Value-at-Risk, Expected Shortfall and Density Forecasting

Kevin Sheppard
http://www.kevinsheppard.com

Oxford MFE

This version: February 27, 2020

February 2020

Risk Measurement Overview

- What is risk?
- What is Value-at-Risk?
- How can VaR be measured and modeled?
- How can VaR models be tested?
- What is Expected Shortfall?
- How can densities be forecasted?
- How can density models be evaluated?
- What is a coherent risk measure?

Error

Mean $y_{t+h} - \hat{y}_{t+h|t}$

Variance $(r_{t+h} - \hat{\mu}_{t+h})^2 - \hat{\sigma}^2_{t+h|t}$

$\underbrace{HIT_{t+h} - \alpha}$

$I[r_{t+h} < -VaR_{t+h|t}]$

Density Forecasting and Evaluation

Density Estimation and Forecasting

- End all be all of risk measurement
- Issues:
  - Equally hard
  - Lots of estimation and model error
    - Can have non obvious effects on nonlinear functions (i.e. options)
  - Not closed under aggregation
    - No multi-step
- Builds off of the GARCH VaR application

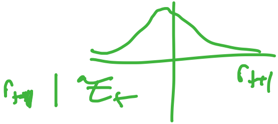
Density forecasts from GARCH models

- Simple constant mean GARCH(1,1)

$$
\begin{aligned}
r_{t+1} &= \mu + \epsilon_{t+1} \\
\sigma^2_{t+1} &= \omega + \gamma \epsilon^2_t + \beta \sigma^2_t \\
\epsilon_{t+1} &= \sigma_{t+1} e_{t+1} \\
e_{t+1} &\overset{\text{i.i.d.}}{\sim} g(0,1).
\end{aligned}
$$

TARCH
EGARCH
ARCH

- g is some known distribution, but not necessarily normal
- Density forecast is simply $g(\mu, \sigma^2_{t+1|t})$
- Flexible through choice of g
- Parsimonious
- Semiparametric works in same way replacing g with the standardized residuals of a "smoothed" estimate

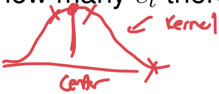


Kernel Densities

- "Smoothed" densities are more precise than rough estimates

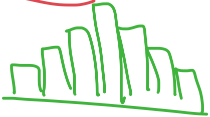
$$g(e) = \frac{1}{Th}\sum_{t=1}^{T} K\left(\frac{\hat{e}_t - e}{h}\right), \quad \hat{e}_t = \frac{y_t - \hat{\mu}_t}{\hat{\sigma}_t} = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t}$$

- Local average of how many $\hat{e}_t$ there are in a small neighborhood of e



- $K(\cdot)$ is a kernel
  - Gaussian

$$K(x) = \frac{1}{\sqrt{2\pi}}\exp(-x^2/2)$$

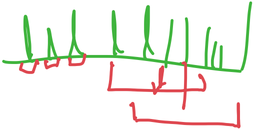


  - Epanechnikov

$$K(x) = \begin{cases} \frac{3}{4}(1 - x^2) & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

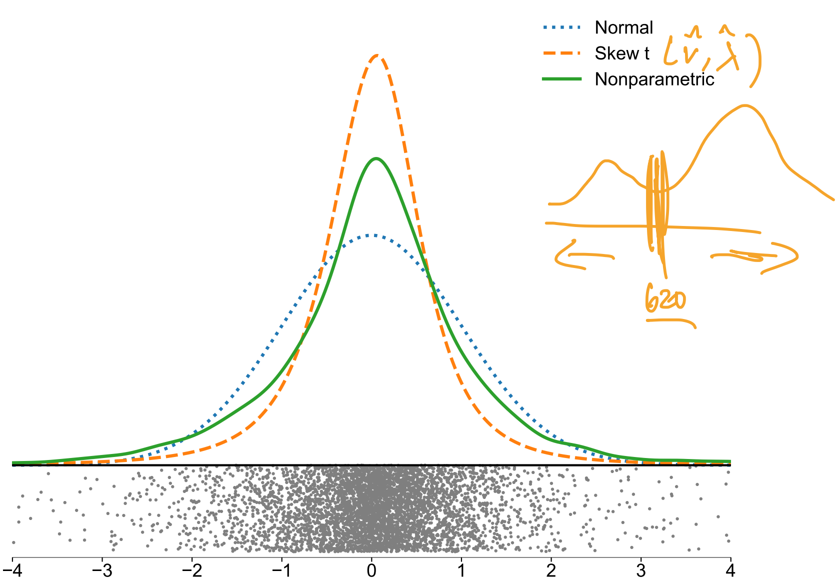
- h: Bandwidth controls smoothing
- Silverman's bandwidth

$$1.06\sigma_x T^{-\frac{1}{5}}$$



  - h too small produces very rough densities (low bias but lots of variance)
  - h too large produces overly smooth (low variance but very biased)

