

THE LONDON QUANTITATIVE FINANCE SEMINAR  
1 MAY 2014



**Foundations of Risk Management and  
Consistency of Risk Measure Estimates**



MARK DAVIS  
Department of Mathematics  
Imperial College London  
`www2.imperial.ac.uk/~mdavis`

Paper: <http://ssrn.com/abstract=2342279>

# AGENDA

## I. Overview of Risk Management

- The VaR vs. CVaR controversy
- Fundamentals of prediction
- External and internal risk measures
- Weather forecasting
- Elicitability

## II. Consistency of risk measure estimates

- Consistent prediction
- Quantile forecasting
- Risk measures involving mean values

## III. Computations

## Financial Risk Management

In this talk losses are positive

Banks have to report various risk management estimates for portfolios of assets, such as

*Value at Risk: VaR*

This is the  $\beta$ -quantile  $q_\beta = F^{-1}(\beta)$  for some conventional confidence level  $\beta$  such as 95% or 99% depending on the application.

NB: in this talk, losses are positive.

*Conditional Value at Risk: CVaR*

Defined by

$$(1) \quad \text{CVaR}_\beta(F) = q_\beta + \frac{1}{1-\beta} \int_{\mathbb{R}} [y - q_\beta]^+ F(dy) = \frac{1}{1-\beta} \int_\beta^1 \text{VaR}_\tau d\tau.$$

*Question:* How can we tell if the values we compute are ‘correct’?

VaR is the tried and tested industry standard, but is criticised on two main grounds

- (i) it takes no account of the magnitude of losses beyond VaR, and
- (ii) it is not a coherent risk measure, implying that diversification does not necessarily reduce risk.

As a result, the Basel Committee is recommending that banks abandon VaR in favour of CVaR. However, there has been a backlash: CVaR in turn has been criticised for

- (i) Instability of computation (Cont, Deguest & Scandolo, QF 2010)
- (ii) Not being ‘elicitable’ (Gneiting, JASA 2010, Ziegel 2013).

*A Revolutionary Suggestion* (Kou, Peng & Heyde, MOR 2013). CVaR is ‘conditional expected loss’. What about ‘conditional median loss’ (CMVaR)?

But

$$\text{CMVaR}_\beta = \text{VaR}_{(1+\beta)/2}$$

so CMVaR gives a reasonable representation of the ‘loss beyond VaR’ at no computational overhead beyond VaR.

## Fundamentals of prediction

In any problem of prediction in time series, the approach taken must depend on the nature of the data, on what it is we are trying to predict and on what we plan to use the prediction for. There is a hierarchy of possibilities.

- (i) Coin tossing.
- (ii) Radioactive emissions.
- (iii) Weather Forecasting
  - Short-term forecasts.
  - Flood barrier design.
- (iv) Car accidents.
- (v) Financial price data
  - Regulatory capital (99% VaR) – “external”.
  - Market risk (95% VaR) (this talk) – “internal”, this talk.

## External and Internal risk measures (Kou, Peng and Heyde, 2013)

**External.** Used for regulation; imposed on all relevant banks.

*Example: Basel II.* Capital charge on day  $t$  is

$$c_t = \max \left\{ \text{VaR}_{t-1}, \frac{k}{60} \sum_{i=1}^{60} \text{VaR}_{t-i} \right\},$$

$k \geq 3$ ,  $\text{VaR}_d = 10$ -day VaR at 99% confidence level calculated on day  $d$ .

Analogous to flood barrier design: ultimate objective is to ensure each bank has an adequate capital cushion. This is an *engineering problem*.

- By definition this is *mark to model*.
- Key technical points relate to robustness and stability of computation.
- More fundamental question: is this a good way to build a ‘flood barrier’?
  - It’s pro-cyclical.
  - It’s expensive: In the UK in 1980 there was 1 regulator per 11,000 employees in the finance industry. In 2011 the figure was 1 in 300. (See A. Haldane ‘The dog and the frisbee’.)

**Internal.** Used within the bank at company-wide and/or trading desk level to monitor risk. Lower significance levels, typically 95%. This means it is feasible to compare predictions with achieved results.

Representative data set: 20 years of weekly values  $S_n$  of the FTSE100 stock index 1994-2013. Associated series of returns  $X_n = (S_n - S_{n-1})/S_{n-1}$  demon-

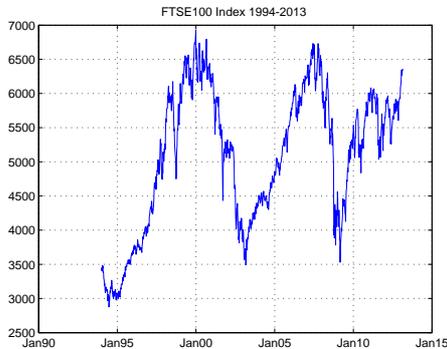


Figure 1: FTSE100 index: weekly values 1994-2013

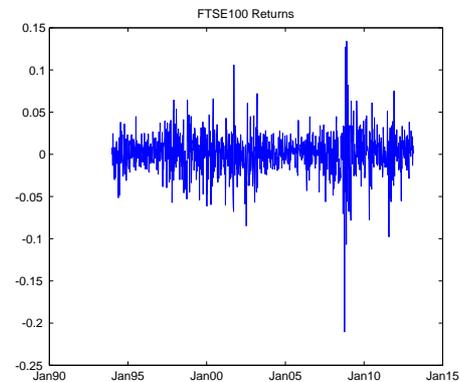


Figure 2: FTSE100 weekly return series.

strates typical stylised features: apparent non-stationarity and highly ‘bursty’ volatility. Empirical distribution has power law tails  $1/x^3$  on both sides.

