Value-at-Risk, Expected Shortfall and Density Forecasting

Kevin Sheppard
`http://www.kevinsheppard.com`

Oxford MFE
This version: February 20, 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?

Risk

- What is risk?
- Market Risk
  - Liquidity Risk
  - Credit Risk
  - Counterparty Risk
  - Model Risk
  - Estimation Risk
- Today's focus: Market Risk
- Tools
  - Value-at-Risk
  - Expected Shortfall
  - Density Estimation

Value-at-Risk

Value-at-Risk

- Value-at-Risk is a standard tool of risk management
  - Basel Accord

Definition (Value-at-Risk)

The α Value-at-Risk (VaR) of a portfolio is defined as the largest number such that the probability that the loss in portfolio value over some period of time is greater than the VaR is α,

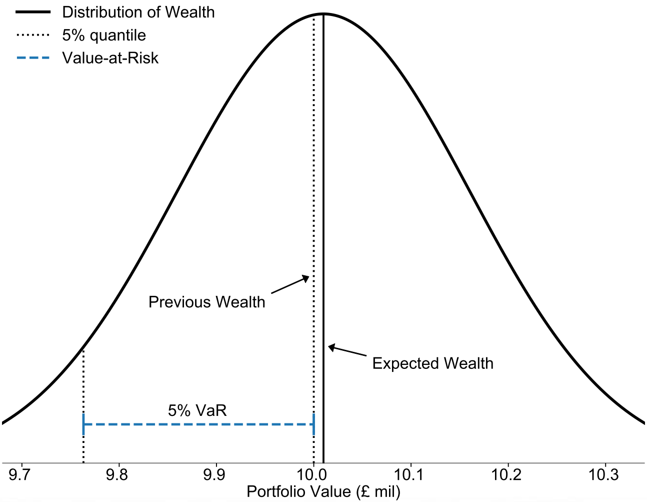
$$Pr(R < -VaR) = \alpha$$

where $R = W_1 - W_0$ is the total return on the portfolio, W_t, $t = 0,\ 1$, is the value of the assets in the portfolio and 1 and 0 measure an arbitrary time span (e.g. one day or two weeks).

- Units are \$, £, ¥
- Almost always *positive*
- It is a **quantile**

Value-at-Risk in a picture

- Returns are $N(.001, .015^2)$
- W_0 is £10,000,000



Percent Value-at-Risk

- Value-at-Risk can be normalized and reported as a %

Definition (Percentage Value-at-Risk)

The α percentage Value-at-Risk (%VaR) of a portfolio is defined as the largest return such that the probability that the return on the portfolio over some period of time is less than -%VaR is α,

$$Pr(r < -\%VaR) = \alpha$$

where r is the percentage return on the portfolio. %VaR can be equivalently defined as %VaR = VaR/W_0.

- Units are returns (no units)
- Also almost always *positive*
- Lets VaR be interpreted without knowing the value of the portfolio, W_0
  - No meaningful loss of information from standard VaR
  - Used throughout rest of lecture in place of formal definition of VaR

Relationship between Quantiles and VaR

- VaR is a quantile
  - Quantile of the future distribution

Definition (α-Quantile)

The α-quantile of a random variable X is defined as the *smallest* number q_α such that

$$\Pr(X \leq q_\alpha) = \alpha$$

- Other "-iles"
  - Tercile
  - Quartile
  - Quintile
  - Decile
  - Percentile

Conditional and Unconditional VaR

- Condiitonal VaR is similar to conditional mean or conditional variance

Definition (Conditional Value-at-Risk)

The conditional α Value-at-Risk is defined as

$$Pr(r_{t+1} < -VaR_{t+1|t}|\mathcal{F}_t) = \alpha$$

where $r_{t+1} = (W_{t+1} - W_t)/W_t$ is the return on a portfolio at time $t+1$.

