Estimating and Backtesting Distortion Risk Measures

Hideatsu Tsukahara

(tsukahar@seijo.ac.jp)

Dept of Economics, Seijo University

November 1, 2013

“Stochastic processes and their statistics in Finance”

in Okinawa

Contents

1. Introduction to Distortion Risk Measures (DRMs)

2. Statistical Estimation

   - Asymptotic results
   - Estimation of asymptotic variance
   - Bias correction

3. Backtesting DRMs

   - Unconditional & Conditional approaches
   - Backtestability & Elicitability

1. Distrotion Risk Measures

A random variable X represents a loss of some financial position

DRM

Any coherent risk measure satisfying law invariance and comonotonic additivity is a **distortion risk measure**:

$$\rho(X) = \rho(F) := \int_{[0,1]} F^{-1}(u)\,\mathrm{d}D(u) = \int_{\mathbb{R}} x\,\mathrm{d}D \circ F(x).$$

where F is the df of X, F^{-1} is the quantile of X, and D is a convex distortion, i.e., a df on $[0, 1]$.

►► a.k.a. spectral risk measure (Acerbi), weighted V@R (Cherny)

Example: *Expected Shortfall (ES)*

The expected loss that is incurred when VaR is exceeded:

$$\mathsf{ES}_\theta(X) := \frac{1}{\theta}\int_{1-\theta}^{1} F^{-1}(u)\,\mathrm{d}u \doteq \mathrm{E}(X \mid X \geq \mathsf{VaR}_\theta(X))$$

Taking distortion of the form

$$D_\theta^{\mathsf{ES}}(u) = \frac{1}{\theta}\big[u - (1-\theta)\big]_+, \quad 0 < \theta < 1$$

yields ES as a distortion risk measure.

►► Typical values for θ are: 0.05, 0.01, ...

Other Examples of DRM:

- *Proportional Hazards*:

$$D_\theta^{\mathsf{PH}}(u) = 1 - (1-u)^\theta,$$

- *Proportional Odds*:

$$D_\theta^{\mathsf{PO}}(u) = \frac{\theta u}{1-(1-\theta)u}$$

- *Gaussian* (Wang transform):

$$D_\theta^{\mathsf{GA}}(u) = \Phi(\Phi^{-1}(u) + \log\theta)$$

★ See Tsukahara (2009) *Mathematical Finance*, vol. 19.

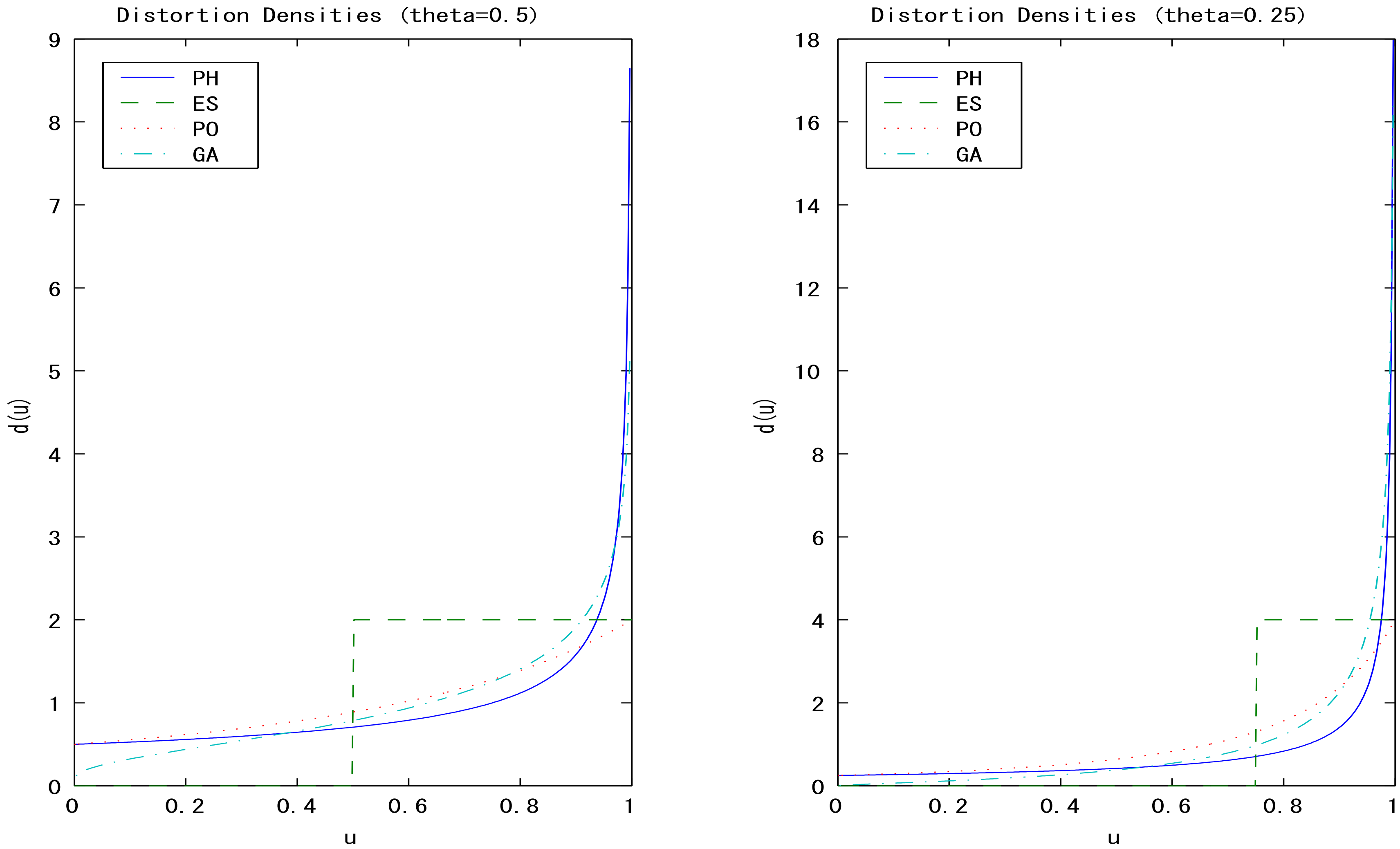

Figure 1: Distortion densities ($\theta = 0.5,\ \ \theta = 0.25$)