S&P 500 Parametric and Nonparametric Densities



Multi-step Density Forecasts

- Densities do not aggregate in general
  - Multivariate normal is special
- Densities from GARCH models do not easily aggregate
- 1-step density forecast from a standard GARCH(1,1)

$$r_{t+1}|\mathcal{F}_t \sim N(\mu, \sigma^2_{t+1|t})$$

- Wrong 2-step forecast from a standard GARCH(1,1)

$$r_{t+2}|\mathcal{F}_t \sim N(\mu, \sigma^2_{t+2|t})$$

?? Monte Carlo

- Correct 2-step forecast from a standard GARCH(1,1)

$$r_{t+2}|\mathcal{F}_t \sim \int_{-\infty}^{\infty} \phi(\mu, \sigma^2(e_{t+1})_{t+2|t+1})\phi(e_{t+1})\mathrm{d}e_{t+1}.$$

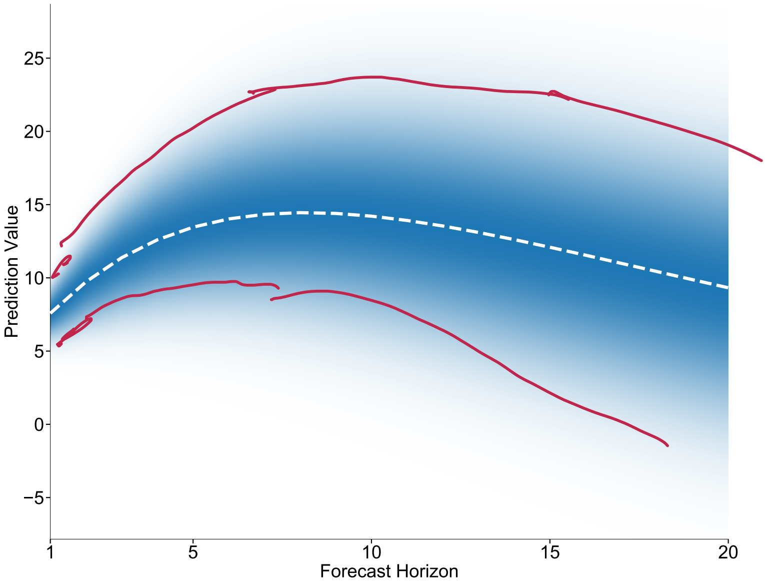
"t+1"

- Must integrate out the variance uncertainty between $t+1$ and $t+2$
- Easy fix: directly model $t+2$ (or $t+h$)

The Fan plot

- Hard to produce time-series of densities
- Solution is the Fan Plot
- Popularized by the Bank of England
- Horizontal axis (x) is the number of time-periods ahead
- Vertical axis (y) is the vale the variable might take
- Density is expressed using varying degrees of color intensity.
  - Dark color indicate the highest probability
  - Progressively lighter colors represent decreasing likelihood
  - Essentially a plot of many quantiles of the distribution through time
- A lot of “wow”
- Not necessarily a lot of content

A fan plot for an AR(2)



Density “Standardized” Residuals

- Consider a generic stochastic process $\{y_t\}$
  - Residuals from mean models:

$$\hat{\epsilon}_t = y_t - \hat{\mu}_t$$

  - Residuals from variance models:

$$\hat{e}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t} = \frac{y_t - \hat{\mu}_t}{\hat{\sigma}_t}$$

z

e_t

  - Residuals from Value-at-Risk models:

$$HIT_t = \underline{I_{[y_t < q_t]}} - \alpha$$

  - Residual from density models:

$$\hat{u}_t = F_t(y_t)$$

CDF

- Known as the Probability Integral Transformed Residuals.
- One very useful property: If $y_t \sim F$ then $u_t \equiv F(y_t) \sim U(0, 1)$

Probability Integral Transform

Theorem (Probability Integral Transform)

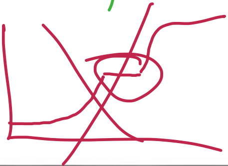
Let a random variable X have a continuous, increasing CDF $F_X(x)$ and define $Y = F_X(X)$. Then Y is uniformly distributed and $\Pr(Y \leq y) = y$, $0 < y < 1$.