- t is an arbitrary measure of time $\Rightarrow t+1$ also refers to an arbitrary unit of time
  - day, two-weeks, 5 years, etc.
- Incorporates all information available at time t to assess risk at time $t+1$
- Natural extension of conditional expectation and conditional variance to conditional quantile

Models for Value-at-Risk

Conditional VaR: RiskMetrics

- Industry standard benchmark
- Restricted GARCH(1,1)

$$\sigma_{t+1}^2 = (1-\lambda)r_t^2 + \lambda\sigma_t^2$$

- *Exponentially Weighted Moving Average (EWMA)*:

$$\sigma_{t+1}^2 = \sum_{i=0}^{\infty}(1-\lambda)\lambda^i r_{t-i}^2$$

$$VaR_{t+1} = -\sigma_{t+1}\Phi^{-1}(\alpha)$$

- $\Phi^{-1}(\cdot)$ is the inverse normal CDF
- Advantages
  - No parameters to estimate
  - λ =.94 (daily data), .97 (weekly), .99 (monthly)
  - Easy to extend to portfolios (see notes)
- Disadvantages
  - No parameters to estimate
  - No leverage effect
  - Random Walk VaR

Conditional VaR: GARCH models for Value-at-Risk

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

- Value-at-Risk:

$$VaR_{t+1} = -\hat{\mu} - \hat{\sigma}_{t+1}F^{-1}_{\alpha}$$

- F^{-1}_{α} is the α quantile of the distribution of e_{t+1}
  - For example, 1.645 for the 5% from a normal
- Advantages
  - Flexible volatility model and easy to estimate
- Disadvantages
  - Must chose f (know f to get the correct VaR)
  - Location-Scale families

Conditional VaR: Semiparametric/Filtered HS

- Parametric GARCH + Nonparametric Density $\rightarrow$ Semi-parametric VaR

$$e_{t+1} \overset{\text{i.i.d.}}{\sim} g(0,1), \quad g\text{unknown distribution}$$

- Implementation
  1. Fit an ARCH model using Normal QMLE
  2. $\hat{e}_t = \frac{\hat{\epsilon}_t}{\hat{\sigma}_t}$
  3. Order residuals

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

- Quantile is residual $\alpha \times N$ residual (N=T).

$$VaR_{t+1}(\alpha) = -\hat{\mu} - \hat{\sigma}_{t+1}\hat{G}_{\alpha}^{-1}$$

- Advantages
  - All advantages of GARCH
  - Quantile converges to true quantile
- Disadvantages
  - Location-Scale families
  - Quantile convergence is *slow*

Conditional VaR: CaViaR

- Conditional Autoregressive Value-at-Risk (ARCVaR)
  - Conditional quantile *regression*
  - Directly parameterize quantile $F_\alpha^{-1} = q$ of the return distribution

$$q_{t+1} = \omega + \gamma HIT_t + \beta q_t$$

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

$$VaR_{t+1} = -q_{t+1}$$

- Advantages
  - Focuses on quantile
  - Flexible specification
- Disadvantages
  - Hard to estimate
  - Which specification?
  - Out-of-order VaR: 5% can less than 10% VaR

Estimation of CaViaR models

- Many CaViaR specifications
  - Symmetric
$$q_{t+1} = \omega + \gamma HIT_t + \beta q_t.$$
  - Symmetric absolute value,
$$q_{t+1} = \omega + \gamma |r_t| + \beta q_t.$$
  - Asymmetric absolute value
$$q_{t+1} = \omega + \gamma_1 |r_t| + \gamma_2 |r_t| I_{[r_t<0]} + \beta q_t$$
  - Indirect GARCH
$$q_{t+1} = \left(\omega + \gamma r_t^2 + \beta q_t^2\right)^{\frac{1}{2}}$$
- Estimation minimizes the “tick” loss function
$$\underset{\boldsymbol{\theta}}{\operatorname{argmin}} \; T^{-1} \sum_{t=1}^{T} \alpha (r_t - q_t)(1 - I_{[r_t<q_t]}) + (1-\alpha)(q_t - r_t) I_{[r_t<q_t]}$$
  - Non-differentiable
  - Requires “derivative-free” optimizers (e.g. simplex optimizers)

Weighted Historical Simulation

- Uses a weighted empirical CDF
- Weights are exponentially decaying

$$w_i = \lambda^{t-i}(1-\lambda)/\left(1-\lambda^t\right), \;\; i = 1, 2, \ldots, t$$