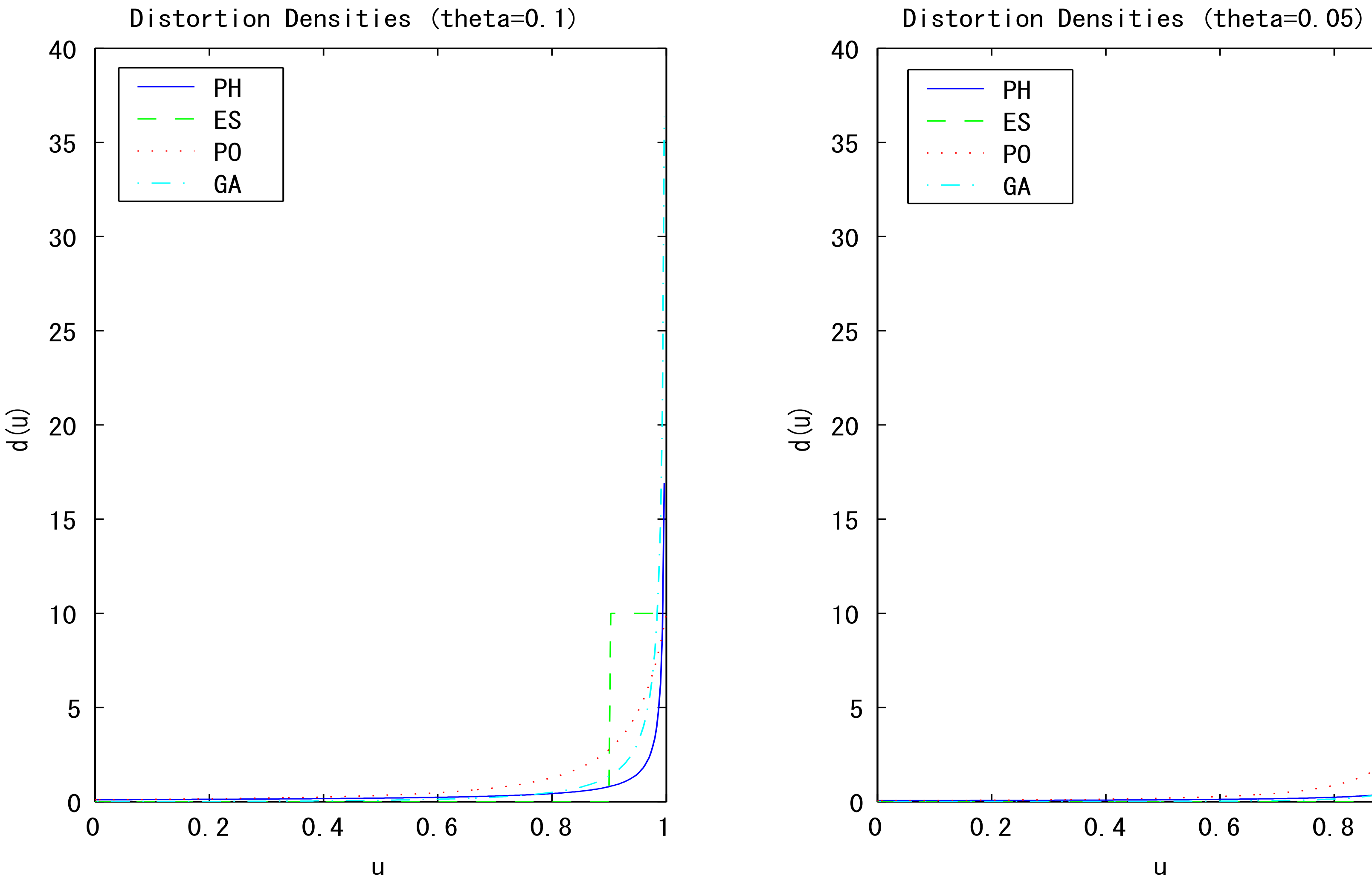

Figure 2: Distortion densities ($\theta = 0.1$, $\theta = 0.05$)

2. Statistical Estimation

$(X_n)_{n\in\mathbb{N}}$: strictly stationary process with $X_n \sim F$

$\mathbb{F}_n$: empirical df based on the sample $X_1, \ldots, X_n$

A natural estimator of $\rho(F)$ is

$$\begin{aligned}\widehat{\rho}_n &= \int_0^1 \mathbb{F}_n^{-1}(u)\, dD(u) \\ &= \sum_{i=1}^{n} c_{ni} X_{n:i}, \qquad c_{ni} := D\left(\frac{i-1}{n}, \frac{i}{n}\right]\end{aligned}$$

This type of statistics is called L-statistics

Strong consistency

Let $d(u) = \frac{\mathrm{d}}{\mathrm{d}u} D(u)$ for a convex distortion D, and $1 \leq p \leq \infty$, $1/p + 1/q = 1$. Suppose

- $(X_n)_{n \in \mathbb{N}}$ is an ergodic stationary sequence
- $d \in L^p(0,1)$ and $F^{-1} \in L^q(0,1)$

Then

$$\widehat{\rho}_n \longrightarrow \rho(F), \quad \text{a.s.}$$

For a proof, see van Zwet (1980, AP)

[All we need is SLLN and Glivenko-Cantelli Theorem].