For any $y \in (0,1)$, $Y = F_X(X)$, and so

$$\begin{aligned}
\Pr(Y \leq y) &= \Pr(F_X(X) \leq y) \\
&= \Pr(F_X^{-1}(F_X(X)) \leq F_X^{-1}(y)) && \text{Since } F_X^{-1} \text{ is increasing} \\
&= \boxed{\Pr(X \leq F_X^{-1}(y))} && \text{Invertible since strictly increasing} \\
&= F_X(F_X^{-1}(y)) && \text{Definition of } F_X \\
&= y
\end{aligned}$$

Q Pr(X ≤ Q)

F_x(Q)



Evaluating Density Forecasts: QQ Plots

- Quantile-Quantile Plots
- Plots the data against a hypothetical distribution

A

$$\hat{e}_1 < \hat{e}_2 < \ldots < \hat{e}_{N-1} < \hat{e}_N$$

$\frac{1}{N+1}, \frac{1}{T+1}$

T/T+1

F

(Normal)

  - $N = T$ but used to indicate that the index is not related to time
- e_n against $F^{-1}(\frac{j}{T+1})$

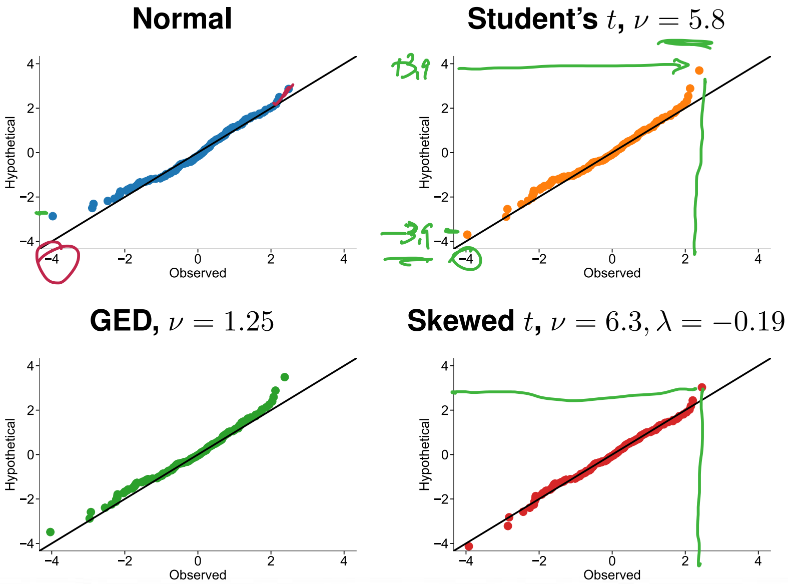
B

$$F^{-1}\left(\frac{1}{T+1}\right) < F^{-1}\left(\frac{2}{T+1}\right) < \ldots < F^{-1}\left(\frac{T-1}{T+1}\right) < F^{-1}\left(\frac{T}{T+1}\right)$$

- F^{-1} is inverse CDF of distribution being used for comparison
- Should lie along a 45^o line
- No confidence bands

QQ Plots for the S&P 500

Monthly Returns



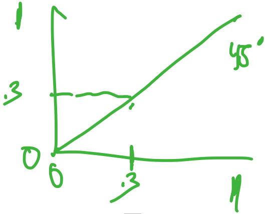
Evaluating Density Forecasts: Kolmogorov-Smirnov

- Formalizes QQ plots
- Key property
  - If $x \sim F$, then $u \equiv F(x) \sim U(0,1)$
  - Can test $U(0,1)$

$F \sim N(\mu, \sigma^2)$

$N(\hat{\mu}, \hat{\sigma}^2)$

Lilliefors



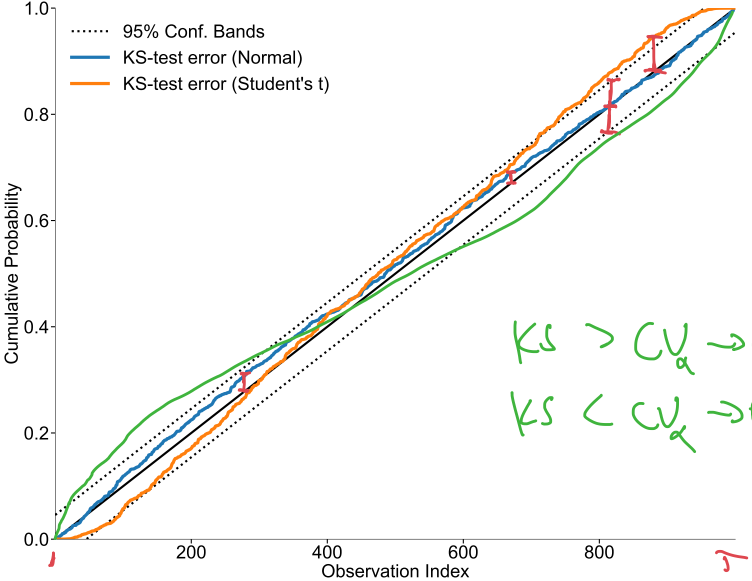
- KS tests maximum deviation from U(0,1)