- Weighted Empirical CDF

$$\hat{G}_t(r) = \sum_{i=1}^{t} w_i I_{[r_i < r]}$$

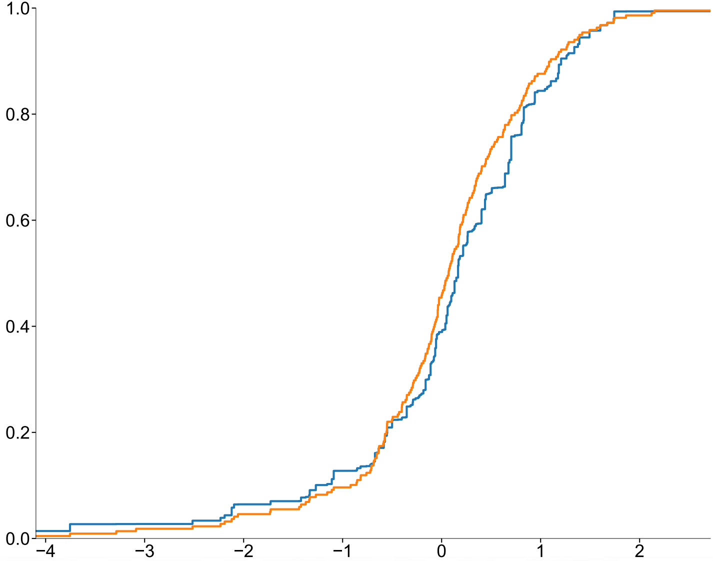
- Conditional VaR is solution to

$$\mathrm{VaR}_{t+1} = \min_r \hat{G}(r) \geq \alpha$$

- Example uses $\lambda = 0.975$

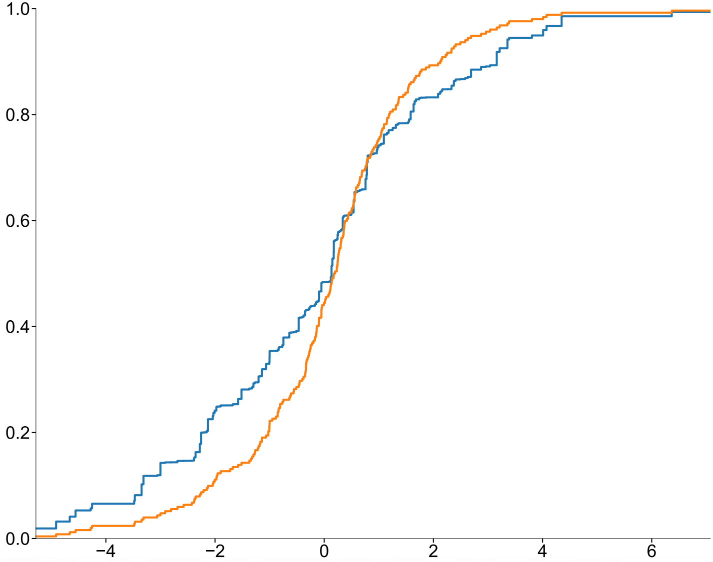
Weighted Historical Simulation

2018

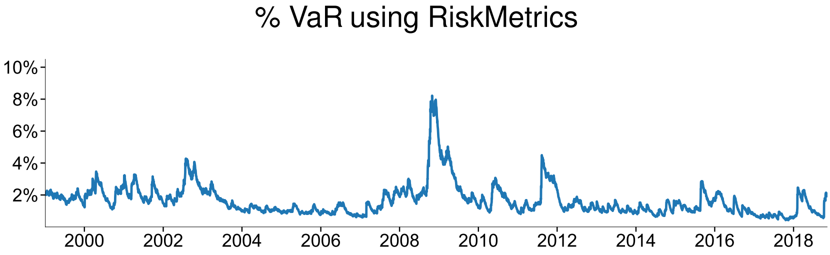


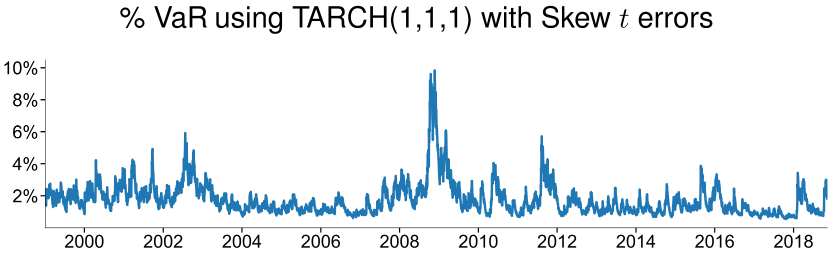
Weighted Historical Simulation

2009

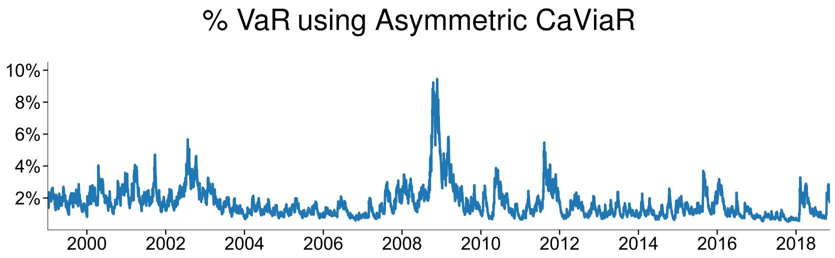


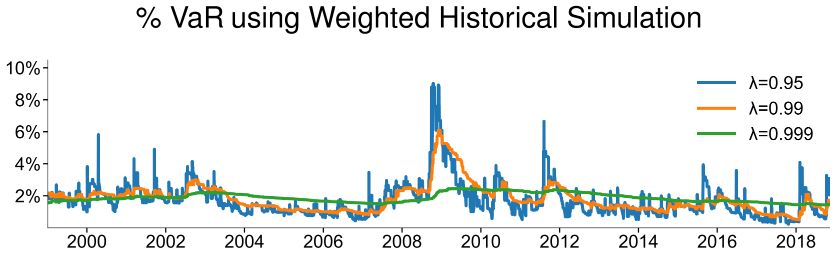
S&P 500 VaR Estimates, $\alpha = 5\%$





S&P 500 VaR Estimates, $\alpha = 5\%$





Unconditional VaR

- Parametric Estimation
  - Specify some fully parametric model for returns
  - Estimate the parameters by MLE
  - VaR is the α-quantile of the fit distribution
- Nonparametric Estimation (Historical Simulation)
  - Nonparametric estimation of the density of returns using raw data
  - Identical to previous density estimation
  - Can “smooth” to reduce variance
- Parametric Monte Carlo
  - Estimate a conditional model for short horizon returns
  - Simulate the model for many periods
  - Use a nonparametric estimate of the density of the simulated returns