## Falsifiability

In risk management we're interested in computing the *conditional distribution*  $F$  of future returns, or some statistic  $\mathfrak{s}(F)$  such as a quantile  $q_\beta(F)$ . But in this universe of highly non-stationary data, do the predictive distributions  $F$  implied by statistical models have any meaning at all?

Karl Popper: a statement is *meaningful* if and only if it is *falsifiable*, i.e. evidence could in principle be produced that would show the statement to be false. Equivalently: a statement is meaningless if it can't be tested.

The concept is related to the basic asymmetry between proof and counterexample: to show  $A \Rightarrow B$  we have to show that in *every* case where  $A$  holds,  $B$  holds too, whereas to show  $A \not\Rightarrow B$  we only have to find *one* case where  $A$  holds but  $B$  does not.

Now consider the following statement

Ⓔ: 'The conditional distribution of the FTSE100 return  $X_n$ , given data up to time  $n - 1$ , is  $F$ ', where  $F$  is a specified distribution function.

According to Popper's criterion, statement  $\mathfrak{S}$  is surely meaningless. We compute  $F$  at time  $n - 1$ , and at time  $n$  we get a single number  $X_n = x$ ; so was  $F$  correct?

$\mathfrak{S}$  is falsified at time  $n$  only if  $x$  lies outside the support of  $F$ , which will never be the case in practice, where the support is invariably specified as  $\mathbb{R}$  (or  $\mathbb{R}^+$  for long-only portfolios).

Even if the return sequence is deemed stationary, the *conditional* distributions of subsequent data points  $X_{n+1}, X_{n+2}, \dots$  are all different and provide no evidence about the correctness of  $F$ . In any case *post hoc* data is not germane, since decisions have to be made on the basis of calculations at time  $n - 1$  and history cannot be rewritten afterwards.

Consequently  $\mathfrak{S}$  is not falsifiable.

What is needed here is a shift of perspective. Instead of asking whether our model is correct, we should ask whether our objective in building the model has been achieved.

## Weather forecasting

On day  $i - 1$ , forecaster gives a quantised ‘probability’  $p_i$  of rain on day  $i$ .

The outcome is  $a_i = \mathbf{1}_{(\text{Precipitation}_i \geq 0.5\text{mm})}$ . Example

Probability $p_i$	0.4	0.6	0.3	0.2	0.6	0.3	0.4	0.5	0.6	0.2	0.6	0.4	0.3	0.5
Outcome $a_i$	0	0	1	0	1	0	1	1	1	0	1	0	0	1

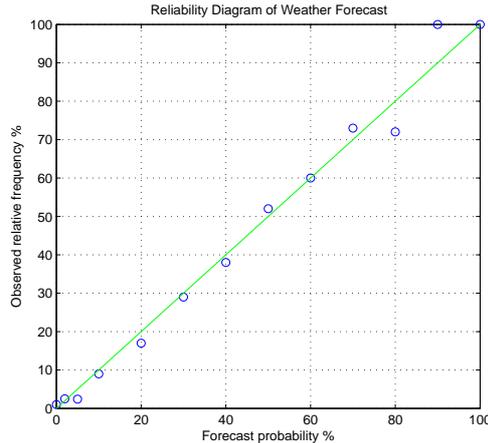
At each value of  $p$  the relative frequency is

$$\bar{a}_p = \frac{\sum_i a(i) \mathbf{1}_{(p_i=p)}}{\sum_i \mathbf{1}_{(p_i=p)}},$$

giving us the calibration table

Probability $p$	0.2	0.3	0.4	0.5	0.6
Relative frequency $\bar{a}_p$	0	0.33	0.33	1	0.75

Here's the reliability diagram for 2820 12-hour forecasts by a single forecaster in Chicago, 1972-1976. (Average  $\sim 200$  forecasts per probability value.)



## APPLICATION TO VALUE AT RISK

Here we want to predict quantiles of the return distribution for an asset or portfolio. This is a slightly different problem:

*Weather forecasting:* Same event “rain”, different forecast probabilities  $p_n$ .

*Risk management:* Same probability  $p = 10\%$ , different events “return  $\geq q_n$ ”.

We have to forecast  $q_n$ .

## Elicitability

Recent papers by Gneiting (JASA 2010) and Ziegel (Math Finance, to appear, arXiv 2013). Original idea: decision-theoretic framework due to L.J. Savage (1971). The ingredients are

- A set  $\mathcal{P}$  of probability measures on  $(\mathbb{R}, \mathcal{B})$ , where  $\mathcal{B}$  is the Borel  $\sigma$ -field.
- A function  $\mathfrak{s} : \mathcal{P} \rightarrow \mathcal{B}$ . Often  $\mathfrak{s}(P)$  is a single-point set, so  $\mathfrak{s}$  is equivalent to a real-valued function  $\hat{\mathfrak{s}} : \mathcal{P} \rightarrow \mathbb{R}$ , a ‘statistic’.
- A score function  $\ell : \mathbb{R} \times \mathbb{R} \rightarrow \mathbb{R}^+$ .