$$\max_{\tau} \left| \frac{1}{T}\left(\sum_{i=1}^{\tau} I_{[u_i < \frac{\tau}{T}]}\right) - \frac{\tau}{T} \right|, \quad \tau = 1, 2, \ldots, T$$

$\frac{\tau}{T} \in (0,1)$

  - $\frac{1}{T}\sum_{i=1}^{\tau} I_{[u_j < \frac{\tau}{T}]}$: *Empirical percentage of* u below τ/T
  - τ/T: How many *should* be below τ/T
- Nonstandard distribution
- Parameter estimation error
  - Parameter Estimation Error (PEE) causes significant size distortions
  - Using a 5% CV will only reject 0.1% of the time
  - Solution is to simulate the needed critical values

The Kolmogorov-Smirnov Test



$KS > CV_\alpha \rightarrow$ reject

$KS < CV_\alpha \rightarrow$ FTR

Addressing PEE in a KS test

- Model is a complete model so can be easily simulated
- Exact KS distribution tabulated

$r_t = \mu + \epsilon_t$

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

$\epsilon_t = \sigma_t e_t$

$e_t \sim g(0,1)$

Algorithm (Correct CV for KS test with PEE)

1. *Estimate model and save* $\hat{\theta}$

   $\mu, \omega, \alpha, \beta$

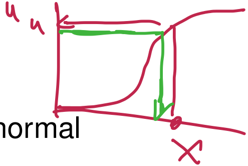
2. *Repeat many times (1000+)*
   a. *Simulate artificial series from model using* $\hat{\theta}$ *with same number of observations as original data*
   b. *Estimate parameters from simulated data,* $\ddot{\theta}$
   c. *Compute KS test statistic on simulated data using* $\ddot{\theta}$ *and save as* $KS_i, i = 1, 2, \ldots,$
3. *Sort the* KS_i *values and use the* $1 - \alpha$ *quantile for get correct CV for* α *size test*

   (5%) KS_{950}

$KS_1 < KS_2 < \ldots < KS_{1000}$

Evaluating Density Forecasts: Berkowitz Test

- Berkowitz Test extends KS to evaluation of conditional densities
- Exploits probability integral transform property



$$\hat{u}_t = F(y_t)$$

- But then *re-transforms* the data to a standard normal

$$\hat{\eta}_t = \Phi^{-1}(\hat{u}_t) = \Phi^{-1}(F(y_t))$$

Inv CDF

  - Since $\hat{u}_t \overset{\text{i.i.d.}}{\sim} U(0,1)$, $\hat{\eta}_t \overset{\text{i.i.d.}}{\sim} N(0,1)$
- Test is a likelihood ratio test using an AR(1)

$$\hat{\eta}_t = \phi_0 + \phi_1 \hat{\eta}_{t-1} + \nu_t$$

- If the model is correctly specified
  - $\phi_0 = 0$, $\phi_1 = 0$, $\sigma^2 = \mathrm{V}[\nu_t] = 1$
- Likelihood ratio

$$2\left(l(\eta_t|\hat{\phi}_0, \hat{\phi}_1, \hat{\sigma}^2) - l(\eta_t|\phi_0 = 0, \phi_1 = 0, \sigma^2 = 1)\right) \sim \chi^2_3$$