Evaluation of Value-at-Risk Models

Evaluating VaR models

- Basic instrument for testing VaR is the "Hit"

$$ge_t = I_{[r_t < F_t^{-1}]} - \alpha = HIT_t$$

- Is the *generalized error* from the "tick" loss function
- If the VaR is correct,
$$\mathrm{E}_{t-1}[HIT_t] = 0$$
- Leads to a standard Generalized Mincer-Zarnowitz evaluation framework
- Hit Regression

$$HIT_{t+h} = \gamma_0 + \gamma_1 VaR_{t+h|t} + \gamma_2 HIT_t + \gamma_3 HIT_{t-1} + \ldots + \gamma_K HIT_{t-K+1}$$

  - Null is $H_0 : \gamma_0 = \gamma_1 = \ldots = \gamma_K = 0$
  - Alternative is $H_1 : \gamma_j \neq 0$ for some j
- As always, GMZ can be augmented with any time t measurable variable

Unconditional Evaluation of VaR using the Bernoulli

- $\widetilde{HIT}$s from a correct VaR model have a Bernoulli distribution
  - 1 with probability α
  - 0 with probability $1-\alpha$
- Likelihood for T Bernoulli random variables $x_t \in \{0,1\}$

$$f(x_t;p) = \prod_{t=1}^{T} p^{x_t}(1-p)^{1-x_t}$$

- Log-likelihood is

$$l(p;x_t) = \sum_{t=1}^{T} x_t \ln p + (1-x_t)\ln 1 - p$$

- In terms of α and $\widetilde{HIT}_t$

$$l(\alpha;\widetilde{HIT}_t) = \sum_{t=1}^{T} \widetilde{HIT}_t \ln \alpha + \left(1-\widetilde{HIT}_t\right)\ln 1 - \alpha$$