We denote by  $Y$  the canonical random variable  $Y(x) = x$  and by  $F_P$  the distribution of  $Y$  under  $P \in \mathcal{P}$ . The function  $\mathfrak{s}$  is a risk measure such as

$$\begin{aligned}\text{VaR}_\beta(Y) &= -\inf\{x \in \mathbb{R} : F_P(x) \geq \beta\} \\ \text{CVaR}_\beta(Y) &= \frac{1}{1-\beta} \int_\beta^1 \text{VaR}_\tau(Y) d\tau.\end{aligned}$$

The basic question is whether we are able to determine the value of  $\mathfrak{s}(P)$  from observation of the ‘outcomes’  $Y \sim P$ . The score function is interpreted as a penalty  $\ell(x, Y)$  paid when we declare that  $x \in \mathfrak{s}(P)$  and the outcome is  $Y$ .

**Definition 1**  $\ell$  is a consistent scoring function for  $\mathfrak{s}$  if for all  $P \in \mathcal{P}$

$$(2) \quad \mathbb{E}_P[\ell(t, Y)] \leq \mathbb{E}_P[\ell(x, Y)] \quad \forall t \in \mathfrak{s}(P), x \in \mathbb{R}.$$

$\ell$  is strictly consistent if it is consistent and equality in (2) implies  $x \in \mathfrak{s}(P)$ .

**Definition 2**  $\mathfrak{s}$  is elicitable for  $\mathcal{P}$  if there exists a strictly consistent scoring function  $\ell$ .

Gneiting and Ziegel show that VaR (or equivalently, quantiles  $q_\alpha$ ) are elicitable while CVaR is *not* elicitable. Gneiting shows that strictly consistent scoring functions for  $q_\alpha$  take the form

$$\ell(x, Y) = \begin{cases} (1 - \alpha)(g(x) - g(Y)), & Y \leq x \\ -\alpha(g(x) - g(Y)), & Y > x \end{cases}$$

where  $g$  is an increasing function (for example, we could take  $g(x) = x$ .)

*The point:* elicibility is a property that holds across a potentially huge class of data-generating mechanisms  $\mathcal{P}$ .

## Consistent Prediction

We observe a real-valued price series  $X(1), \dots, X(n)$  and an  $\mathbb{R}^r$ -valued series of other data  $H(1), \dots, H(n)$  and wish to compute some statistic relating to the conditional distribution of  $X(n+1)$  given  $\{X(k), H(k), k = 1, \dots, n\}$ . A *statistic* of a distribution  $F$  is some functional of  $F$  such as a quantile or the CVaR. Let  $\mathfrak{s}(F)$  denote the value of this statistic for a candidate distribution function  $F$ . For example, if  $\mathfrak{s}$  is the mean then

$$\mathfrak{s}(F) = \int_{\mathbb{R}} xF(dx), \text{ defined for } F \in \mathcal{P}^1 = \left\{ F : \int_{\mathbb{R}} |x|F(dx) < \infty \right\}.$$

A *model* for the data is a discrete-time stochastic process  $(\tilde{X}(k), \tilde{H}(k))$  defined on a stochastic basis  $(\Omega, \mathcal{F}, (\mathcal{F}_k), \mathbb{P})$ . We always take  $(\Omega, \mathcal{F}, (\mathcal{F}_k))$  to be the canonical space for an  $\mathbb{R}^{1+r}$ -valued process, i.e.  $\Omega = \prod_{k=1}^{\infty} \mathbb{R}_{(k)}^{1+r}$  (where each  $\mathbb{R}_{(k)}^{1+r}$  is a copy of  $\mathbb{R}^{1+r}$ ) equipped with the  $\sigma$ -field  $\mathcal{F}$ , the product  $\sigma$ -field generated by the Borel  $\sigma$ -field in each factor.

For  $\omega \in \Omega$  we write

$$\omega = (\omega_1, \omega_2, \dots) \equiv ((\tilde{X}(1, \omega), \tilde{H}(1, \omega)), (\tilde{X}(2, \omega), \tilde{H}(2, \omega)), \dots).$$

The filtration  $(\mathcal{F}_k)$  is then the natural filtration of the process  $(\tilde{X}(k), \tilde{H}(k))$ . With this set-up, different models amount to different choices of the probability measure  $\mathbb{P}$ . Below we will consider families  $\mathcal{P}$  of probability measures, and we will use the notation  $\mathcal{P} = \{\mathbb{P}^m, m \in \mathfrak{M}\}$ , where  $\mathfrak{M}$  is an arbitrary index set, to identify different elements  $\mathbb{P}^m$  of  $\mathcal{P}$ . The expectation with respect to  $\mathbb{P}^m$  is denoted  $\mathbb{E}^m$ .

