximum and minimum volatilities have the p-th autoregressive effect. Hence, we can obtain the one-period forecast maximum volatility, minimum volatility and volatility uncertainty index.

Discussion

- As shown in this study, we find that the volatility uncertainty index $\kappa_t = \frac{\underline{\sigma}_t}{\overline{\sigma}_t}$ is a powerful tool to represent the inherent uncertainty of the time series.
- Furthermore, it is interesting to develop a mean uncertainty index, and a related autoregressive model. Future work on the relationship between the volatility uncertainty and the mean uncertainty should be considered and investigated.

THANKS !