Probability, Statistics and their Applications
Aarhus, 15-19 June 2015

■

Prequential Elicitability

■

Mark Davis
Department of Mathematics
Imperial College London
www2.imperial.ac.uk/∼mdavis

Paper: arXiv:1410.4382

AGENDA

- Financial Risk Management
- Elicitability
- Dynamic models
- Prequential Statistics
- Consistency
- Quantile forecasting
- Risk Measures involving Mean Values
- The basic problem with ES.

Financial Risk Management

Banks have to report various risk management estimates for portfolios of assets, based an estimated distribution F of the per-period return Y.
[NB sign convention: losses are positive.]

Primary examples:

Value at Risk VaR

This is the minimum β-quantile $q_\beta = \inf\{y : F(y) \geq \beta\}$ for some conventional confidence level β such as 95% or 99.5% depending on the application.

Expected Shortfall ES

Defined by

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

Question: How can we tell if the values we compute are 'correct'?

When the 10-day period has elapsed, we observe *one number*, the actual portfolio value. Since returns are non-stationary, future data beyond the 10-day horizon provides no (or very little) extra information. Also, *post hoc* information is not germane since decisions are made on the basis of estimates when they are calculated.

Conclusion: 'correctness' can only be evaluated by examining long-run performance.

There has been considerable debate about the relative merits of VaR and ES. 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 ES. However, there has been a backlash: ES in turn has been criticised for

(i) Instability of computation (Cont, Deguest & Scandolo, QF 2010)

(ii) Not being 'elicitable' (Gneiting, JASA 2011, Ziegel 2013).

A Revolutionary Suggestion (Kou, Peng & Heyde, MOR 2013). ES is 'conditional expected loss'. What about 'conditional median loss' (MS)? But

$$\mathrm{MS}_\beta = \mathrm{VaR}_{(1+\beta)/2}$$

so MS (Expected Median Shortfall) gives a reasonable representation of the 'loss beyond VaR' at no computational overhead beyond VaR.

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.

Elicitability

This circle of ideas goes back at least L.J. Savage (1971). The concept itself is due to Osborn and Reichelstein (1985) and the name was coined by Lambert, Pennock and Shoham (2008). See Gneiting (2011) for a wide-ranging exposition.

If $Y \in L_2(\mathbb{R}, \mathcal{B}, \mathbb{P})$ then the function $f(x) = \mathbb{E}[(x-Y)^2]$ achieves its minimum at $x = \mathbb{E}[Y]$ and this is true whatever the distribution F of Y within the L_2 class. Elicitability is concerned with generalizing this characterization of the mean value to other statistics $\mathfrak{s}(F)$ of the distribution function.

For a given statistic $\mathfrak{s}(F)$, can we find a *score function* $S(x, y)$ such that $x \mapsto \mathbb{E}[S(x, Y)] = \int S(x, y)F(dy)$ is minimized at $x = \mathfrak{s}(F)$ for all F in some wide class $\mathcal{F}$ of distributions? In general $\mathfrak{s}(F)$ may be set-valued; for instance the β-*quantile*, $\beta \in [0, 1]$ is the interval $[q_\beta^-, q_\beta^+]$, where q_β^-, q_β^+ are the minimum and maximum values of q such that

$$\mathbb{P}[Y \geq q] \geq 1 - \beta \quad \text{and} \quad \mathbb{P}[Y \leq q] \geq \beta.$$

A score function is a measurable functions $S : \mathbb{R}^2 \to \mathbb{R}$ satisfying

(i) $S(x, y) \geq 0$ with equality if $x = y$

(ii) For each $y \in \mathbb{R}$ the function $x \mapsto S(x, y)$ is continuous, and is continuously differentiable if $x \neq y$.

S is a *consistent score function for a statistic* $\mathfrak{s}$ relative to a class $\mathcal{F}$ of distribution functions F if whenever $Y \sim F \in \mathcal{F}$

$$\mathbb{E}[S(t, Y)] \leq \mathbb{E}[S(x, Y)] \qquad \forall\ t \in \mathfrak{s}(F), x \in \mathbb{R}. \tag{2}$$

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

Definition 1 *A statistic* $\mathfrak{s}$ *is* elicitable for $\mathcal{F}$ *if there exists a strictly consistent score function* S.

