Robust Measurement of (Heavy-Tailed) Risks: Theory and Implementation

Judith C. Schneider, Finance Center Münster, U Münster
10.9.2014

joint work with Nikolaus Schweizer, U Duisburg-Essen

Model Risk

- Wherever people base their decisions on mathematical models, there is model risk
- Risk that the assumptions underlying the model are not correct...
- In finance: Increased interest post-2008

Model Risk

- Naive "worst cases" tend to be infinite or hard to pin down
- Divergence ball approach to model risk
- Assume that the true data generating process lies within a ball around the model you work with (nominal model)
- Take worst-case expected values among all models within such a ball
- Our goal: apply this to financial applications

Our questions

- How does the choice of the divergence measure influence the potential worst case distributions?
- How can we quantify the amount of model risk in an economically and statistically meaningful way?
- How can we actually compute these worst cases in a numerically reliable way?

A huge literature

How the worst-case can be calculated is answered

- Robustness literature in Operations Research (Whittle, Ben-Tal,...)
- Robustness literature in Macroeconomics (Hansen, Sargent,...)
- Ambiguity (Gilboa/Schmeidler, Maccheroni/Marinacci/Rustichini)
- Finance: Friedman (2002), Glasserman/Xu (Quant. Fin. 2013), Ben-Tal et al (Man. Sci. 2013), Breuer/Csiszar (Math. Fin. 2013)

Our main results

- We identify for which reference models which f-divergence (α-divergence, KL divergence) implies meaningful potential worst case distributions
- We present a method to bound the amount of model risk built on model confidence sets (Hansen/Lunde/Nason Ecta 2011)
- Lastly, we show that simple numerical schemes like important sampling fail in natural applications and present solutions based on MCMC and SMC

The Setting

- Y some random quantity of interest (e.g. hedging error, insurance claims...)
- ν probability distribution for Y, $Y \sim \nu$, "nominal model" (e.g. Y log-normal)
- L loss/disutility function
- Nominal risk:

$$\Lambda_{\nu}(\mathrm{Y}) = E_{\nu}[\mathrm{L}(\mathrm{Y})] < \infty.$$

Model Risk

- Fix a functional $D_\nu(\cdot)$ which measures the distance of alternative models from ν and some radius κ.
- Then, the worst-case problem

$$\Lambda^{\kappa}{}_{\nu}(Y) = \sup_{\eta : D_\nu(\eta) \le \kappa} E_\eta[L(Y)] = \sup_{\eta : D_\nu(\eta) \le \kappa} E_\nu\left[\frac{\mathrm{d}\eta}{\mathrm{d}\nu}(Y)\, L(Y)\right]$$

is surprisingly tractable.

Model Risk

- Fix a functional $D_\nu(\cdot)$ which measures the distance of alternative models from ν and some radius κ.
- Then, the worst-case problem

$$\Lambda^{\kappa}{}_{\nu}(Y) = \sup_{\eta: D_\nu(\eta) \le \kappa} E_\eta[L(Y)] = \sup_{\eta: D_\nu(\eta) \le \kappa} E_\nu \left[\frac{\mathrm{d}\eta}{\mathrm{d}\nu}(Y)\, L(Y) \right]$$

is surprisingly tractable.

- *Our plan:* Apply this theory to compute worst-case values of (financial) quantities ("actual numbers"), e.g.,

$$E_\nu[L(Y)] \approx 5 \quad \text{and} \quad E_{\eta^{wc}}[L(Y)] \approx 7$$

Example: KL-divergence/ α-divergence

Let D_ν^{KL} be Kullback-Leibler divergence (relative entropy), i.e.

$$D_\nu^{KL}(\eta) = E_\nu \left[\frac{\mathrm{d}\eta}{\mathrm{d}\nu}(Y) \log \left(\frac{\mathrm{d}\eta}{\mathrm{d}\nu}(Y) \right) \right]$$

Let D_ν^{α} be the α-divergence, i.e.

