1. general information

- emphasis on market portfolios
- focus on statistics and simulations
- Genesis of risk
  - fat tails and skeweness
  - non-linearities
  - wrong size
  - wrong horizon
  - Path dependency (your idea eventually works out but you are out of a job or without clients by then)
  - Model risk

2. statistics

- pdf,cdf,invcdf
- p(y|θ) as common in Bayesian statistics
- functions of random variables(rv) ★ see paper notes
- the next term deteaild at "res"
- Stationarity:
  - Stationarity: the unconditional distribution of any observation can be defined independently of time
  - covariance stationary: the unconditional mean and variance exist
  - If we can isolate processes that are at least approximately stationary, we can build better models and decisions.
- The autocorrelation function ρ(h): the correlation between yt and yt
- white noise if it is covariance stationary and uncorrelate
  - $e_t = u_t u_{t-1}^2$ is not strict white noise
  - see ps01
- NB! The unconditional distribution of yt is theoretically defined only for stationary processes
- conditional distribution: $p(y_{t+h}|\Omega_t)$
  - eg ★ see paper notes
- skewness and kurtosis ★ see paper notes
- Some common measures of "risk":
  - Standard deviation
  - Semi-deviation,
  - mean absolute deviation(MD or MAD)
    - $\sum |y - mean(y)|$
  - semi-mad
  - VaR (value-at-risk)
  - ES (expected short-fall)
    - ES is a more relevant and better (better properties) measure of risk than VaR
    - ES is even harder to estimate accurately than VaR
    - heavy-tailed: ES is more relevant and also cannot estimated accurately
  - Drawdown
  - Stress test (or "scenarios")
  - Convexity heuristics
    - ★ see paper notes

3. maximum likelihood

- OLS
  - however, normality is a very poor assumption for most financial returns
  - as the sample gets sufficiently large, MLE is better (less variance of errors)
- Maximum (log) likelihood
  - Probability density function

    - Linear model with Gaussian errors
    - ★ see paper notes
    - OLS is asymptotically optimal in a well-defined sense for Gaussian
- Asymptotic distribution of ML estimators
    - $\theta_{ML} \sim N(\theta_0, I(\theta_0)^{-1})$ where $\theta_0 = \theta_{TRUE}$
    - $\hat{I}(\theta_{ML}) = -\sum_{i=1}^{n} \frac{\partial^2 I(\theta_{ML}, y_i)}{\partial\theta_{ML}\partial\theta'_{ML}}$
- Maximum likelihood in practice
    - ...

4. normal, log-normal, CLT, fat tails

- z is risk-factor, and risk is on portfolio value v = f(z)
    - risk increase when:
        1. z is thick-tailed
        2. f(.) is non-linear (concave for right and convex for left) ★ see paper notes
    - an example of left-tail and thick-tail
        - z is normal
        - $y = \beta_0 + \beta_1 z + \beta_2 z^2 * I(z < 0)$
    -
        - an example of thick-tailed distribution
            - $r_t = u_t u_{t-1}^2$ with normal u
- Why is the Gaussian distribution "normal"?
    - the sum of independent normals is still normal
    - central-limit-theorems
        - sums of su¢ ciently many non-normal r.v. are approximately normal.
        - the speed to normal depends on the individual random variables differ from normal
    - break or slow down CLT
        - infinite variance
        - not iid rv, but strong dependence
    - geometric average, CLT not hold (such as price)
        - but may be log normal
        - but can use log: (log (product) = sum (log)) ~ normal
- log normal
    - $mean = exp(\mu + \frac{1}{2}\sigma^2)$
    - for small SD, approx to normal
    - right skewed and leptokurtic
    - Thinner tail if negatively affected by size
    - Thicker tail if postively affected by size
- extreme value theory

5. financial time series

- tails are much thicker(fat!)
    - in QQ plot, left lower and right higher
    - * skew < 0, kurtosis >> 3
- not continuous
    - jump
- **clustered volatility**: Strong and long-lived serial correlation of |r| and r^2
- (many but not all) return series are strongly left-skewed
    - indexes are more skew than individual
    - monthly are more skew than daily
- CLT in most cases works very slowly
    - CLT: be normal from day->week->month->quarter->year
    - but slowly
    - **but: Skew in particular can initially grow with the horizon**

- reversion in some important variables (profit margins, valuations, oil prices...)
- often: use σ(h)=$\sqrt{h}$σ(1) to compute VaR(h), but:

  1. not normal
  2. more skewed at longer horizons
  - result underestimate

6. student-t

- portfolio with J assets
  - multivariate model
    - $R_t \sim MN(\mu, \sum)$
    - $R_{p,t} \sim N(w^T\mu, w^T \sum w)$
  - $R_{p,t} = \sum_{j=1}^{J} w_j R_{j,t}$
- $\chi^2(df = v) = z_1^2 + z_2^2 + \ldots + z_n^2$
- $F(m, n) = \frac{\chi^2(n)/n}{\chi^2(m)/m}$
- student-t
  - $t(m, s, v) = N(m, s) * \sqrt{\frac{v}{\chi^2}}$
  - !!!! s^2 is not variance, but scale
  - Thicker tails for lower v
  - $var(t) = s^2 \frac{v}{v-2}$
  - pdf of t
- maximum Log-likelihood with student t errors
  - log-likelihood is extremely flat in v for v > 30 (hard to know the maximum one)
  - Using log(v) in estimation is an improvement
  - make sure to initialize the optimization from a low value of v, say, 4-8
  - cannot model skeweness
- The t distribution can generate very fat tails but we must be aware of its limitations:
  - it cannot model skeweness
  - somes prefer to think of extreme events as generated by a different process
  - One way to move past both limitations is by using mixtures

7. optimization; mixtures of normals and t

- optimization (to maximum likelihood)
  - General advice
    - parameters have widely different size: standardize covariates and take logs of parameters
    - to transform parameters so that an unconstrained optimization: ln(π/(1-π)
    - if have constraints: use Bayesian or quasi-Bayesian solutio
  - deterministic
    - Newton-Raphon
      - NR has maximum efficiency in the best-case scenario (which we expect in large samples), but is fragile in less rosy scenarios
    - quasi-NR (BFGS) !!!
      - method = "BFGS" provides a much better default for most maximum likelihood problems
    - conjugate gradient (CG)
    - Nelder-Mead
  - stochastic (derivative-free, stochastic optimization)
    - gradient based stochastic approximation
    - random search
    - simulated annealing
    - can escape local modes
    - MUCH slower

- Analytical vs numerical derivatives
  - evaluating the log-likelihood numerically for small changes in parameters
  - p or 2p likelihood evaluations
  - Analytical derivatives will make estimation faster and more reliable
  - and the Hessian (and VCV (θ)) much more reliable.
  - but, need time to code, and some dont need hessian
- mixtures of normals and t
- Extensions
  - Regression model where the errors are a mixture of normals insted of normals
  - Each component has di§erent regression parameters
  - πt instead of π, with πt a function of covariates xt and/or an exogenous process.
- fewer and more complex is usually superior (when mistures)\
- Bonds returns are roughly symmetric
- Other returns are asymmetric, common structure
- Computing risk measures by simulation
  - a) mixture of t does a good job at reproducing the empirical VaR,
  - b) Gaussian VaR is OK at 0.95 but increasingly poor further out in the tail,
  - c) ES>>VaR even at 0.999 due to low dof.

8. Modeling conditional volatility

- Volatility clustering: |r| and r^2 autocorrelated and has long memory
- VIX
  - VIX from option prices on the SP500, as an approximate market estimate of the annualized volatility (std) of SP500 returns in the next month
  - Changes in volatility are strongly negatively correlated with SP500 returns
  - extremely skewed, could use log VIX (also skewed)
  - High persistence
  - Can change (especially increase) rapidly
  - has little predictive power for future returns
  - horizom ~= 21
  - standardized returns: simple return/VIX, could assume normal or student-T
  - getting the distribution right can be more important than getting volatility right
  - VIX is a fairly good measure of volatility one month ahead
  - does not predict volatility at long horizons
  - low volatility breeds high volatility
  - we take a bigger position in SP500 when VIX is low, and then reduce the position when VIX increases
- two-stage analysis
- model volatility
  - realized volatility
    - $\sigma_t^2 = \frac{1}{h}\sum_{i=1}^{h} r_{t-i}^2$
    - h=21 for one month
  - Exponential smoothing (exponential moving averages)
    - $\sigma_t^2 = \frac{1}{\sum_{i=1}^{H}\delta^i}\sum_{i=1}^{h}\delta^i r_{t-i}^2$
    - $\sigma_t^2 = \delta\sigma_{t-1}^2 + (1-\delta) r_{t-1}^2$
  - Two-stage analysis with the Risk Metrics approach
    - compute all σ^2
    - compute standardized return $z = (p_{t+1} - p_t)/\sigma_{t+1}$
    - forcast z
    - get forcast σ^2
    - estimate is biased: can can use a MAD estimate to achieve a better fit
- Volatility clustering always generates thicker tails
- GARCH(1,1) model
- Threshold GARCH(1,1)

- that negative returns increase the variance more than positive returns
- negative return use big beta, positive return use small beta
- This gives corr(σt+1,rt) < 0 and generates skeweness in cumulative returns