Examples (Gneiting, 2011, §3)

1. Mean value. Here $\mathcal{F}$ is L_2 and the score function $S(x,y) = (x-y)^2$ is continuously differentiable. We can characterize optimality by noting that

$$\frac{\partial}{\partial x}\mathbb{E}[S(x,Y)] = \mathbb{E}\left[\frac{\partial}{\partial x}S(x,Y)\right] = x - \mathbb{E}[Y], \tag{3}$$

confirming that the expected score is indeed minimized at the mean value $\mathbb{E}[Y]$. $S = (x-y)^2$ is not the only score function eliciting the mean value.

In general, a function $V(x,y)$ is called an *identification function* for a statistic $\mathfrak{s}$ if

$$\mathbb{E}_F[V(x,Y)] = 0 \qquad \Leftrightarrow \qquad x = \mathfrak{s}(F).$$

2. Quantiles. Here $\mathcal{F}$ is the set of all probability distributions on some interval $I \subset \mathbb{R}$. Then the β-quantile, $\beta \in (0,1)$ is elicitable. If I is compact then a score function S satisfying conditions (i), (ii) above is strictly consistent for the β-quantile if and only if it takes the form

$$S(x,y) = (\mathbf{1}_{(x \geq y)} - \beta)(g(x) - g(y)) \tag{4}$$

where g is a strictly increasing function. Score functions S as in (4) are strictly consistent without the compactness assumption in the class of distributions for which the random variable $g(Y)$ is integrable. An obvious choice is $g(y) = y$, corresponding to $F \in L_1$.

Suppose g is continuously differentiable and let $\mathcal{F}_c$ be the class of continuous distribution functions. Then S is continuously differentiable except at $x = y$ and

$$\frac{\partial S}{\partial x} = g'(x)[\mathbf{1}_{(x \geq y)} - \beta]. \tag{5}$$

Since the event $(Y = x)$ has probability 0 for all $F \in \mathcal{F}_c$ we see that

$$\mathbb{E}\left[\frac{\partial}{\partial x} S(x, Y)\right] = g'(x)[F(x) - \beta], \tag{6}$$

which is equal to zero if and only if x is a β-quantile.

Taking $g(x) = x$ we see that $V(x, y) = \mathbf{1}_{(x \geq y)} - \beta$ is an identification function for the β-quantile.

3. Expectiles. For $\tau \in (0,1)$ and $F \in L_1$ the τ-expectile is the unique solution m_τ to the equation

$$\tau \int_{(x,\infty)} (y-x)F(dy) = (1-\tau) \int_{(-\infty,x)} (x-y)F(dy).$$

If ϕ is a C^1 strictly convex function, the score function

$$S(x,y) = (\tau \mathbf{1}_{(x<y)} + (1-\tau)\mathbf{1}_{(x\geq y)})(\phi(y) - \phi(x) - \phi'(x)(y-x))$$

is strictly consistent for the τ-expectile in the class of F such that Y and $\phi(Y)$ are F-integrable. The natural choice is $\phi(x) = x^2$ when $(\phi(y) - \phi(x) - \phi'(x)(y-x)) = (y-x)^2$. If $\phi \in C^2$ then

$$\frac{\partial S}{\partial x} = \phi''(x)[\tau \mathbf{1}_{(x<y)} + (1-\tau)\mathbf{1}_{(x\geq y)}](x-y), \tag{7}$$

and hence

$$\mathbb{E}\left[\frac{\partial}{\partial x} S(x,Y)\right] = -\phi''(x)\left[\tau \int_{(x,\infty)} (y-x)F(dy) - (1-\tau)\int_{(-\infty,x)} (x-y)F(dy)\right]$$

so that $\mathbb{E}[(\partial S/\partial x)(x,Y)] = 0 \Leftrightarrow x = m_\tau$. This characterization only requires Y to be F-integrable. Note that the mean is the $\frac{1}{2}$-expectile.

VaR vs. ES

The facts are as follows.

(i) VaR is elicitable as long as the quantile is unique (i.e. there is just one x such that $F(x) = \beta$).

(ii) ES is not elicitable. (It fails to have 'convex level sets'.)

(iii) The pair (VaR, ES) is jointly elicitable. This was very recently shown by Fissler & Ziegel.