$$D_\nu^{\alpha}(\eta) = \frac{\left(E_\nu \left[\left(\frac{\mathrm{d}\eta}{\mathrm{d}\nu}(Y) \right)^{\alpha} \right] - 1 \right)}{\alpha(\alpha - 1)}.$$

For both, worst cases can be calculated theoretically in a tractable way. Good reference: Breuer/Csiszar (Mathematical Finance 2013)

Choosing the distance measure D_ν

- Conditions to ensure that the worst case exists

KL-divergence

$$E_\nu[\exp(\varepsilon L(Y))] < \infty$$

α-divergence

$$E_\nu[(L(Y)^{\frac{\alpha}{\alpha-1}}] < \infty$$

- But what does this imply for the potential worst case distribution?

Balls around the lognormal distribution

- Let ν be a one-dimensional lognormal distribution with parameter σ (or have the same tail behavior)
- For all $\gamma > 1$, every KL-divergence ball around ν contains distributions with density $f(y) \sim y^{-\gamma}$ for large y

Balls around the lognormal distribution

- Let ν be a one-dimensional lognormal distribution with parameter σ (or have the same tail behavior)
- For all $\gamma > 1$, every KL-divergence ball around ν contains distributions with density $f(y) \sim y^{-\gamma}$ for large y ($\gamma \leq 2 \Rightarrow$ non-existence of expected values)

Balls around the lognormal distribution

- Let ν be a one-dimensional lognormal distribution with parameter σ (or have the same tail behavior)
- For all $\gamma > 1$, every KL-divergence ball around ν contains distributions with density $f(y) \sim y^{-\gamma}$ for large y ($\gamma \leq 2 \Rightarrow$ non-existence of expected values)
- No α-divergence ball around ν contains any distributions with polynomial tail behavior $f(y) \sim y^{-\gamma}$

Balls around the lognormal distribution

- Let ν be a one-dimensional lognormal distribution with parameter σ (or have the same tail behavior)
- For all $\gamma > 1$, every KL-divergence ball around ν contains distributions with density $f(y) \sim y^{-\gamma}$ for large y ($\gamma \leq 2 \Rightarrow$ non-existence of expected values)
- No α-divergence ball around ν contains any distributions with polynomial tail behavior $f(y) \sim y^{-\gamma}$
- α-divergence balls around ν contain some further lognormal distributions, but only those with not too large volatility $\hat{\sigma}$

Balls around the lognormal distribution

- Let ν be a one-dimensional lognormal distribution with parameter σ (or have the same tail behavior)
- For all $\gamma > 1$, every KL-divergence ball around ν contains distributions with density $f(y) \sim y^{-\gamma}$ for large y ($\gamma \leq 2 \Rightarrow$ non-existence of expected values)
- No α-divergence ball around ν contains any distributions with polynomial tail behavior $f(y) \sim y^{-\gamma}$
- α-divergence balls around ν contain some further lognormal distributions, but only those with not too large volatility $\hat{\sigma}$
- Example: For $\alpha = 4$ and $\sigma = 0.2$, lognormal distributions with volatility $\hat{\sigma} > 0.231$ are considered infinitely different from ν

Wrap-Up

- For log-normal nominal models, working with α-divergence essentially means excluding tail behavior from the worst-case analysis
- "Garch models are infinitely different from Black-Scholes..."
- What about the true data-generating process? What about tail-behavior during financial crises?

Quantifying model risk

- The choice of the radius κ is essential – and has received surprisingly little attention...
- Hansen/Sargent approach:
  - The worst case should only be taken over models which are statistically indistinguishable from the nominal model
  - If we had enough data to identify superior models, we would use them
  - Thus, κ can be determined from the speed at which data comes in

Quantifying model risk

- Hansen/Sargent have in mind macroeconomic problems ...
- In finance, model choice is dictated as much by (numerical) tractability as by realism
- The frequency of incoming data might be extremely high - and is a poor proxy for model risk

Quantifying model risk

- Ben-Tal et al. propose to estimate κ from independent realizations of Y that have been observed
- Their estimator only works for finite state space
- If no independent observations are available, we propose to