- Easy to conduct a LR test

$$LR = 2(l(\hat{\alpha};\widetilde{HIT}) - l(\alpha_0;\widetilde{HIT})) \sim \chi^2_1$$

- $\hat{\alpha} = T^{-1}\sum_{t=1}^{T}\widetilde{HIT}_t$, α_0 is the α from the VaR

Evaluation of Conditional VaR using the Bernoulli

- Can also be extended to testing conditional independence of HITs
- Define

$$n_{00} = \sum_{t=1}^{T-1} (1 - \widetilde{HIT}_t)(1 - \widetilde{HIT}_{t+1}), \qquad n_{10} = \sum_{t=1}^{T-1} (1 - \widetilde{HIT}_t)\widetilde{HIT}_{t+1}$$

$$n_{01} = \sum_{t=1}^{T-1} \widetilde{HIT}_t(1 - \widetilde{HIT}_{t+1}), \qquad n_{11} = \sum_{t=1}^{T-1} \widetilde{HIT}_t\widetilde{HIT}_{t+1}$$

- The log-likelihood for the sequence two VaR exceedences is

$$l(p; \widetilde{HIT}) = n_{11} \ln(p_{11}) + n_{01} \ln(1 - p_{11}) + n_{00} \ln(p_{00}) + n_{10} \ln(1 - p_{00})$$

Evaluation of Conditional VaR using the Bernoulli

- Null is $H_0 : p_{11} = 1 - p_{00} = \alpha$
- MLEs are

$$\hat{p}_{00} = \frac{n_{00}}{n_{00} + n_{10}}, \ \ \hat{p}_{11} = \frac{n_{11}}{n_{11} + n_{01}}$$

- Tested using a likelihood ratio test

$$LR = 2(l(\hat{p}_{00}, \hat{p}_{11}; \widetilde{HIT}) - l(p_{00} = 1 - \alpha, p_{11} = \alpha; \widetilde{HIT}))$$

- Test statistic follows a χ^2_2 distribution

Relationship to Probit/Logit

- Standard GMZ regression is not an ideal test
- Ignores special structure of a HIT
- A HIT is a limited dependant variable
    - Only takes one of two values
- Define a modified hit $\widetilde{HIT}_t = I_{[r_t < F_t^{-1}]}$
    - Takes the value 1 with probability α and 0 with probability $1 - \alpha$
    - Name that distribution $\rightarrow$
- Leads to a modified regression framework known as a probit or logit
    - Probit:
$$\widetilde{HIT}_{t+1} = \Phi\left(\gamma_0 + \mathbf{x}_t \boldsymbol{\gamma}\right)$$
        - If model is correct, $\gamma_0 = \Phi^{-1}(\alpha)$ and $\boldsymbol{\gamma} = \mathbf{0}$
        - Estimated using Bernoulli Maximum Likelihood
        - Easy to compute Likelihood ratio
- Accounts for the limited range of the variable and that the density is non-normal
- Allows for simple-yet-powerful likelihood ratio tests under the null

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)

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

- 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

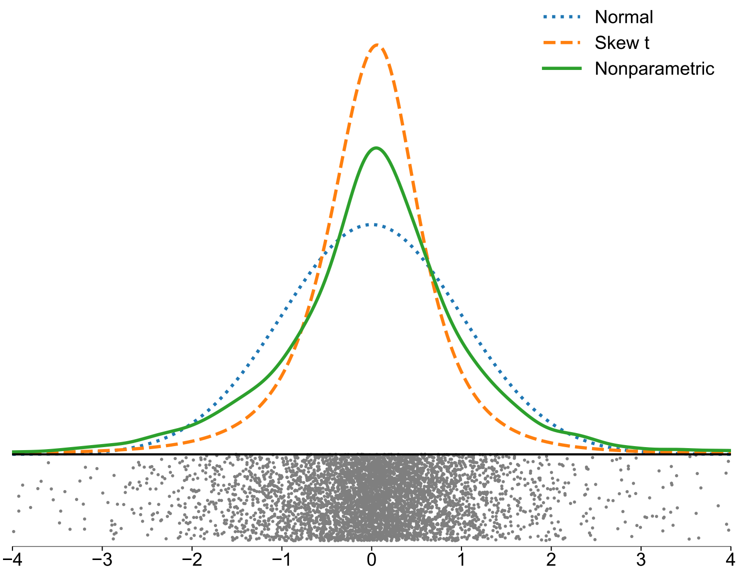
$$K(x) = \begin{cases} \frac{3}{4}(1-x^2) & -1 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