In view of (iii), objections to ES on the grounds that is is not elicitable fall away. There are other objections, however ..

A possible score function for joint elicitability is

$$S(x, z, y) = (\mathbf{1}_{x \geq y} - \beta)(x - y) + \frac{1}{1 - \beta} e^{-z} \mathbf{1}_{x<y}(y - x) + e^{-z}(x - z - 1).$$

For this function we find that

$$\mathbb{E}\left[\frac{\partial S}{\partial x}(x,z,Y)\right], \qquad \mathbb{E}\left[\frac{\partial S}{\partial z}(x,z,Y)\right]$$

are equal to zero when

$$F(x)=\beta, \qquad z=x+\frac{1}{1-\beta}H(x), \tag{8}$$

where $H(x)=\mathbb{E}[Y-x]^+$, i.e., when $x=q_\beta$ and

$$z=q_\beta+\frac{1}{1-\beta}\int_{q_\beta}^{\infty}(y-q_\beta)F(dy),$$

the β-ES.

Dynamic Models

Suppose we observe not just one r.v. Y with distribution function F but a sequence $Y_1, Y_2, \ldots$, i.e. a discrete-time process, for which we denote by $F_k(y)$ the conditional distribution of Y_k given $Y_1, \ldots, Y_{k-1}$:

$$F_k(y) = \mathbb{P}[Y_k \leq y | Y_1, \ldots, Y_{k-1}].$$

Suppose that, for some class $\mathcal{F}$ of distributions and for all sequences $y_1, y_2, \ldots$

(i) $F_k(\cdot\,; y_1, \ldots, y_{k-1}) \in \mathcal{F}$;

(ii) A given statistic $\mathfrak{s}$ is elictable for $\mathcal{F}$ and there is an identification function V such that $x \in \mathfrak{s}(F) \Leftrightarrow \mathbb{E}_F[V(x, Y)] = 0$.

Then when $x_k = \mathfrak{s}(F_k)$ we have

$$\mathbb{E}[V(x_k, Y_k) | Y_1, \ldots, Y_{k-1}] = 0,$$

i.e. $B_k \stackrel{\Delta}{=} V(x_k, Y_k)$ is a *martingale difference sequence.*

Prequential Statistics (Dawid, JRSS 1984; Dawid and Vovk, Bernoulli 1999)

Combines *probability forecasting* with *sequential prediction*.

Perfect example: 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

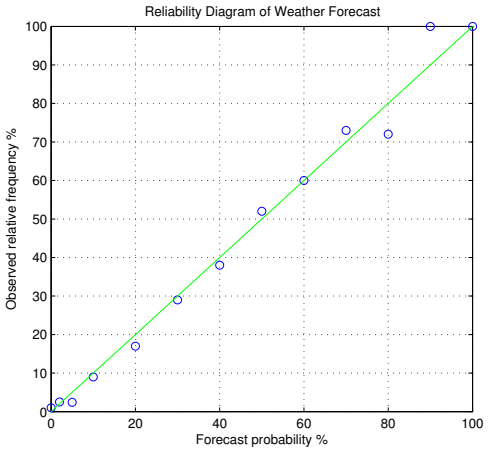
At each value of p the relative frequency is

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

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 produce *forecast thresholds* q_n.

Principles of Prequential Statistics (Dawid & Vovk)

Weak prequential principle: Evaluation of forecasting systems should be based only on the observed data and the numerical values of the forecasts produced (not on the algorithm that produced them).

Strong prequential principle: Criteria for correct prediction should only depend on agreement between Nature and Forecaster on the stochastic law $\mathbb{P}$ generating the data, not on what that law is (within some specified class $\mathcal{P}$).

Consistent Prediction

We observe a real-valued price series $Y(1), \ldots, Y(n)$ and an $\mathbb{R}^r$-valued series of other data $H(1), \ldots, H(n)$ and wish to compute some statistic relating to the conditional distribution of $Y(n+1)$ given $\{Y(k), H(k), k = 1, \ldots, n\}$.