Quantifying model risk

- Ben-Tal et al. propose to estimate κ from independent realizations of Y that have been observed
- Their estimator only works for finite state space
- If no independent observations are available, we propose to
    1. gather the model(s) which mimic(s) the data best in a statistically reliable way
    2. estimate the divergence between these alternatives and the nominal model in a consistent way
    3. determine the radius for the worst case analysis as the maximum of these divergences

Quantifying model risk

1. Model confidence sets: Hansen et al. (2011)
   - $M_0 = \{m_{0,1}, ..., m_{0,k}\}$ finite set of starting models
   - $M^* = \{i \in M_0 : E[d_{ij}] \leq 0 \forall j \in M_0\}$ with $d_{ij} = l(e_j) - l(e_i)$
   - sequential testing and elimination of models from the set M_0 to reach confidence set $\hat{M}$ for M^*

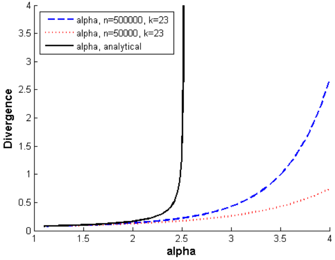
Quantifying model risk

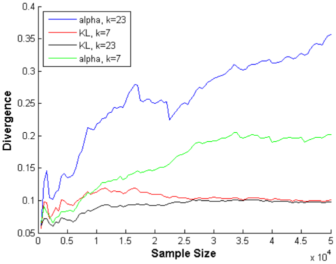
1. Model confidence sets: Hansen et al. (2011)
   - $M_0 = \{m_{0,1}, ..., m_{0,k}\}$ finite set of starting models
   - $M^* = \{i \in M_0 : E[d_{ij}] \leq 0 \forall j \in M_0\}$ with $d_{ij} = l(e_j) - l(e_i)$
   - sequential testing and elimination of models from the set M_0 to reach confidence set $\hat{M}$ for M^*
2. estimate the divergence $\hat{D}_\nu(\eta)$ relying on kNN-estimators (Poczos/Schneider)

Quantifying model risk

1. Model confidence sets: Hansen et al. (2011)
   - $M_0 = \{m_{0,1}, ..., m_{0,k}\}$ finite set of starting models
   - $M^* = \{i \in M_0 : E[d_{ij}] \leq 0 \forall j \in M_0\}$ with $d_{ij} = l(e_j) - l(e_i)$
   - sequential testing and elimination of models from the set M_0 to reach confidence set $\hat{M}$ for M^*
2. estimate the divergence $\hat{D}_\nu(\eta)$ relying on kNN-estimators (Poczos/Schneider)
3. $\hat{\kappa}_\nu = \max_{\eta \in \hat{M}} \hat{D}_\nu(\eta)$

Divergence measure and divergence estimation





Wrap-Up

- α-divergence is often infinite, hard to estimate and it places a strong and non-transparent condition on the potential worst-case distributions

Wrap-Up

- α-divergence is often infinite, hard to estimate and it places a strong and non-transparent condition on the potential worst-case distributions
- Our proposal: Use KL-divergence and restrict attention to bounded functions L, e.g.

$$L_c(y) := \max(-c, \min(c, L(y)))$$

Wrap-Up

- α-divergence is often infinite, hard to estimate and it places a strong and non-transparent condition on the potential worst-case distributions
- Our proposal: Use KL-divergence and restrict attention to bounded functions L, e.g.

$$L_c(y) := \max(-c, \min(c, L(y)))$$

- Work in progress: Divergence measures which lie in between α-divergence and KL-divergence

Numerics

- Recall

$$E_{\eta^{wc}}[L(Y)] = E_{\nu}\left[L(Y)\frac{\exp(\theta L(Y))}{E_{\nu}[\exp(\theta L(Y))]}\right].$$