**Lemma 1** *Let  $\mathbb{P}^m$  be any probability measure on  $(\Omega, \mathcal{F}, (\mathcal{F}_k))$  as defined above. Then for each  $k \geq 2$  there is a conditional distribution of  $\tilde{X}(k)$  given  $\mathcal{F}_{k-1}$ , i.e. a function  $F_k^m : \mathbb{R} \times \Omega \rightarrow [0, 1]$  such that (i) for a.e.  $\omega$ ,  $F_k(\cdot, \omega)$  is a distribution function on  $\mathbb{R}$  and (ii) for each  $x \in \mathbb{R}$ ,*

$$F_k(x, \omega) = \mathbb{P}^m[X_k \leq x | \mathcal{F}_{k-1}] \quad \text{a.s. } (d\mathbb{P}^m).$$

REMARK: For  $k = 1$  we denote  $F_1^m(x) = \mathbb{P}^m[\tilde{X}(1) \leq x]$ , the unconditional distribution function. □

## Consistency

Consistency is defined for a statistic  $\mathfrak{s}$  relative to a class of models  $\mathcal{P}$ .

Let  $\mathfrak{B}(\mathcal{P})$  denote the set of strictly increasing predictable processes  $(b_n)$  on  $(\Omega, (\mathcal{F}_k))$  such that  $\lim_{n \rightarrow \infty} b_n = \infty$  a.s.  $\forall \mathbb{P}^m \in \mathcal{P}$ ; in this context, ‘predictable’ means that for each  $k$ ,  $b_k$  is  $\mathcal{F}_{k-1}$ -measurable. Often,  $b_k$  will actually be deterministic.

A *calibration function* is a measurable function  $\ell : \mathbb{R}^2 \rightarrow \mathbb{R}$  such that

$$\mathbb{E}^m[\ell(\tilde{X}(k), \mathfrak{s}(F_k^m)) | \mathcal{F}_{k-1}] = 0 \quad \text{for all } \mathbb{P}^m \in \mathcal{P}.$$

**Definition 3** A statistic  $\mathfrak{s}$  is  $(\mathcal{P}, \ell, b)$ -consistent, where  $\mathcal{P}$  is a set of probability measures on  $(\Omega, \mathcal{F})$ ,  $\ell$  is a calibration function, and  $b \in \mathfrak{B}(\mathcal{P})$ , if

$$(3) \quad \lim_{n \rightarrow \infty} \frac{1}{b_n} \sum_{k=1}^n \ell(\tilde{X}(k), \mathfrak{s}(F_k^m)) = 0 \quad \mathbb{P}\text{-a.s. for all } \mathbb{P} \in \mathcal{P}.$$

## Quantile forecasting

Here  $\mathfrak{s}(F) = q_\beta(F)$ , the  $\beta$ -quantile.

Possible choices for  $\ell$  and  $b$  are  $\ell(x, q) = \mathbf{1}_{(-\infty, q]}(x) - \beta$  and  $b_n = n$ , so we examine convergence of

$$\frac{1}{n} \sum_{k=1}^n (\mathbf{1}_{(\tilde{X}(k) \leq q_\beta^k)} - \beta),$$

i.e. we examine the difference between  $\beta$  and the average frequency of times the realized value  $\tilde{X}(k)$  lies below the quantile  $q_\beta^k$  predicted at time  $k - 1$ , over the time interval  $1, \dots, n$ .

The key point is that the criterion only depends on realized values of data and numerical values of predictions; this is the ‘weak prequential principle’ in P. Dawid’s theory of Prequential Statistics. The criterion in elicibility theory does *not* satisfy this principle.

In practice we observe the data sequence  $X(1), \dots, X(n-1)$  and produce an estimate  $\pi(n)$ , based on some algorithm, for what we claim to be  $\mathfrak{s}(F_n)$ . We evaluate the quality of this prediction by calculating

$$J_n(X, \pi) = \frac{1}{b_n} \sum_{k=1}^n \ell(X(k), \pi(n)).$$

Consistency is a ‘reality check’: it says that if  $X_i$  were actually a sample function of some process and we did use the correct predictor  $\pi(i) = \mathfrak{s}(F_i)$  then the loss  $J_n$  will tend to 0 for large  $n$ , and this will be true whatever the model generating  $X(i)$ , within the class  $\mathcal{P}$ , so a small value of  $J_n$  is evidence that our prediction procedure is well-calibrated.

The evidence is strongest when  $\mathcal{P}$  is a huge class of distributions and  $b_n$  is the slowest-diverging sequence that guarantees convergence in (3) for all  $\mathbb{P} \in \mathcal{P}$ .

## Quantile forecasting, continued

The set of models is

$$(\Omega, \mathcal{F}, (\mathcal{F}_k), (\tilde{X}(k), \tilde{H}(k), \mathbb{P}^m), \quad \mathbb{P}^m \in \mathcal{P}$$

where  $\mathcal{P}$  is some class of measures and  $F_k^m(x, \omega)$  is the conditional distribution function of  $\tilde{X}_k$  given  $\mathcal{F}_{k-1}$  under measure  $\mathbb{P}^m \in \mathcal{P}$ . Let  $\mathfrak{P}$  be the set of all probability measures on  $(\Omega, \mathcal{F})$ , and define

$$\mathcal{P}^0 = \{\mathbb{P}^m \in \mathfrak{P} : \forall k, F_k^m(x, \omega) \text{ is continuous in } x \text{ for almost all } \omega \in \Omega\}.$$

For risk management applications, the continuity restriction is of no significance; no risk management model would ever predict positive probability for *specific values* of future prices. So  $\mathcal{P}^0$  is the biggest relevant subset of  $\mathfrak{P}$ .

**Proposition 1** Suppose  $\mathbb{P}^m \in \mathcal{P}^0$ . Then the random variables  $U_k = F_k^m(\tilde{X}_k)$ ,  $k = 1, 2, \dots$  are i.i.d. with uniform distribution  $U[0, 1]$ .