9. Kelly criterion

- suggests to maximize the expected value of next period log wealth
- f*=2p-1 (detail ★ see paper notes)
- maximize the median terminal wealth (while f=1 to maximize expected return)
- almost certain of increasing your wealth if you have an edg (p>0.5)
- f=p/a-(1-p)/b (win b and loss a)
- Ruin impossible
- $f^* = \frac{\mu}{\sigma^2}$ (detail ★ see paper notes)
- our portfolio volatility is equal to the portfolio Sharpe ratio
- considering rf, f* the same
- half-Kelly
- mean-variance solution near Kelly's
    - The similarity to MV is only if the quadratic approximation holds. If ruin or extremely large losses are possible, the similarity is broken
    - The similarity is therefore most likely to hold when considering short horizons.
    - The Kelly criterion elucidates (point) the role of under-betting and over-betting, and the importance of avoiding ruin. This is typically lost in MV.
    - Most people and organizations cannot say what their λ is
- Much easier to specify the maximum drawdown (loss) they can tolerate.

10. multivariate modeling

- Multivariate normal
    - marginal distributions are also Gaussian
- Multivariate t
- Mixtures of multivariate normals and of multivariate t
- **Local correlations increasing in some periods (good) generates thicker(fatter) tails in portfolio returns**
- **If local correlations increase in bad periods, modest skew**
- A multivariate normal or even a multivariate t would miss this effect and produce too thin portfolio tails
- Factor modeling *
    - $vcv(R_t) = B[E(f_t f_t^T)]B^T + \sigma^2 I$
    - has many less parameters
    - factors may be explanatory variables like market returns, oil prices, GDP growth, or they may be statistical factors
- Risk-Metrics approach *
    - $V_t = \delta V_{t-1} + (1-\delta) R_{t-1} R_{t-1}^T$

11. mean-variance optimization

- $max(w) \quad [E(R_p) - \lambda\sigma^2(R_p)]$
- the w is the same as the approximate Kelly solution for f if λ =0.5
- Both Kelly and MV are very sensitive to inputs
- (Kelly and MV) easy task due to
    - few assets
    - low or modest cross-correlations
    - knowledge of average returns (which is not available in real time)
- Problems with MV in practice
    - quasi(准)-arbitrage (like arbitrage, similar with arbitrage)
        - MV will try to use assets that arbitrage

        - problems more severe with more assets
    - errors in estimating $m_{t|t+1}$ are particularly detrimental to the performance of MV
    - MV will work best with few assets (asset classes), modest(少量) correlations, and good estimates $m_{t|t+1}$ (knowledge of average returns)
- recent reactions to practical problems with MV
    - Equal weight portfolio (1/n)
    - Minimum-variance portfolio
        - Usually applied within an asset class, typically stocks
        - against the idea that risk should be rewarded
            - examples: high volatility != high return
    - Risk-parity portfolio
        - give the same volatility-weight to each asset class
        - target a desired level of portfolio volatility
    - Sorted portfolios
    - Other approaches
- Some relevant R packages *:
    1. fPortfolio. Large package covering many types of portfolio optimizations
    2. FRAPO. Draw-down constraints (and more)
    3. PerformanceAnalytics. Large package for measuring various portfolio statistics (VaR, ES, drawdowns)

12. Bayesian

- The Bayesian approach
    - Randomness is a state of knowledge, not a state of nature.
- base rates
    - we tend to over-weight the evidence p(B|A) and neglect the base rate **p(A)**
    - The frequentist approach also neglects the base rate
    - neglecting the base rate can lead to large errors
        - p(B|A) does not contain a lot information
        - at low signal-to-noise situations
    - In finance, low signal-to-noise is the rule
- Example: updating estimates of managers' skill (cannot understand)
- Bayesian inference(推导) on model parameters
- Forecasting *
- posterior distribution p(θ|y) (后验概率分布)
- Two types of priors (先验)
    - reflecting knowledge coming from experts' judgment, similar situations, or model's simulation
    - reflecting the idea that the final model will be in some form "simpler", in the sense that some elements of θ will be zero or small.
- Three types of inference *
    - available analytically
    - by computer intensive methods known as MCMC (Markov Chain Monte Carlo)
    - approximation
- Model averaging
    - consider more than one model, Bayesian inference taking a weighted combination of all the models
    - don't pick one, but average the results of both.

read: Rational Decision-Making Under Uncertainty

- introduce the coin experiment
- maybe you know Kelly criterion
- Findings: How Well Did Our Players Play
    - having heard of Kelly did not seem to help two of them
    - some take small and constant bet
    - gambler increases the size of his wagers after losses

- How Much Should You Be Willing to Pay to Play
  - depend on the cap and time
- Similarities to Investing in the Stock Market
- Conclusion

read: Superforecasting

- nothing new under the sun
- Superforecasters are not perfect Bayesian updaters but they are better than most
- generate difierent perspectives
- The more a forecaster embraced probabilitistic thinking, the more accurate she was