- 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)

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

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

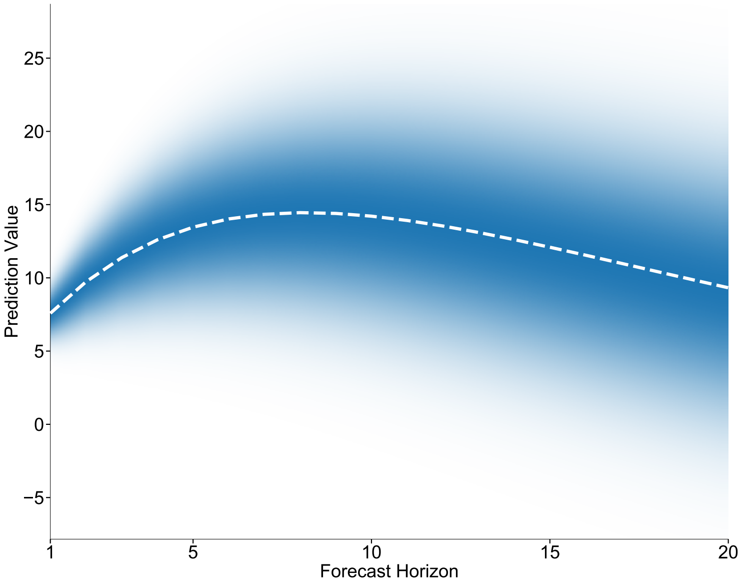
$$r_{t+2}|\mathcal{F}_t \sim \int_{-\infty}^{\infty} \phi(\mu, \sigma^2(e_{t+1})_{t+2|t+1})\phi(e_{t+1})\mathrm{d}e_{t+1}.$$

- 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:

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

  - Residuals from Value-at-Risk models:

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

  - Residual from density models:

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

- 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

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

Evaluating Density Forecasts: QQ Plots

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

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

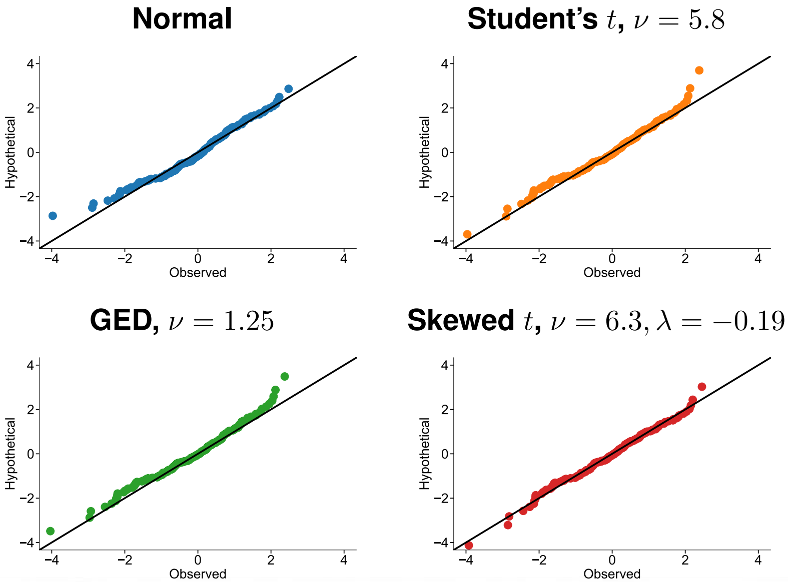
  - $N = T$ but used to indicate that the index is not related to time
- e_n against $F^{-1}(\frac{j}{T+1})$

