HPCFinance Conference

New Thinking in Finance

12-14 February 2014

Foundations of Probability Forecasting and Risk Management

Mark Davis

Department of Mathematics

Imperial College London

www2.imperial.ac.uk/∼mdavis

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

AGENDA

- Statistics and Financial Risk Management
- Fundamentals of Prediction
- Weather Forecasting
- Elicitability
- Consistent Prediction
- Quantile Forecasting
- Risk Measures involving Mean Values
- An Algorithm for Quantile Prediction

Financial Risk Management

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

- *VaR: Value at risk:* Essentially, a quantile of the distribution of returns, say 10 days ahead (More precisely, the conditional distribution of 10-day returns given data up to today.)
- *CVaR: Expected loss beyond VaR:* An option value computed according to the above distribution.

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.

Risk Measures

Value at Risk VaR

This is the β-quantile $q_\beta = F^{-1}(\beta)$ for some conventional confidence level β such as 95% or 99.5% depending on the application

Conditional Value at Risk CVaR

Defined by

$$\text{(1)} \qquad \mathrm{CVaR}_\beta(F) = q_\beta + \frac{1}{1-\beta}\int_{\mathbb{R}} [y - q_\beta]^+ F(dy) = \frac{1}{1-\beta}\int_0^1 \mathrm{VaR}_\tau d\tau.$$

There has been renewed debate about the relative merits of VaR and CVaR.

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

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

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 and on what it is we are trying to predict. There is a heirarchy of possibilities.

(i) Coin tossing.

(ii) Radioactive emissions.

(iii) Weather Forecasting

- Flood barrier design.

(iv) Car accidents

(v) Financial price data.

- Regulatory capital (99.7% VaR)
- Market risk (95% VaR) (this paper.)

As a representative data set we will take the series displayed in Figure 1, 20 years of weekly values S_n of the FTSE100 stock index 1994-2013. The

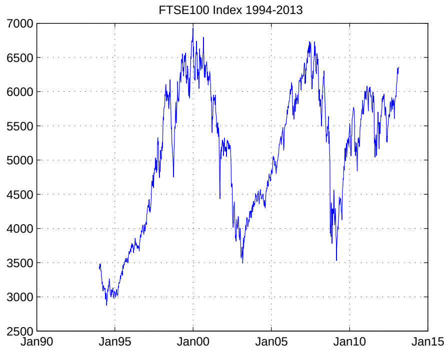


Figure 1: FTSE100 index: weekly values 1994-2013

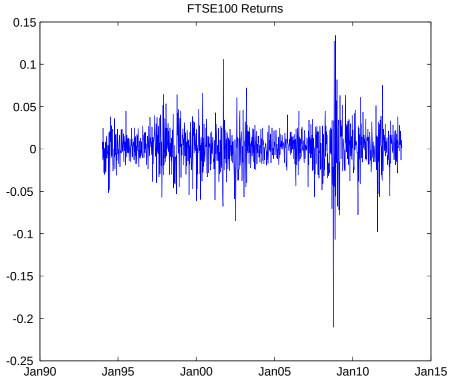


Figure 2: FTSE100 weekly return series.

accompanying Figure 2 shows the associated series of returns $X_n = (S_n - S_{n-1})/S_{n-1}$ and demonstrates the typical stylised features found in financial price data: apparent non-stationarity and highly 'bursty' volatility. The empirical distribution has power law tails $1/x^3$ on both sides.

Analysis of financial series have been the subject of intensive research over at least the last 50 years (e.g. Mandelbrot) and the subject has become mainstream in econometrics and statistics.

However: not so much of this effort has been aimed directly at prediction and when econometric/statistical models have been applied in this direction the results have proved to be poor, as shown by empirical studies.

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 $q_\beta(F)$.

Falisifiability

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

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

$\mathfrak{S}$: '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 one's view is that the return sequence is stationary, the *conditional* distributions of subsequent data points $X_{n+1}, X_{n+2}, \ldots$ are all different and cannot be said to provide any 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

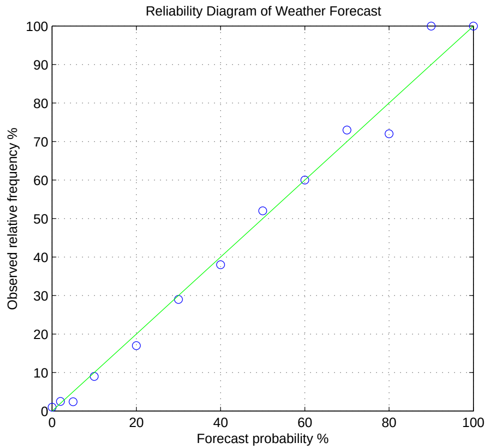
At each value of p the relative frequency is

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

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 (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 σ-field.