Assumptions for asymptotic normality:

- $(X_n)_{n\in\mathbb{N}}$ is strongly mixing with rate

$$\alpha(n) = O(n^{-\theta-\eta}) \quad \text{for some } \theta \geq 1+\sqrt{2},\ \eta > 0$$

- For F^{-1}-almost all u, d is continuous at u

- $|d| \leq B$, $\quad B(u) := Mu^{-b_1}(1-u)^{-b_2}$,

- $|F^{-1}| \leq H, \quad H(u) := Mu^{-d_1}(1-u)^{-d_2}$

Assume b_i, d_i & θ satisfy $b_i + d_i + \dfrac{2b_i+1}{2\theta} < \dfrac{1}{2}, \quad i = 1,2$

Set

$$\sigma(u,v) := [u \wedge v - uv] + \sum_{j=1}^{\infty} [C_j(u,v) - uv] + \sum_{j=1}^{\infty} [C_j(v,u) - uv],$$

$$C_j(u,v) := \mathrm{P}(X_1 \le F^{-1}(u), X_{j+1} \le F^{-1}(v))$$

Theorem (Asymptotic Normality)

Under the above assumptions, we have

$$\sqrt{n}(\widehat{\rho}_n - \rho(F)) \xrightarrow{\mathcal{L}} N(0, \sigma^2),$$

where

$$\sigma^2 := \int_0^1 \int_0^1 \sigma(u,v) d(u) d(v) \, \mathrm{d}F^{-1}(u) \mathrm{d}F^{-1}(v) < \infty$$

- GARCH model:

$X_t = \sigma_t Z_t, \quad (Z_t):$ i.i.d.

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

▶▶ If the stationary distribution has a positive density around 0, then GARCH is strongly mixing with exponentially decaying $\alpha(n)$

- Stochastic Volatility model:

$X_t = \sigma_t Z_t, \quad (Z_t):$ i.i.d., $\quad (\sigma_t):$ strictly stationary positive

(Z_t) and (σ_t) are assumed to be independent

▶▶ The mixing rate of (X_t) is the same as that of (σ_t)

Estimation of Asymptotic Variance

Let

$$Y_n := \int [\mathbf{1}\{X_n \leq x\} - F(x)] d(F(x)) \, \mathrm{d}x, \quad n \in \mathbb{Z}.$$

Then Y_n is also a strictly stationary and strongly mixing sequence with the same mixing coefficient as X_n. Furthermore

$$\mathrm{E}(Y_n) = 0, \qquad \sigma^2 = \sum_{h=-\infty}^{\infty} \gamma(h) < \infty,$$

where $\gamma(h) := \mathrm{E}(Y_n Y_{n+h})$.

Let f be the spectral density of (Y_n). Then

$$\sum_{h=-\infty}^{\infty} \gamma(h) = 2\pi f(0)$$

$\Longrightarrow$ Use a consistent estimator of $f(0)$ (JHB approach)

The lag window estimator is defined by

$$\widehat{f}_n(\lambda) = \frac{1}{2\pi} \sum_{|k|<K_n} w(k/K_n)\widehat{\gamma}_n(k)\cos k\lambda$$

where w is a "lag window", and $\widehat{\gamma}_n(k) := \dfrac{1}{n}\displaystyle\sum_{i=1}^{n-k} Y_i Y_{i+k}$

►► F in the expression of Y_n is unknown, so we replace it with the empirical df. That is, we use

$$Y_{i,n} := \int [\mathbf{1}\{X_i \leq x\} - \mathbb{F}_n(x)] d(\mathbb{F}_n(x)) \, \mathrm{d}x, \quad i = 1, \ldots, n$$

Let

$$\widetilde{\gamma}_n(k) := \frac{1}{n} \sum_{i=1}^{n-k} Y_{i,n} Y_{i+k,n} \quad \text{and} \quad \widetilde{f}_n(0) := \frac{1}{2\pi} \sum_{|k|<K_n} w(k/K_n) \widetilde{\gamma}_n(k)$$

Then $2\pi \widetilde{f}_n(0)$ should give a consistent estimator of the asymptotic variance σ^2

Theorem

In addition to the conditions assumed in the above theorem, suppose that J is Lipschitz, w is a bounded even function which is continuous in $[-1,1]$ with $w(0)=1$ and equals 0 outside $[-1,1]$. Also assume $\mathrm{E}|Y_n|^4 < \infty$ and the fourth-order cumulants

$$\begin{aligned}\kappa(h,i,j) := &\,\mathrm{E}(Y_1 Y_{1+h} Y_{1+i} Y_{1+j}) - \gamma(h)\gamma(i-j) \\ &- \gamma(i)\gamma(h-j) - \gamma(j)\gamma(h-i)\end{aligned}$$

are summable: $\sum_{h,i,j=-\infty}^{\infty} |\kappa(h,i,j)| < \infty$.

Let K_n be a sequence of integers such that $K_n \to \infty$ and $K_n/\sqrt{n} \to 0$ as $n \to \infty$. Then we have

$$2\pi \widetilde{f}_n(0) \xrightarrow{L_1} \sigma^2, \quad n \to \infty$$

Bias of L-statistics

By Fubini, for any df F and any distortion D,

$$\int_{[0,1]} F^{-1}(u)\,\mathrm{d}D(u) = -\int_{\infty}^{0} D(F(x))\,\mathrm{d}x + \int_0^{\infty} [1 - D(F(x))]\,\mathrm{d}x$$

By Fubini and Jensen, for convex D,

$$\begin{aligned}
&\mathrm{E}\left[\int_{[0,1]} \mathbb{F}_n^{-1}(u)\,\mathrm{d}D(u)\right] \\
&= \int_{\infty}^{0} \mathrm{E}(-D(\mathbb{F}_n(x)))\,\mathrm{d}x + \int_0^{\infty} \mathrm{E}[1 - D(\mathbb{F}_n(x))]\,\mathrm{d}x \\
&\leq \int_{\infty}^{0} -D(\mathrm{E}(\mathbb{F}_n(x)))\,\mathrm{d}x + \int_0^{\infty} [1 - D(\mathrm{E}(\mathbb{F}_n(x)))]\,\mathrm{d}x \\
&= \int_{[0,1]} F^{-1}(u)\,\mathrm{d}D(u)
\end{aligned}$$

Therefore

$$\mathrm{E}(\widehat{\rho}_n) - \rho(F) \leq 0$$

$\Longrightarrow$ $\widehat{\rho}_n$ has a negative bias

Need bias correction methods. For the i.i.d. case,

- Xiang (1995): Modify the form of L-statistics
- Kim (2010): Bootstrap-based method

►► The bootstrap methodology is still available in the dependent case (see Lahiri (2003), Example 4.8).