A *model* for the data is a discrete-time stochastic process $(\tilde{Y}(k), \tilde{H}(k))$ 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. $\Omega = \prod_{k=1}^{\infty} \mathbb{R}^{1+r}_{(k)}$ (where each $\mathbb{R}^{1+r}_{(k)}$ is a copy of $\mathbb{R}^{1+r}$) equipped with the σ-field $\mathcal{F}$, the product σ-field generated by the Borel σ-field in each factor.

For $\omega \in \Omega$ we write

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

The filtration $(\mathcal{F}_k)$ is then the natural filtration of the process $(\tilde{Y}(k), \tilde{H}(k))$. Then different models amount to different choices of the probability measure $\mathbb{P}$. When considering families $\mathcal{P}$ of probability measures, we write $\mathcal{P} = \{\mathbb{P}^m, m \in \mathfrak{M}\}$, 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 $\tilde{Y}(k)$ given $\mathcal{F}_{k-1}$, i.e. a function $F_k^m : \mathbb{R} \times \Omega \to [0,1]$ such that (i) for a.e. ω, $F_k(\cdot, \omega)$ is a distribution function on $\mathbb{R}$ and (ii) for each $x \in \mathbb{R}$,*

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

REMARK: For $k = 1$ we denote $F_1^m(x) = \mathbb{P}^m[\tilde{Y}(1) \leq x]$, 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 $\lim_{n\to\infty} b_n = \infty$ 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 $\ell : \mathbb{R}^2 \to \mathbb{R}$ such that

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

Definition 2 *A statistic $\mathfrak{s}$ is* $(\ell, b, \mathcal{P})$-consistent, *where ℓ is a calibration function, $b \in \mathfrak{B}(\mathcal{P})$ and $\mathcal{P}$ is a set of probability measures on $(\Omega, \mathcal{F})$, if*

$$\lim_{n\to\infty} \frac{1}{b_n} \sum_{k=1}^{n} \ell(\tilde{Y}(k), \mathfrak{s}(F_k^m)) = 0 \quad \mathbb{P}-\text{a.s. for all } \mathbb{P} \in \mathcal{P}. \tag{9}$$

In practice we observe the data sequence $Y(1), \ldots, Y(n-1)$ 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

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

Consistency is a 'reality check': it says that if Y_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 $Y(i)$, 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 (9) for all $\mathbb{P} \in \mathcal{P}$.

Quantile forecasting

For $\beta \in [0,1]$ the β-quantile is the set

$$q_\beta = \{x : \mathbb{P}[Y \leq x] \geq \beta \text{ and } \mathbb{P}[Y \geq x] \geq 1-\beta\}.$$

The β-VaR is defined as $\text{VaR}_\beta = q_\beta^- = \inf\{x : x \in q_\beta\}$. The set of models is

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

where $\mathcal{P}$ is some class of measures and $F_k^m(x,\omega)$ is the conditional distribution function of $\tilde{Y}_k$ 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 $U_k = F_k^m(\tilde{Y}_k)$, $k = 1, 2, \ldots$ are i.i.d. with uniform distribution $U[0,1]$.*

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

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

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 β-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 $b_n \in \mathfrak{B}(\mathcal{P})$,*

$$\frac{1}{b_n}\frac{1}{n^{1/2}(\log\log n)^{1/2}}\sum_{k=1}^{n}(\mathbf{1}_{(Y_k \leq q_k^m)} - \beta) \to 0 \quad \text{a.s.}\,(\mathbb{P}^n) \tag{10}$$

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

Proof. By monotonicity of the distribution function, $(Y_k \leq q_k^m) \Leftrightarrow (U_k \leq F_k^m(q_k^m)) \Leftrightarrow (U_k \leq \beta)$. 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,

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

The convergence in (10) follows. □

Of course, if convergence holds in (10) 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

$$\frac{1}{n} \sum_{k=1}^{n} (\mathbf{1}_{(Y_k \leq q_k^m)} - \beta) \tag{11}$$

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 (10) is a tail property, unaffected by any initial segment of the data. Nonetheless, it is practically relevant to compute the relative frequency (11).

As a further practical matter, it is advantageous to augment computation of (11) with statistical tests of the finite-sample hypothesis that the random variables $Y(1), \ldots, Y(n)$ defined above are i.i.d.