*Proof.* By continuity,  $\mathbb{P}^m[U_1 \leq u_1] = \mathbb{P}^m[\tilde{X}_1 \leq (F_1^m)^{-1}(u_1)] = u_1$ , so  $U_1 \sim U[0, 1]$ . Similarly,  $U_k \sim U[0, 1]$  for each  $k > 1$ . Now suppose that  $U_1, \dots, U_n$  are independent for some  $n$ . Then

$$\begin{aligned} \mathbb{P}^m[U_i \leq u_i, i = 1, \dots, n+1] &= \mathbb{E}^m \left[ \left( \prod_{i=1}^n \mathbf{1}_{(U_i \leq u_i)} \right) \mathbb{P}^m[U_{n+1} \leq u_{n+1} | \mathcal{F}_n] \right] \\ &= \mathbb{E}^m \left[ \left( \prod_{i=1}^n \mathbf{1}_{(U_i \leq u_i)} \right) \right] u_{n+1} \\ &= \prod_{i=1}^{n+1} u_i. \end{aligned}$$

Thus all finite-dimensional distributions of  $(U_i)$  are i.i.d.  $U[0, 1]$ . □

This result is used by Diebold, Gunther and Tay (Int. Econ. Rev. 98) in a different way to the application here.

For  $\beta \in (0, 1)$  let  $q_k^m$  denote the  $\beta$ 'th quantile of  $F_k^m$ , i.e.  $q_k^m = \inf\{x : F_k^m(x) \geq \beta\}$ .  $q_k^m$  is of course an  $\mathcal{F}_{k-1}$ -measurable random variable for each  $k > 0$ .

**Theorem 1** *For each  $\mathbb{P}^m \in \mathcal{P}^0$ , for any sequence  $b_n \in \mathfrak{B}(\mathcal{P})$ ,*

$$(4) \quad \frac{1}{b_n} \frac{1}{n^{1/2}(\log \log n)^{1/2}} \sum_{k=1}^n (\mathbf{1}_{(X_k \leq q_k^m)} - \beta) \rightarrow 0 \quad \text{a.s. } (\mathbb{P}^n)$$

*Thus the quantile statistic  $\mathfrak{s}(F) = q_\beta$  is  $(\mathcal{P}^0, \ell, b')$ -consistent in accordance with Definition 3, where  $l(x, q) = \mathbf{1}_{(x \leq q)} - \beta$  and  $b'_k = b_k(k \log \log k)^{1/2}$ .*

*Proof.* By monotonicity of the distribution function,

$$(X_k \leq q_k^m) \Leftrightarrow (U_k \leq F_k^m(q_k^m)) \Leftrightarrow (U_k \leq \beta).$$

The result now follows from Proposition 1 and by applying the Law of the Iterated Logarithm (LIL) to the sequence of random variables  $Y_k = \mathbf{1}_{(U_k \leq \beta)} - \beta$ , which are i.i.d with mean 0 and variance  $\beta(1 - \beta)$ .

Indeed, define

$$\zeta(n) = \frac{1}{\sigma(2n \log \log n)^{1/2}} \sum_{k=1}^n (\mathbf{1}_{(U_k \leq \beta)} - \beta)$$

where  $\sigma = \sqrt{\beta(1 - \beta)}$ . Then the LIL asserts that, almost surely,

$$\limsup_{n \rightarrow \infty} \zeta(n) = 1, \quad \liminf_{n \rightarrow \infty} \zeta(n) = -1.$$

The convergence in (4) follows. □

Of course, if convergence holds in (4) then it also holds if we replace the sequence  $b$  by  $b''$  such that  $b''_n \geq b_n$  for all  $n$ . In particular, the conventional relative frequency measure

$$(5) \quad \frac{1}{n} \sum_{k=1}^n (\mathbf{1}_{(X_k \leq q_k^m)} - \beta)$$

converges under the same conditions; this also follows directly from the Strong Law of Large Numbers (SLLN).

The striking thing about Theorem 1 is that consistency of quantile forecasting is obtained under essentially *no conditions* on the mechanism generating the data. As we shall see below, we cannot expect any such strong result in estimating other risk measures.

Theorem 1 is a ‘theoretical’ result in that (4) is a tail property, unaffected by any initial segment of the data. Nonetheless, it is practically relevant to compute the relative frequency (5), as we show later.

As a further practical matter, it might be advantageous to replace computation of (5) by statistical tests of the finite-sample hypothesis that the random variables  $Y(1), \dots, Y(n)$  defined above are i.i.d.

## Risk Measures Involving Mean Values

Risk measures such as CVaR involve integration with respect to the conditional distribution functions  $F_k^m$ . In this section we will consider the straight prediction problem of estimating the conditional means

$$(6) \quad \mu_k^m = \int_{\mathbb{R}} x F_k^m(dx).$$

We must assume that the class of candidate models is at most

$$\mathcal{P}^1 = \left\{ \mathbb{P}^m \in \mathfrak{P} : \forall k, \int_{\mathbb{R}} |x| F_k^m(dx) < \infty \right\}.$$

In fact, this problem is general enough to include risk measures of the form  $\int f(x) F_k^m(dx)$  for general functions  $f$ : we can simply define a new model class  $(\tilde{X}', \tilde{H}')$  where  $\tilde{X}'(k) = f(X(k))$  and  $\tilde{H}'(k) = (X(k), H(k))$ . Some modification is required when  $f$  is an option-like function such as  $f(x) = (x - K)^+$  since then  $f(\tilde{X}(k)) = 0$  with positive probability for some measures  $\mathbb{P}^m$ , so these measures are no longer in the class  $\mathcal{P}^0$  as previously defined.