Moving Block Bootstrap (MBB)

- Data: $X_1, \ldots, X_n$
- Block size: ℓ, $\quad \#$ of blocks: $N := n - \ell + 1$
- Blocks: $\mathcal{B}_i = (X_i, \ldots, X_{i+\ell-1})$, $\; i = 1, \ldots, N$

Resample $k = [n/\ell]$ blocks from $\{\mathcal{B}_1, \ldots, \mathcal{B}_N\}$ with replacement to get $\mathcal{B}_1^*, \ldots, \mathcal{B}_k^*$

Write $\mathcal{B}_i^* = (X_{(i-1)\ell+1}^*, \ldots, X_{i\ell}^*)$

$\Longrightarrow X_1^*, \ldots, X_{k\ell}^*$: MBB sample

MBB version of $\widehat{\rho}_n$ is

$$\widehat{\rho}_n^* = \frac{1}{n}\sum_{i=1}^{n} c_{ni} X_{n:i}^*, \qquad c_{ni} := D\left(\frac{i-1}{n}, \frac{i}{n}\right]$$

Validity of MBB follows from an argument specific to our case.

▶▶ The approach based on Hadamard differentiability of L-functional

$$T(F) := \int_0^1 h(F^{-1}(u))J(u)\,\mathrm{d}u$$

is not convenient. See Boos (1979, AS), Lahiri (2003), Section 12.3.5.

Simulation example: inverse-gamma SV model

$$X_t = \sigma_t Z_t$$

Z_t i.i.d. N(0,1) and $V_t = 1/\sigma_t^2$ satisfies

$$V_t = \rho V_{t-1} + \varepsilon_t,$$

where $V_t \sim$ Gamma(a, b) for each t, (ε_t) i.i.d. rv's, and $0 \leq \rho < 1$

$\Rightarrow X_t$ has scaled t-distribution with $\nu = 2a$, $\sigma^2 = b/a$

►► Lawrance (1982): the distribution of ε_t is compound Poisson

►► Can be shown that (X_t) is geometrically ergodic

Simulation results for estimating VaR, ES & PO risk measures with inverse-gamma SV observations ($n = 500$, # of replications $= 1000$)

$X_t = \sigma_t Z_t$, where $V_t = 1/\sigma_t^2$ follows AR(1)
with gamma(2,16000) marginal & $\rho = 0.5$, Z_t i.i.d. N(0,1)

		VaR		ES		PO	
	θ	bias	RMSE	bias	RMSE	bias	RMSE
	0.1	0.0692	10.9303	−2.2629	22.1361	−1.7739	17.5522
SV	0.05	2.5666	17.6755	−1.2168	37.2719	−2.0200	28.5053
	0.01	14.9577	61.2290	−11.9600	103.9269	−15.7888	73.7147
	0.1	0.7976	10.5893	−1.2914	19.5756	−1.3574	15.3271
i.i.d.	0.05	0.7974	16.1815	−2.6346	31.3166	−2.8342	23.9933
	0.01	10.6838	53.2567	−12.9355	95.9070	−15.8086	69.5425

Simulation results for estimating variance and bias of PO risk measure

($n = 500$, $K_n = 5$, Parzen kernel $w(x) = 1 - x^2$, block size$= 5$, $\#$ of bootstrap replicates $= 800$, $\#$ of replications $= 10000$)

	ρ	θ	MC bias	MC s.e.	$\widehat{\text{A-s.e.}}$	BS bias	BS s.e.
IG-SV		0.1	−0.8328	15.4456	14.0956	−0.8151	13.9829
$\alpha = 2$	0.1	0.05	−2.0580	24.6961	20.9719	−1.8170	20.6863
$\beta = 16000$		0.01	−13.3608	68.9197	46.6943	−10.2030	46.0788
IG-SV		0.1	−0.3345	10.7979	10.4231	−0.6812	10.3933
$\alpha = 4$	0.1	0.05	−1.3663	15.1946	14.0623	−1.3511	13.9725
$\beta = 48000$		0.01	−6.8659	34.4725	26.4183	−6.0749	26.4446
IG-SV		0.1	−0.5432	9.0853	8.8370	−0.6048	8.8281
$\alpha = 10$	0.1	0.05	−1.1786	11.7923	11.2289	−1.1263	11.2003
$\beta = 144000$		0.01	−5.8673	22.9686	18.7767	−4.4474	18.9614

	ρ	θ	MC bias	MC s.e.	$\widehat{\text{A-s.e.}}$	BS bias	BS s.e.
IG-SV		0.1	-1.0054	17.5469	15.0711	-0.8793	14.6925
$\alpha = 2$	0.5	0.05	-2.2714	27.1465	22.0852	-1.9450	21.4374
$\beta = 16000$		0.01	-13.9208	74.8887	47.7943	-10.6541	46.8379
IG-SV		0.1	-0.5791	11.4856	10.7162	-0.6957	10.5906
$\alpha = 4$	0.5	0.05	-1.3472	15.7116	14.4718	-1.3994	14.2658
$\beta = 48000$		0.01	-7.4680	35.1014	26.7575	-6.1939	26.7115
IG-SV		0.1	-0.8213	9.2632	8.9299	-0.6062	8.8957
$\alpha = 10$	0.5	0.05	-1.0663	11.9443	11.3608	-1.1368	11.2996
$\beta = 144000$		0.01	-5.7987	23.1130	18.8147	-4.4769	18.9896