Risk Measures Involving Mean Values

Risk measures such as ES 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

$$\mu_k^m = \int_{\mathbb{R}} x F_k^m(dx). \tag{12}$$

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 $(\tilde{Y}', \tilde{H}')$ where $\tilde{Y}'(k) = f(Y(k))$ and $\tilde{H}'(k) = (Y(k), H(k))$. Some modification is required when f is an option-like function such as $f(x) = (x-K)^+$ since then $f(\tilde{Y}(k)) = 0$ with positive probability for some measures $\mathbb{P}^m$, so these measures are no longer in the class $\mathcal{P}^0$ as previously defined.

Martingale analysis

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

$$\breve{Y}(k) = \tilde{Y}(k) - \mu_k^n, \qquad S(n) = \sum_{k=1}^{n} \breve{Y}(k) \tag{13}$$

with $S(0) = 0$, then $S(n)$ is a zero-mean $\mathbb{P}^m$-martingale since $\mathbb{E}^m[\breve{Y}(k)|\mathcal{F}_{k-1}] = 0$. 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 $u_n = \sum_{k=1}^{n} x_n/b_n$. If $u_n \to u_\infty$ for some finite u_∞ then*

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

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 $S(n) \to S(\infty)$ a.s. where $S(\infty)$ is a random variable such that $\mathbb{E}|S_\infty| < \infty$.

Now let $\breve{Y}(k), S(k)$ be as defined at (13) 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 $Y_k^Z = \breve{Y}(k)/Z(k)$ and $S^Y(n) = \sum_1^n Y^Z(k)$. Then S_n^Y is a martingale, since

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

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

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

We have shown

Proposition 2 *Under the above conditions, the statistic* $\mathfrak{s}(F) = \int xF(dx)$ *is* $(\ell, Z, \mathcal{P}^1)$*-consistent, according to the Definition* (2)*, 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 $S^2(n)$. The following is standard (Williams, *Probability with Martingales*).

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

Proposition 3 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, which requires that the norming sequence be calculable using only observed data and numerical values of estimates.

To achieve a calculable norming sequence, we follow a line of reasoning pursued by Hall and Heyde's *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 σ^2. Then $\langle S \rangle_n = \sigma^2 n$ and by the SLLN

$$\lim_{n \to \infty} \frac{Q_n}{\langle S \rangle_n} = \frac{1}{\sigma^2} \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} Y^2(k) = 1 \quad \text{a.s.} \tag{14}$$

In the general, martingale, case we may or may not have convergence as in (14). We do not go into this here but simply present the following definition.

Definition 3 *Let $\mathcal{P}^e \subset \mathfrak{P}$ be the set of probability measures $\mathbb{P}^m$ such that*
(i) $\forall k,\ \tilde{Y}(k) \in L_2(\mathbb{P}^m)$.
(ii) $\lim_{n\to\infty}\langle S\rangle_n = \infty$ a.s. $\mathbb{P}^m$, *where $S(n)$ is defined at* (13).
(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 $(l, Q_n, \mathcal{P}^e)$-consistent, where*

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

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

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

The result follows. □

The basic problem with ES (or any mean value) estimation

Recall that

$$\mathrm{ES}_\beta = \frac{1}{1-\beta}\int_\beta^1 q_\tau d\tau.$$

Most financial data exhibits power tails. Consider the following proposition, in which F is supposed to have *exact* power tail with index κ.

Proposition 4 *Let $0 < \beta < \eta < 1$ and F be a distribution function on $\mathbb{R}^+$ such that for $x \geq q_\eta^+$*

$$F(x) = 1 - (1-\eta)\left(\frac{x}{q_\eta}\right)^{-\kappa}$$

where $\kappa > 1$. Then

$$\mathrm{ES}_\beta(F) = \frac{1}{1-\beta}\left(\int_\beta^\eta q_\tau d\tau + \frac{\kappa}{\kappa-1}(1-\eta)q_\eta\right).$$

$$\text{ES}_\beta(F) = \frac{1}{1-\beta}\left(\underbrace{\int_\beta^\eta q_\tau d\tau}_{\text{known}} + \underbrace{\frac{\kappa}{\kappa-1}(1-\eta)}_{\text{unknown}}\, q_\eta\right). \tag{15}$$