## Universality

The first question to ask is whether we can get any ‘universal’ result, similar to Theorem 1, for estimating  $\mu_k^m$ , by using the i.i.d. sequence  $U_k$ . The answer appears to be no. What makes Theorem 1 work is the equality

$$\mathbf{1}_{(\tilde{X}(k) \leq q_k^n)} - \beta = \mathbf{1}_{(U(k) \leq \beta)} - \beta,$$

so by transforming the variables we obtain the universal calibration function  $l(u, \beta) = \mathbf{1}_{(u \leq \beta)} - \beta$ . In the case of expected value prediction the natural criterion is

$$\frac{1}{n} \sum_{k=1}^n (\tilde{X}(k) - \mu_k^m) \rightarrow 0.$$

Mapping the terms in the  $k$ th summand through the distribution function  $F_k^m$  gives us a summand

$$U(k) - F_k^m(\mu_k^m).$$

This translates into a universal calibration function if and only if there is a constant  $c$  such that

$$(7) \quad F_k^m(\mu_k^m) = c \quad \text{a.s. for all } \mathbb{P}^m,$$

meaning that  $\mu_k^m$  coincides with a *fixed quantile*  $c$  of  $F_k^m$ . But if that is the case the problem reduces to quantile estimation and the results of Theorem 1 apply. The only natural example of this is the situation where each distribution function  $F_k^m$  is symmetric around its mean value, when (7) holds with  $c = \frac{1}{2}$ , but this is totally unrealistic for risk management.

## Martingale analysis

To proceed further, we need to make use of martingale properties. If we define

$$(8) \quad Y(k) = \tilde{X}(k) - \mu_k^n, \quad S(n) = \sum_{k=1}^n Y(k)$$

with  $S(0) = 0$ , then  $S(n)$  is a zero-mean  $\mathbb{P}^m$ -martingale since  $\mathbb{E}^m[Y(k)|\mathcal{F}_{k-1}] = 0$ . We want to determine calibration conditions by using the SLLN for martingales. In this subject, a key role is played by the *Kronecker Lemma* of real analysis.

**Lemma 2** *Let  $x_n, b_n$  be sequences of numbers such that  $b_n > 0$ ,  $b_n \uparrow \infty$ , and let  $u_n = \sum_{k=1}^n x_k/b_k$ . If  $u_n \rightarrow u_\infty$  for some finite  $u_\infty$  then*

$$\lim_{n \rightarrow \infty} \frac{1}{b_n} \sum_{k=1}^n x_k = 0.$$

The *martingale convergence theorem* states that if  $S(n)$  is a zero-mean martingale on a filtered probability space and there is a constant  $K$  such that  $\mathbb{E}|S(n)| \leq K$  for all  $n$ , then  $S(n) \rightarrow S(\infty)$  a.s. where  $S(\infty)$  is a random variable such that  $\mathbb{E}|S_\infty| < \infty$ .

Now let  $Y(k), S(k)$  be as defined at (8) above, and let  $Z(k)$  be a *predictable* process, i.e.  $Z(k)$  is  $\mathcal{F}_{k-1}$ -measurable, such that  $Z(k) > 0$  and  $Z(k) \uparrow \infty$  a.s. Let

$$Y_k^Z = \frac{Y(k)}{Z(k)} \quad \text{and} \quad S^Z(n) = \sum_1^n Y^Z(k).$$

Then  $S_n^Z$  is a martingale (martingale transform, stochastic integral), since

$$\mathbb{E}^m[Y^Z(k)|\mathcal{F}_{k-1}] = \frac{1}{Z(k)} \mathbb{E}^m[Y(k)|\mathcal{F}_{k-1}] = 0.$$

If we can find  $Z(k)$  such that  $\mathbb{E}^m|S^Z(n)| < c_Z^m$  for some constants  $c_Z^m$  then  $S^Z$  converges a.s. ( $\mathbb{P}^m$ ) and hence by the Kronecker lemma

$$\frac{1}{Z(n)}S(n) = \frac{1}{Z(n)} \sum_{k=1}^n (\tilde{X}(k) - \mu_k^n) \rightarrow 0 \quad \text{a.s.}$$

We have shown ...

**Proposition 2** *Under the above conditions, the statistic  $\mathfrak{s}(F) = \int xF(dx)$  is  $(\mathcal{P}^1, \ell, Z)$ -consistent, according to the Definition (3), where  $\ell(x, \mu) = x - \mu$ .*

This Proposition is of course useless as it stands, because no systematic way to specify the norming process  $Z(k)$  has been provided. We can partially resolve this problem by moving to a setting of *square-integrable martingales*. If  $S(n) \in L_2$  we define the ‘angle-brackets’ process  $\langle S \rangle_n$  by

$$\langle S \rangle_n = \sum_1^n \mathbb{E}[Y^2(k) | \mathcal{F}_{k-1}].$$

This is the increasing process component in the Doob decomposition of the submartingale  $S^2(n)$ .