	ρ	θ	MC bias	MC s.e.	$\widehat{\text{A-s.e.}}$	BS bias	BS s.e.
IG-SV		0.1	-2.0408	28.2224	15.5015	-0.9609	14.7212
$\alpha = 2$	0.9	0.05	-4.8204	42.1005	22.1388	-2.0483	20.9685
$\beta = 16000$		0.01	-23.5844	106.4374	43.6402	-10.1556	42.4681
IG-SV		0.1	-1.1973	14.9586	11.1112	-0.7274	10.8092
$\alpha = 4$	0.9	0.05	-2.2346	20.8199	14.8937	-1.4366	14.4566
$\beta = 48000$		0.01	-10.2968	42.5085	26.3137	-6.1439	26.0855
IG-SV		0.1	-0.5956	10.3666	9.1248	-0.6262	9.0293
$\alpha = 10$	0.9	0.05	-1.4212	13.6534	11.5934	-1.1609	11.4494
$\beta = 144000$		0.01	-6.3827	25.2688	18.8986	-4.4824	19.0079

	ρ	θ	MC bias	MC s.e.	$\widehat{\text{A-s.e.}}$	BS bias	BS s.e.
		0.1	−0.5734	8.2886	8.0638	−0.5619	8.0667
$N(0, 126.5^2)$	iid	0.05	−1.1557	10.1327	9.8175	−1.0116	9.8117
		0.01	−4.4730	18.1714	14.9659	−3.6136	15.2192
		0.1	−0.9038	15.3536	13.9544	−0.8121	13.8815
$t_4(0, 126.5^2)$	iid	0.05	−1.8468	24.3247	20.8781	−1.7928	20.6468
		0.01	−12.5608	73.3170	46.9313	−10.2243	46.3147
		0.1	−0.5538	10.7575	10.3154	−0.6687	10.2909
$t_8(0, 126.5^2)$	iid	0.05	−1.4518	14.9271	13.9883	−1.3379	13.9033
		0.01	−6.8385	34.8496	26.4076	−6.8385	26.4531
		0.1	−0.5470	9.0123	8.8209	−0.5985	8.8127
$t_{20}(0, 126.5^2)$	iid	0.05	−1.1266	11.6915	11.2178	−1.1176	11.1965
		0.01	−5.5631	22.9298	18.7808	−4.4588	18.9697

3. Backtesting

Purpose of Backtesting:

1. Monitor the performance of the model and estimation methods for risk measurement

2. Compare relative performance of the models and methods

Idea

ex ante risk measure forecasts from the model

vs.

ex post realized portfolio loss

Setup

Entire observations: $X_1, \ldots, X_T$

Estimation window size $= n$, $\quad m := T - n$

	data	estimand	realized loss
1.	$X_1, \ldots, X_n$	$\rho(X_{n+1})$	X_{n+1}
2.	$X_2, \ldots, X_{n+1}$	$\rho(X_{n+2})$	X_{n+2}
$\vdots$	$\vdots$	$\vdots$	$\vdots$
m.	$X_{T-n}, \ldots, X_{T-1}$	$\rho(X_T)$	X_T

Two approaches to risk measurement

Assume that the loss process $(X_t)_{t\in\mathbb{Z}}$ is a stationary time series with stationary df F. At time t, we have two options:

I. Unconditional Approach

Look at the risk measure associated with $F(x) = \mathrm{P}(X_{t+1} \leq x)$

(For a large time horizon; credit risk and insurance)

II. Conditional Approach

For a given filtration $\mathscr{F}_t$, look at the risk measure associated with the conditional df $F_t(x) := \mathrm{P}(X_{t+1} \leq x \,|\, \mathscr{F}_t)$,

(For a short time horizon; market risk)

In the case of VaR

- Unconditional VaR, denoted by VaR_α, satisfies

$$\mathrm{E}\big(\mathbf{1}\{X_{t+1} \geq \mathrm{VaR}_\alpha\}\big) = \alpha$$

But $\mathbf{1}\{X_{t+1} \geq \mathrm{VaR}_\alpha\}$'s might not be independent

- Conditional VaR, denoted by VaR_α^t, satisfies

$$\mathrm{E}\big(\mathbf{1}\{X_{t+1} \geq \mathrm{VaR}_\alpha^t\} \,|\, \mathscr{F}_t\big) = \alpha$$

By Lemma 4.29 of MFE, if (Y_t) is a sequence of Bernoulli rv's adapted to $(\mathscr{F}_t)$ and if $\mathrm{E}(Y_{t+1} \,|\, \mathscr{F}_t) = p > 0$, then (Y_t) must be i.i.d.