- If Y is high-dimensional, this expectation must be approximated by Monte Carlo
- Friedman (2002) and Glasserman and Xu (2012) propose to use an Importance Sampling (IS) approach
- Both papers contain numerical examples where IS appears to converge to a finite value despite the fact that the true worst case is infinite
- Problem: Lognormal tails, unbounded L, KL-divergence...
- Thus, the demand for more sophisticated numerical methods is not an artefact from cutting-off

Importance Sampling

- Draw M samples Y_i from ν and write

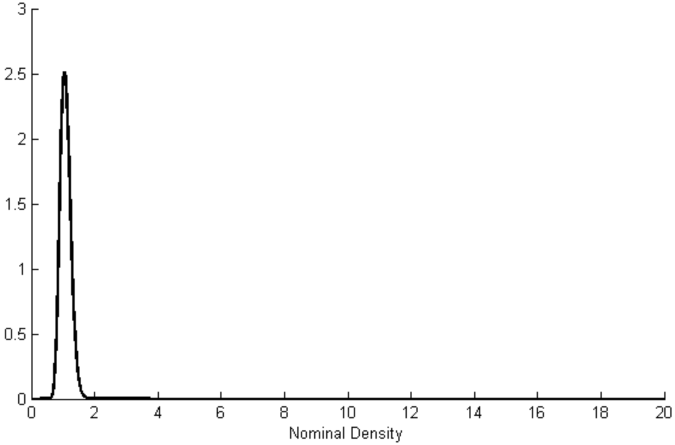
$$E_{\eta^{wc}}[L(Y)] \approx \frac{\frac{1}{M}\sum_i L(Y_i)\exp(\theta L(Y_i))}{\frac{1}{M}\sum_i \exp(\theta L(Y_i))}$$

- Usually: IS used as a variance reduction technique, proposal distribution more heavy-tailed than target
- Here: ν is used to approximate the more heavy-tailed η^{wc} and important regions in the tails might be missed...

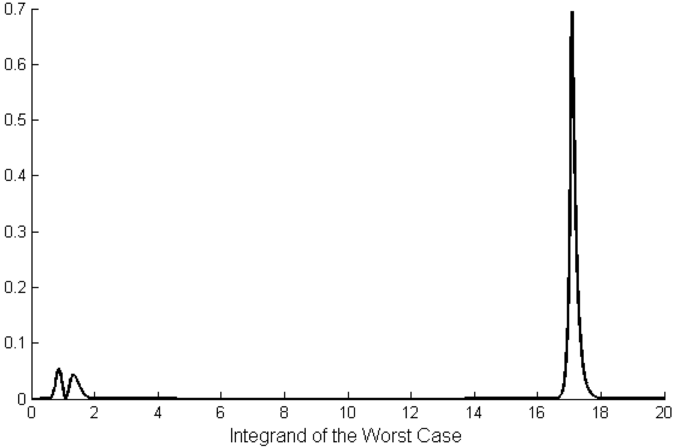
Toy example

- Y lognormal
- $L(y) = |y - E_\nu[Y]|$, cut-off at some finite value
- How well can the nominal density approximate the integrand of the worst case expectation?
- A visual inspection is sufficient this case...

Nominal density



Worst-case integrand



Toy example

- IS means approximating the second picture by the first...
- With moderately many Monte Carlo samples, the IS estimator appears to converge to a value slightly above the nominal mean
- Of course, the estimator does not literally converge to a wrong value, it just converges very slowly...
- Rough calculation: After about 10^{60} years of simulation the value will be about right with high probability...

Illustration

- Suppose that $X_i = 1000$ with probability $1/100$ and 0 otherwise.
- Then $E[X_i] = 10$ and
$$\frac{1}{M}\sum_{i=1}^{M} X_i$$
must converge there by the law of large numbers
- But:
$$P\left[\frac{1}{10}\sum_{i=1}^{10} X_i = 0\right] > 0.9$$

Numerics: MCMC

- What can we do instead?
- Replace IS by Markov Chain Monte Carlo (MCMC)? MCMC samplers are less tied to the nominal distribution and might be an alternative...
- But: (Standard) MCMC samplers do not work well in applications (like ours) where the state space is composed of several disjoint important regions