**Proposition 3** *If  $S(n)$  is a square-integrable martingale then  $S(n)/\langle S \rangle_n \rightarrow 0$  on the set  $\{\omega : \langle S \rangle_\infty(\omega) = \infty\}$ .*

This shows that in the square-integrable case we can take  $Z = \langle S \rangle$  in Proposition 2. However, we cannot use  $\langle S \rangle$  as it stands because it does not satisfy the weak prequential principle.

To achieve a calculable norming sequence, we follow a line of reasoning pursued by Hall and Heyde *Martingale Limit Theory and its Application*, relating the predictable quadratic variation  $\langle S \rangle_n$  to the realized quadratic variation

$$Q_n = \sum_{k=1}^n (S(k) - S(k-1))^2 = \sum_{k=1}^n Y^2(k).$$

As Hall and Heyde point out, the two random variables have the same expectation, and we are interested in the ratio  $Q_n/\langle S \rangle_n$ . To get the picture, consider the case where the  $Y(k)$  are i.i.d. with variance  $\sigma^2$ . Then  $\langle S \rangle_n = \sigma^2 n$  and

$$(9) \quad \lim_{n \rightarrow \infty} \frac{Q_n}{\langle S \rangle_n} = \frac{1}{\sigma^2} \lim_{n \rightarrow \infty} \frac{1}{n} \sum_{k=1}^n Y^2(k) = 1 \quad \text{a.s.}$$

by the SLLN. In the general, martingale, case we may or may not have convergence as in (9). We do not go into this here but simply present the following definition.

**Definition 4** Let  $\mathcal{P}^e \subset \mathfrak{P}$  be the set of probability measures  $\mathbb{P}^m$  such that

(i)  $\forall k, \tilde{X}(k) \in L_2(\mathbb{P}^m)$ .

(ii)  $\lim_{n \rightarrow \infty} \langle S \rangle_n = \infty$  a.s. ( $\mathbb{P}^m$ ), where  $S(n)$  is defined at (8).

(iii) There exists  $\epsilon_m > 0$  such that  $Q_n / \langle S \rangle_n > \epsilon_m$  for large  $n$ , a.s.  $\mathbb{P}^m$ .

We can now state our final result.

**Theorem 2** The mean statistic  $\mathfrak{s}(F) = \int xF(dx)$  is  $(\mathcal{P}^e, \ell, Q_n)$ -consistent, where

$$\ell(x, \mu) = x - \mu.$$

*Proof.* Suppose  $\mathbb{P}^m \in \mathcal{P}^e$ . Conditions (i) and (ii) of Definition 4 imply that  $S(n) / \langle S \rangle_n \rightarrow 0$  by Proposition 3. Using condition (iii) we have

$$\left| \frac{S(n)}{Q_n} \right| = \left| \frac{\langle S \rangle_n}{Q_n} \right| \left| \frac{S(n)}{\langle S \rangle_n} \right| \leq \frac{1}{\epsilon_m} \left| \frac{S(n)}{\langle S \rangle_n} \right| \quad \text{for large } n.$$

The result follows. □

## Remarks

(i) Significant conditions must be imposed to secure consistency of mean-type estimates, in contrast to the situation for quantile estimates where almost *no* conditions are imposed.

(ii) Of the conditions in Definition 4, (i) and (ii) are harmless—few will object to assumptions of finite variance and persistent volatility—but condition (iii) is uncheckable and this condition or something equivalent to it cannot be dispensed with.

(iii) This seems to indicate that verifying the validity of mean-based estimates is essentially more problematic the same problem for quantile-based statistics.

(iv) Theorem 1 is a LIL-based result whereas Theorem 2 is based on the SLLN. There is a sizable literature on the LIL for martingales (see Hall & Heyde), but a number of quite intricate conditions are required, none of which would be checkable in the context of mean estimation.

# An algorithm for quantile forecasting

20 years of weekly values  $S_n$  of the FTSE100 stock index 1997-2012.

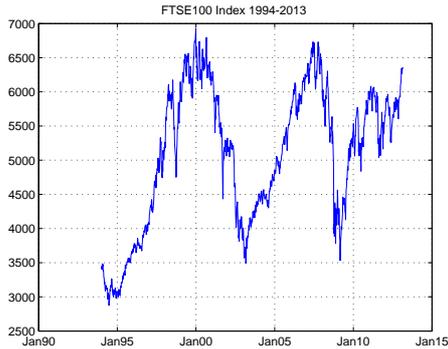


Figure 3: FTSE100 index: weekly values 1994-2013

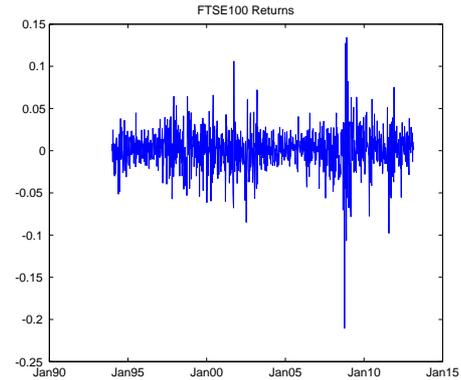


Figure 4: FTSE100 weekly return series.