Therefore $\mathbf{1}\{X_{t+1} \geq \mathrm{VaR}_\alpha^t\}$, $t = n, \ldots, T-1$ are i.i.d. Bernoulli rv's.

$$\Downarrow$$

This gives the grounds for backtesting using $\mathbf{1}\{X_{t+1} \geq \widehat{\mathrm{VaR}}_\alpha^t\}$, where $\widehat{\mathrm{VaR}}_\alpha^t$ is an estimate of the VaR associated with the conditional df $F_t(x) := \mathrm{P}(X_{t+1} \leq x \,|\, \mathscr{F}_t)$. Namely,

(i) Test $\displaystyle\sum_{t=n}^{T-1} \mathbf{1}\{X_{t+1} \geq \widehat{\mathrm{VaR}}_\alpha^t\} \sim \mathrm{Bin}(m, \alpha)$

(ii) Test independence of $\mathbf{1}\{X_{t+1} \geq \widehat{\mathrm{VaR}}_\alpha^t\}$, $t = n, \ldots, T-1$
(e.g., runs test)

Backtesting DRMs

Note that, with $d(u) = \frac{\mathrm{d}}{\mathrm{d}u} D(u)$ and $X \sim F$,

$$\rho(X) = \int_{-\infty}^{\infty} x \,\mathrm{d}D \circ F(x) = \int_{-\infty}^{\infty} x d(F(x)) \,\mathrm{d}F(x)$$

$$= \mathrm{E}[X d(F(X))]$$

Thus $X d(F(X)) - \rho(X)$ has mean 0 unconditionally.

►► In the conditional case, $\mathrm{E}[X_{t+1} d(F_t(X_{t+1})) \mid \mathscr{F}_t] = \rho_t(X_{t+1})$, but this does not help much.

I.I.D. case (rough-and-ready)

If $X_1, \ldots, X_T$ are i.i.d. with df F, then we can base the backtesting of our method/model on

$$X_{n+1} d(\widehat{\mathbb{F}}_{1:n}(X_{n+1})) - \widehat{\rho}_{(1:n)},$$

$$\vdots$$

$$X_T d(\widehat{\mathbb{F}}_{T-n:T-1}(X_T)) - \widehat{\rho}_{(T-n:T-1)}$$

where $\widehat{\mathbb{F}}_{k:l}$ and $\widehat{\rho}_{(k:l)}$ are estimates based on the sample $X_k, \ldots, X_l$

►► If we have dependent data or we use the conditional approach, it is necessary to introduce more explicit time series models.

Conditional Approach

Write $\rho_t(X_{t+1})$ for a distortion risk measure with a distortion D for the conditional df $F_t(x) := \mathrm{P}(X_{t+1} \le x \mid \mathscr{F}_t)$, $\mathscr{F}_t := \sigma(X_s \colon s \le t)$:

$$\rho_t(X_{t+1}) := \int_{[0,1]} F_t^{-1}(u) \, \mathrm{d}D(u)$$

Assumption

Suppose that for $\mathscr{F}_{t-1}$-measurable μ_t and σ_t,

$$X_t = \mu_t + \sigma_t Z_t,$$

where (Z_t) is i.i.d. with finite 2nd moment.

Example: ARMA(p_1, q_1) with GARCH(p_2, q_2) errors

Let (Z_t) be i.i.d. with finite 2nd moment.

$$\begin{aligned}
X_t &= \mu_t + \sigma_t Z_t, \\
\mu_t &= \mu + \sum_{i=1}^{p_1} \phi_i (X_{t-i} - \mu) + \sum_{j=1}^{q_1} \theta_j (X_{t-j} - \mu_{t-j}), \\
\sigma_t^2 &= \alpha_0 + \sum_{i=1}^{p_2} \alpha_i (X_{t-i} - \mu_{t-i})^2 + \sum_{j=1}^{q_2} \beta_j \sigma_{t-j}^2,
\end{aligned}$$

where $\alpha_0 > 0$, $\alpha_i \geq 0$, $i = 1, \ldots, p_2$, $\beta_j \geq 0$, $j = 1, \ldots, q_2$.

Usually, it is assumed that (X_t) is covariance stationary, and $\sum_{i=1}^{p_2} \alpha_i + \sum_{j=1}^{q_2} \beta_j < 1$.

By (conditional) translation equivariance and positive homogeneity,

$$\rho_t(X_{t+1}) = \mu_{t+1} + \sigma_{t+1}\rho(Z)$$

where Z is a generic rv with the same df G as Z_t's.

(i) If G is a known df, $\rho(Z)$ is a known number.
We need to estimate μ_{t+1} and σ_{t+1} based on $X_{t-n+1}, \ldots, X_t$ using some specific model and method (e.g., ARMA with GARCH errors using QML). Then the risk measure estimate is given by

$$\widehat{\rho}_t(X_{t+1}) := \widehat{\mu}_{t+1} + \widehat{\sigma}_{t+1}\rho(Z)$$

Observe that

$$\rho(Z) = \mathrm{E}\left[Z_{t+1} d(G(Z_{t+1}))\right]$$

$$\Downarrow$$

$$\mathrm{E}\left[(Z_{t+1} - \rho(Z)) d(G(Z_{t+1}))\right] = 0$$

Defining

$$R_{t+1} := Z_{t+1} - \rho(Z) = \frac{X_{t+1} - \rho_t(X_{t+1})}{\sigma_{t+1}}$$

one sees that $(R_t d(G(Z_t)))_{t \in \mathbb{Z}}$ is i.i.d.

This suggests that in practice, we may perform backtesting by examining mean-zero behavior of $\widehat{R}_{t+1} d(G(\widehat{Z}_{t+1}))$, $t = n, \ldots, T-1$, where

$$\widehat{R}_{t+1} := \frac{X_{t+1} - \widehat{\rho}_t(X_{t+1})}{\widehat{\sigma}_{t+1}}$$

and

$$\widehat{Z}_{t+1} = \frac{X_{t+1} - \widehat{\mu}_{t+1}}{\widehat{\sigma}_{t+1}} = \widehat{R}_{t+1} + \rho(Z)$$

►► Bootstrap test can be used

(ii) When G is unknown, we need to estimate G in addition to μ_{t+1} and σ_{t+1}.