Numerics: SMC

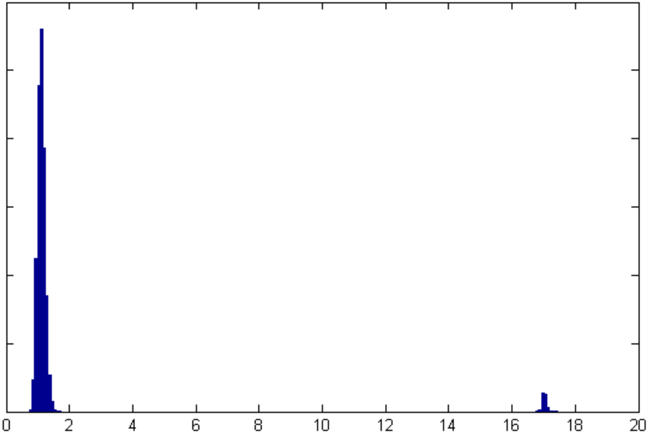
- Sequential Monte Carlo (SMC) samplers are an alternative to MCMC for multimodal target distributions
- Idea: Approximate the target distribution by a sequence of distributions, beginning with one that is easy to sample
- Perform parallel copies of MCMC for each distribution in the sequence, moving samples from one distribution to the next using importance sampling and resampling

Illustration

Parameter Set

$Y \sim LN(0.1, 0.1)$, $L = min(c^2, (y-m)^2)$, $c = 16$;
$m = 1.105; \kappa = 0.05; \theta = 1.4631$;
$\sqrt{E[L]}$: nominal $= 0.110$;
worst case analytical $= 0.2163$;
IS $= 0.1130$;
MCMC with start in 1 $= 0.1124$;
MCMC with start in 15 $= 15.76$;
SMC $= 0.2153$;

Illustration



Summary

- We now have everything in place for our calculation of worst-case quantities...
- Use KL-divergence in combination with cut-off or bounded risk functionals L
- Estimate the amount of model risk by comparing the nominal model to the "best" models we have
- Use SMC to calculate expected values

Summary

- We now have everything in place for our calculation of worst-case quantities...
- Use KL-divergence in combination with cut-off or bounded risk functionals L
- Estimate the amount of model risk by comparing the nominal model to the "best" models we have
- Use SMC to calculate expected values
- Finally: An application

Discrete Hedging

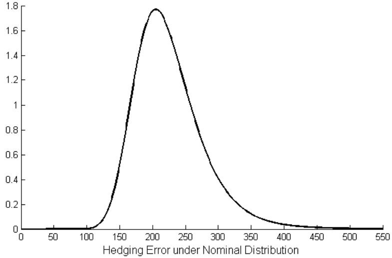
- We want to hedge a simple Call option, i.e., $[S_T - K]^+$
- We consider the (cut-off) quadratic hedging error as the risk measure
- As the nominal model we choose a Black-Scholes model where hedging takes place at discrete points in time
- Y: hedging error,
- Apply the model confidence set to sophisticated (fitted) Garch models (and a BS model with missestimated volatility)
- Contract parameters: $\mu = 0.097$; $\sigma = 0.189$; $T = 1$; $S_0 = 100$; $K = 100$; $r = 0.03$;

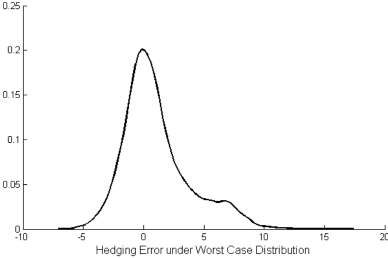
Estimation of the distance

	HS-EGARCH	t-EGARCH	EGARCH	GARCH	GJR
κ	0.20	0.23	0.10	0.28	0.34
θ	0.063	0.066	0.052	0.070	0.074
$E_{\eta^{wc}}[L^b(Y)]^{\frac{1}{2}}$	2.55	2.64	2.19	2.77	2.93

$E_\nu[L(Y)] = 1.32$

Nominal and worst-case distribution





Thank You!