For financial data, quantile estimation is something that can be achieved convincingly for significance levels out to 95% at least. Suppose we wish to compute ES_β and can reliably estimate quantiles q_τ for $\tau \leq \eta$ but not beyond η where the data has dried up. Then the first term on the right of (15) and the value of q_η are known, but the result also depends on the value of κ, and

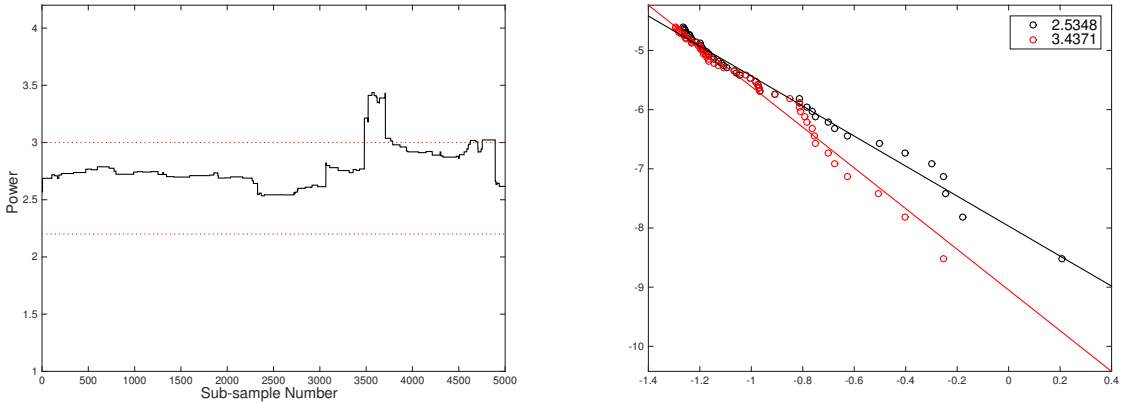
$$\text{ES}_\beta(F) \to +\infty \quad \text{as} \quad \kappa \downarrow 1.$$

To place an upper bound on ES requires a reliable estimate for the tail index κ but by definition this is impossible to obtain.

Conclusion: Any estimate of ES depends on *a priori* assumptions about tail behaviour that cannot be verified on the basis of any finite data set, however large.

Simulated Example

We simulate a 10000-point 'return' series such that all data points have same 99% VaR $q_{0.99}$. Tail index κ switches between 2.2 and 3.0 at jump times of discrete-time Markov chain with $\mathbb{P}[3.0|2.2] = 0.03, \quad \mathbb{P}[2.2|3.0] = 0.01$. Tail beyond $q_{0.99}$ is exactly $1/y^{\kappa}$. Thus the tail index is 3 roughly 3/4 of the time.



Left panel shows empirical tail estimates for 5000-length windows $[i, i+4999]$ for $i = 1, 2, \ldots, 5001$. Right panel shows the empirical tails on a log-log plot for the two windows producing the biggest and smallest tail indices.

ES calculations based on these empirical estimates would be wrong every time and (for this particular data set) underestimates the true ES about 1/4 of the time.

.

Tillykke med fødelsdagen, Ole!

References

Cont, R., R. Deguest, and G. Scandolo (2010). Robustness and sensitivity analysis of risk measurement procedures. *Quantitative Finance* **10**, 593–606.

Davis, M. H. A. (2015). Consistency of internal risk measure estimates. Available at `arxiv.org/abs/1410.4382`.

Dawid, A. P. (1984). Present position and potential developments: some personal views. statistical theory: the prequential approach (with discussion). *J. Roy. Statist. Soc. A* **147**, 278–292.

Dawid, A. P. and V. Vovk (1999). Prequential probability: principles and properties. *Bernoulli* **5**, 125–162.

Fissler, T. and J. F. Ziegel (2015, March). Higher-order elicitability and Osband's principle. Available at `arXiv:1503.08123v1`.

Gneiting, T. (2011). Making and evaluating point forecasts. *J. Amer. Statist. Assoc.* **106**, 746–762.

Hall, P. and C. Heyde (1980). *Martingale Limit Theory and its Application*. Academic Press, NY.

Kou, S., X. Peng, and C. Heyde (2013). External risk measures and Basel accords. *Mathematics of Operations Research* **38**, 393–417.

Ziegel, J. F. (2014). Coherence and elicitability. *Mathematical Finance*. (Published online September 2014).