In ARMA with GARCH errors model, we could use the empirical df based on the residuals $\widetilde{Z}_s$'s: for $s = t-n+1, \dots, t$,

$$\widetilde{Z}_s = \widetilde{\varepsilon}_s / \widetilde{\sigma}_s, \quad \widetilde{\varepsilon}_s : \text{residual from ARMA part}$$

and

$$\widetilde{\sigma}_s^2 = \widehat{\alpha}_0 + \sum_{i=1}^{p_2} \widehat{\alpha}_i \widetilde{\varepsilon}_{s-i}^2 + \sum_{j=1}^{q_2} \widehat{\beta}_j \widehat{\sigma}_{s-j}^2,$$

Then

$$\widetilde{G}_t(z) = \frac{1}{n} \sum_{s=t-n+1}^{t} \mathbf{1}\{\widetilde{Z}_s \le z\},$$

Simulation study

Simulate GARCH(1,1) process:

$$Y_t = \sigma_t Z_t, \quad Z_t \sim N(0,1) \text{ i.i.d.}$$

$$\sigma_t^2 = 0.01 + 0.9\sigma_{t-1}^2 + 0.08 Y_{t-1}^2$$

Set $T = 1000$, $n = 500$ and $\theta = 0.05$

For $t = n+1, \ldots, T$, plot

(i) $X_t d(\widehat{\mathbb{F}}_{t-n:t-1}(X_t)) - \widehat{\rho}_{(t-n:t-1)}$ (historical, unconditional)

(ii) $\widehat{R}_t d(G(\widehat{Z}_t))$ (normal-GARCH based, conditional)

(i) mean $= -0.0286$, std $= 2.073$

(ii) mean $= -0.0185$, std $= 1.019$

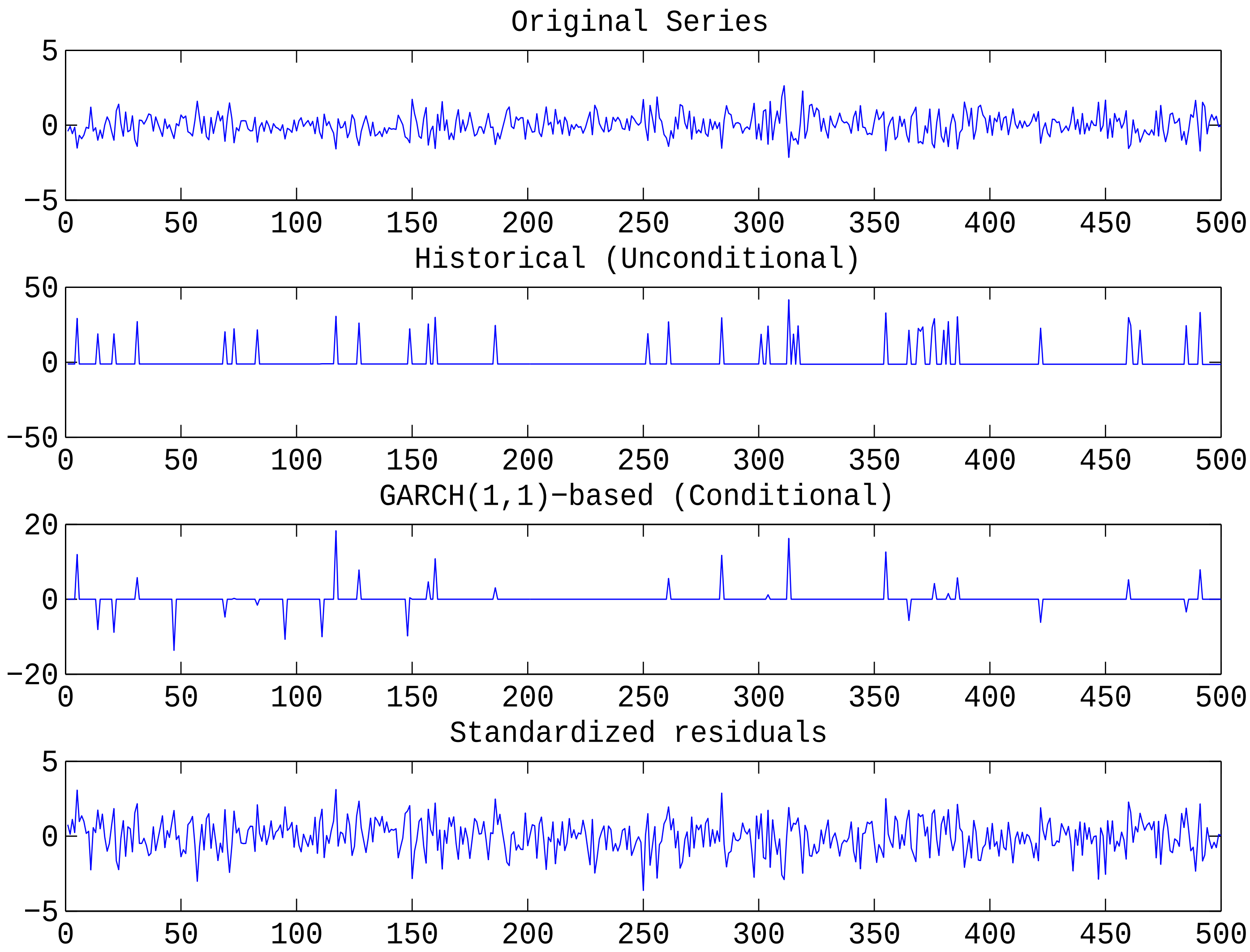

Figure 3: Backtesting results for expected shortfall ($\theta = 0.05$)