  - Critical values wrong if F has estimated parameters

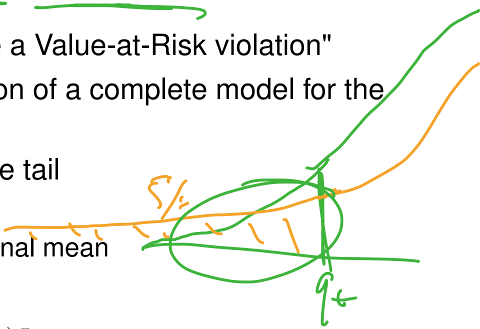
Expected Shortfall

Expected Shortfall

- Conditional Expected Shortfall (ES, also called Tail VaR)

$$ES_{t+1} = \mathrm{E}_t[r_{t+1} | r_{t+1} < -VaR_{t+1}]$$

- "Expected Loss given you have a Value-at-Risk violation"
- Usually requires the specification of a complete model for the conditional distribution
- Uses all of the information in the tail
- Evaluation
  - Standard Problem, a conditional mean
  - GMZ regression



$$(ES_{t+1|t} - R_{t+1}) I_{[R_{t+1} < -VaR_{t+1|t}]} = \mathbf{x}_t \boldsymbol{\gamma}$$

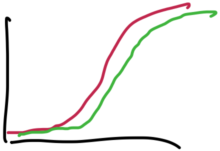
  - $H_0 : \boldsymbol{\gamma} = \mathbf{0}$

- Difficult to test since relatively few observations

Idealized Risk Measures

Coherent Risk Measures

- Coherence is a desirable property for a risk measure
  - But not completely necessary
- ρ is the required capital necessary according to some measure of risk (VaR, ES, Standard Deviation, etc.)
- P, P_1 and P_2 are portfolios of assets
- A Coherent measure satisfies:



Drift Invariance

$$\rho(P + c) = \rho(P) - c$$

Homogeneity

$$\rho(\lambda P) = \lambda \rho(P) \quad \text{for any } \lambda > 0$$

Monotonicity If P_1 first order stochastically dominates P_2, then

$$\rho(P_1) \leq \rho(P_2)$$

Subadditivity

$$\rho(P_1 + P_2) \leq \rho(P_1) + \rho(P_2)$$

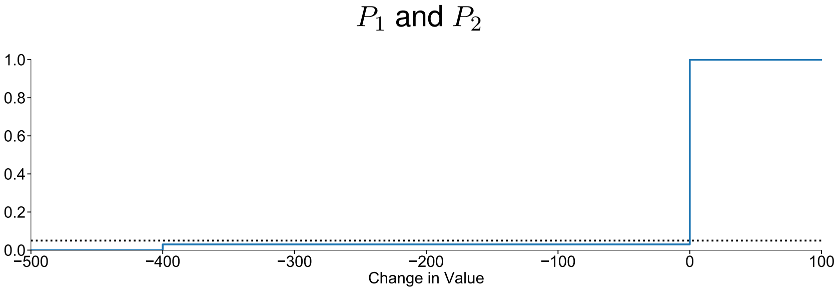
Coherent Risk Measures

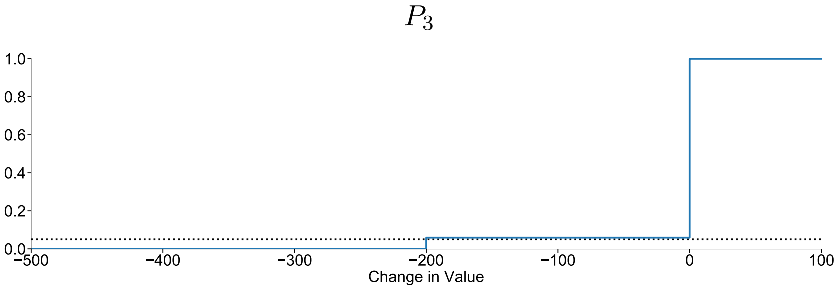
- VaR is *not* coherent
  - Because VaR is a quantile it may not be subadditive

VaR is Not Coherent

- Two portfolios P_1 and P_2 holding a bond
  - Each paying 0%, par value of $1,000
  - Default probability 3%, recovery rate 60%
  - Two companies, defaults are independent
- Value-at-Risk of P_1 and $P_2$is $0
- Value-at-RIsk of $P_3 = 50\% \times P_1 + 50\% \times P_2 = \$200$
  - 5.91% that one or both default

Coherent Risk Measures





Coherent Risk Measures

- ES is coherent
  - Doesn't mean much
  - VaR still has a lot of advantages
  - More importantly VaR and ES agree in most realistic settings

ES is coherent

- ES of P_1 and P_2 is \$240
  - Given in lower 5% of distribution, 60% chance of a loss of \$400
- ES of P_3
  - Given in lower 5% of distribution:
    - ${}^{0.0009}/_{0.05} = .018$ probability of \$400 loss (2 defaults)
    - ${}^{0.0491}/_{0.05} = .982$ probability of \$200 loss (1 default)
    - ES of \$7.20 + \$196.40 = \$203.60

- ES is subadditive when VaR is not


h correct
too small