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

- 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



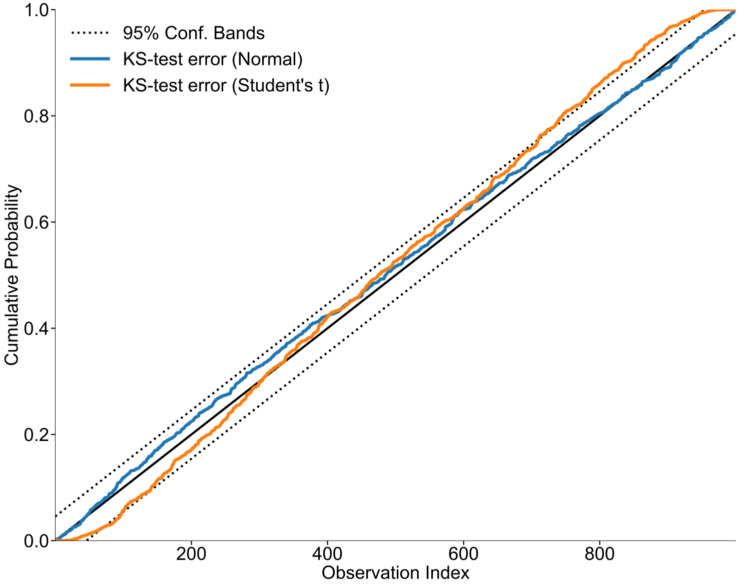
Evaluating Density Forecasts: Kolmogorov-Smirnov

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

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

  - $\frac{1}{T} \sum_{i=1}^{\tau} I_{[u_j < \frac{\tau}{T}]}$: *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



Addressing PEE in a KS test

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

Algorithm (Correct CV for KS test with PEE)

1. *Estimate model and save* $\hat{\boldsymbol{\theta}}$
2. *Repeat many times (1000+)*
   a. *Simulate artificial series from model using* $\hat{\boldsymbol{\theta}}$ *with same number of observations as original data*
   b. *Estimate parameters from simulated data,* $\ddot{\boldsymbol{\theta}}$
   c. *Compute KS test statistic on simulated data using* $\ddot{\boldsymbol{\theta}}$ *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*

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

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

  - 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

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

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

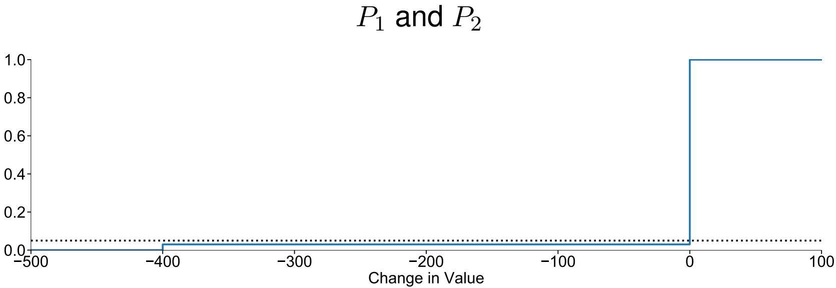
Coherent Risk Measures

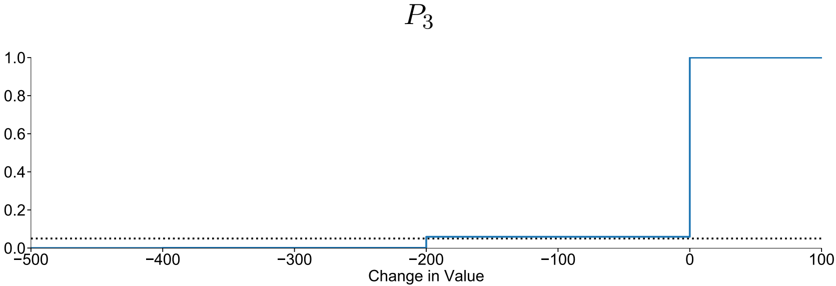
- VaR is *not* coherent
  - Because VaR is a quantile it may not be subadditive

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

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