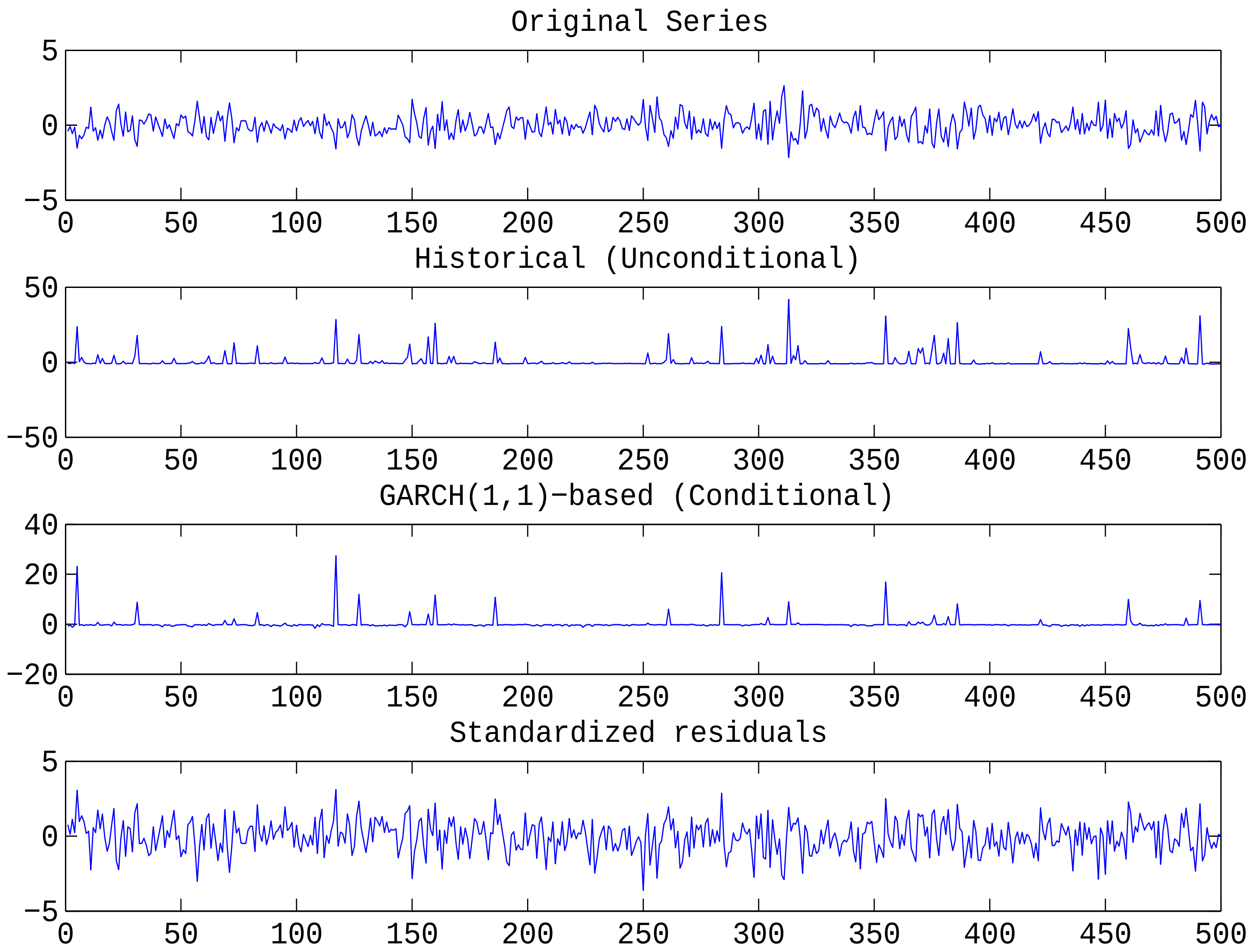

Figure 4: Backtesting results for proportional odds distortion ($\theta = 0.05$)

Issue: Backtestability

"*It is more difficult to backtest a procedure for calculating expected shortfall than it is to backtest a procedure for calculating VaR*" (Yamai & Yoshiba, Hull, Daníelsson, among others)

1. Because the existing tests for ES are based on
   - parametric assumptions for the null distribution
   - asymptotic approximation for the null distribution
2. Because testing an expectation is harder than testing a single quantile.

Elicitability

"*Expected shortfall (and spectral risk measures) cannot be backtested because it fails to satisfy* **elicitability** *condition*" (Paul Embrechts, Mar 2013, *Risk Magazine*)

Def (Osband 1985; Gneiting 2011, JASA)

A statistical functional $T(F)$ is called **elicitable** r.t. $\mathscr{F}$ if $T(F)$ is a unique minimizer of $t \mapsto E^F[S(t, Y)]$ for some scoring function S, $\forall F \in \mathscr{F}$.

Examples

- $\mathrm{VaR}_\theta(F) = F^{-1}(1-\theta)$ is the unique minimizer for

$$S(t,y) = [\mathbf{1}\{t \le y\} - \theta](y-t)$$
$$= \begin{cases} \theta|y-t| & \text{if } t > y \\ (1-\theta)|y-t| & \text{if } t \le y \end{cases}$$

$\mathscr{F} = \{F\colon$ absolutely continuous, $\int |y| \, \mathrm{d}F(y) < \infty\}$.

- Mean functional $T(F) = \int y \, \mathrm{d}F(y)$ is the unique minimizer for

$$S(t,y) = (y-t)^2$$

$\mathscr{F} = \{F\colon \int y^2 \, \mathrm{d}F(y) < \infty\}$.

It is useful when one wants to compare and rank several estimation procedures: With forecasts x_i and realizations y_i, use

$$\frac{1}{n}\sum_{i=1}^{n} S(x_i, y_i)$$

as a performance evaluation criterion.

►► But there seems to be no clear connection with backtestability

e.g., mean cannot be backtested nonparametrically based on the sum of squared errors without invoking asymptotic approximation or assuming parametric distribution.

Concluding Remarks

- Estimation of DRMs is possible with time series data, but for some DRMs, we do not get nice asymptotic properties.

- Backtesting procedure can be performed with DRMs. May need more rigorous/effective procedures.

- Euler capital allocation based on DRMs are easy to compute and widely applicable (with importance sampling)

- Most of the estimation part is published in *Journal of Financial Econometrics* (2013, online)