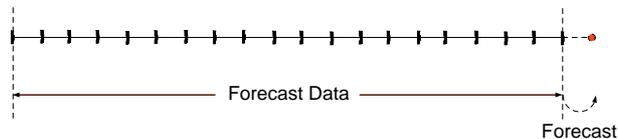
## Computing the quantile forecast

### 1. *Econometrics.*

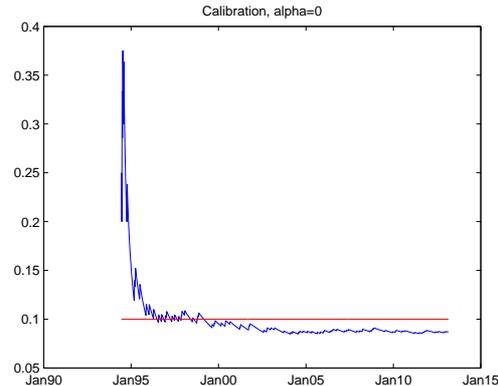
- Choose a model (say, GARCH(1,1))
- Estimate parameters by ML for some window of data.
- Compute conditional 1-week ahead distribution with estimated parameters
- Find 10% upper quantile.

### 2. *Data-driven algorithm*

- Find the 2nd largest of the most recent 20 return values (estimates 10% quantile).
- Use this as the forecast.



... not bad, but slightly miscalibrated.

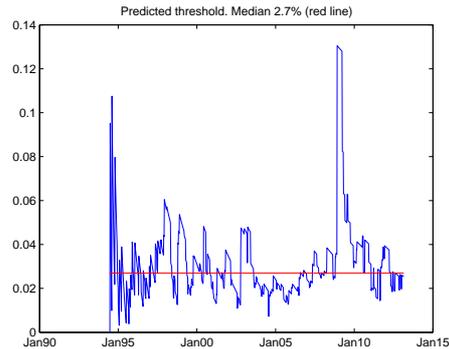
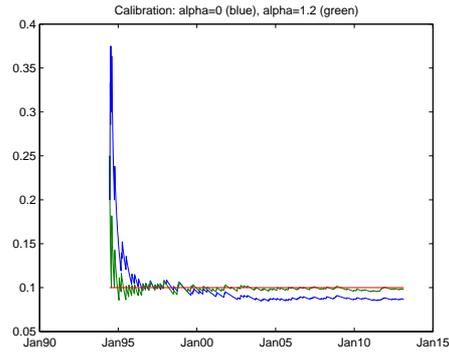


*Remedy:* take 1-week ahead forecast  $\tilde{f}_{n+1}$  given data up to week  $n$  as

$$\tilde{f}_{n+1} = f_n + \alpha(d_n - 0.1)$$

where  $f_n$  is the 20-week estimate as before,  $d_n$  is the observed proportion of above-threshold returns up to time  $n$  and  $\alpha$  is a parameter.

Result—almost perfect calibration. Lower graph shows the sequence of thresholds produced by the algorithm.



# Testing the LIL

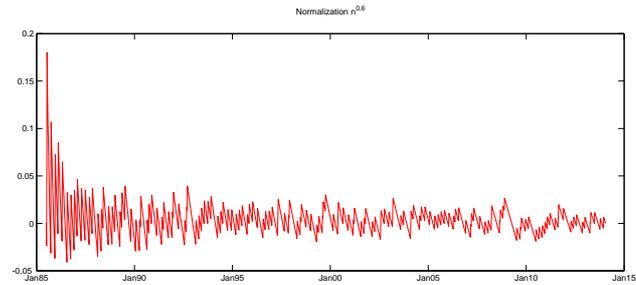


Figure 5: Long data series, normalization  $n^{0.6}$ .

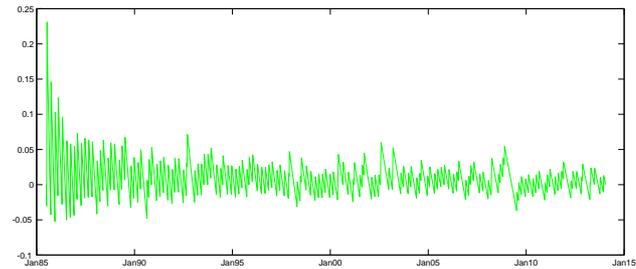


Figure 6: Long data series, normalization  $n^{0.5}$ .

# Running Performance

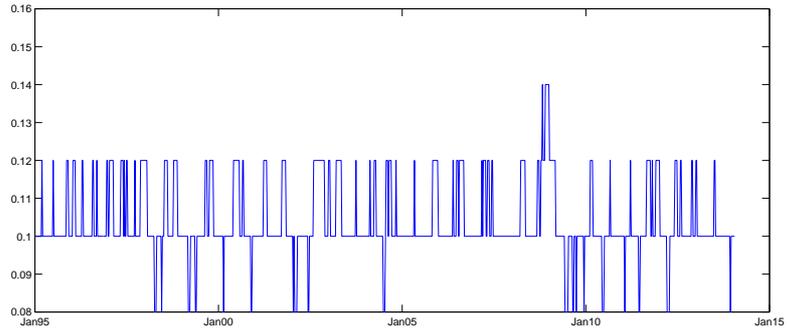


Figure 7: Running 50-week performance of feedback algorithm

## *Statistics*

0.08: 42  
0.10: 744  
0.12: 207  
0.14: 7