– A function $\nu : \mathcal{P} \to \mathcal{B}$. In many cases the values $\nu(P)$ are single-point sets $\{x\}$, so ν is equivalent to a real-valued function $\hat{\nu} : \mathcal{P} \to \mathbb{R}$.

– A score function $s : \mathbb{R} \times \mathbb{R} \to \mathbb{R}^+$.

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 ν is a risk measure such as

$$\mathrm{VaR}_\alpha(Y) = -\inf\{x \in \mathbb{R} : F_P(x) \geq \alpha\}$$

$$\mathrm{CVaR}_\alpha(Y) = (1/\alpha) \int_0^\alpha \mathrm{VaR}_\tau(Y) d\tau.$$

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

Definition 1 *s is a* consistent scoring function for *ν if*

$$\mathbb{E}_P[s(t,Y)] \le \mathbb{E}_P[s(x,Y)] \qquad \forall\, t \in \nu(P), x \in \mathbb{R}. \tag{2}$$

s is strictly consistent *if it is consistent and equality in* (2) *implies $x \in \nu(P)$.*

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

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

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

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

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

$$\mathfrak{s}(F) = \int_{\mathbb{R}} xF(dx), \quad \text{for } F \text{ such that } \int_{\mathbb{R}} |x|F(dx) < \infty.$$

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. $\Omega = \prod_{k=1}^{\infty} \mathbb{R}^{1+r}_{(k)}$ (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

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

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 $F_k^m : \mathbb{R} \times \Omega \to [0,1]$ such that (i) for a.e. ω, $F_k(\cdot, \omega)$ is a distribution function on $\mathbb{R}$ and (ii) for each $x \in \mathbb{R}$,*

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

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

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 $l : \mathbb{R}^2 \to \mathbb{R}$ such that

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

Definition 3 *A statistic* $\mathfrak{s}$ *is* $(l, b, \mathcal{P})$-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*

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

Quantile forecasting

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

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

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

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

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..

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

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

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* $U_k = F_k^m(\tilde{X}_k)$, $k = 1, 2, \ldots$ *are i.i.d. with uniform distribution* $U[0,1]$.

Proof. By continuity, $\mathbb{P}^m[U_1 \leq u_1] = \mathbb{P}^m[\tilde{X}_1 \leq (F_1^m)^{-1}(u_1)] = u_1$, so $U_1 \sim U[0,1]$. Similarly, $U_k \sim U[0,1]$ for each $k > 1$. Now suppose that $U_1, \ldots, U_n$ 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'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 $b_n \in \mathfrak{B}(\mathcal{P})$,*

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

Thus the quantile statistic $\mathfrak{s}(F) = q_\beta$ is $(l, b', \mathcal{P}^0)$-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

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

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 (4) follows. □

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

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

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 $Y(1), \ldots, Y(n)$ 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

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

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 μ_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

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

Mapping the kth term 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

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

meaning that μ_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

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

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 $u_n = \sum_{k=1}^n x_n/b_n$. 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 $Y(k), S(k)$ 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 $Y_k^Z = Y(k)/Z(k)$ and $S^Y(n) = \sum_1^n Y^Z(k)$. Then S_n^Y is a martingale, since

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

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

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

We have shown

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

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* $S(n)/\langle S \rangle_n \to 0$ *on the set* $\{\omega : \langle S \rangle_\infty(\omega) = \infty\}$*.*

Proof (See D. Williams, *Probability with Martingales*.)

Define the martingale

$$W(n) = \sum_{k \le n} \frac{Y(k)}{1 + \langle S \rangle_k},$$

for which

$$\begin{aligned} \mathbb{E}[(W(n) - W(n-1))^2 | \mathcal{F}_{n-1}] &= \frac{1}{(1 + \langle S \rangle_n)^2}(\langle S \rangle_n - \langle S \rangle_{n-1}) \\ &\le \frac{1}{1 + \langle S \rangle_{n-1}} - \frac{1}{1 + \langle S \rangle_n} \quad \text{a.s.} \end{aligned}$$

Thus $\langle W \rangle_\infty \le 1$. From Williams, Theorem 12.13, this implies that $\lim_n W_n$ exists, and hence from the Kronecker lemma that $S(n)/\langle S \rangle_n \to 0$ as long as $\langle S \rangle_n \uparrow \infty$. □

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 *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

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

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) $\lim_{n\to\infty}\langle S\rangle_n = \infty$ 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 $(l, Q_n, \mathcal{P}^e)$-consistent, where*

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

Proof. Suppose $\mathbb{P}^m \in \mathcal{P}^e$. Conditions (i) and (ii) of Definition 4 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. □

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.

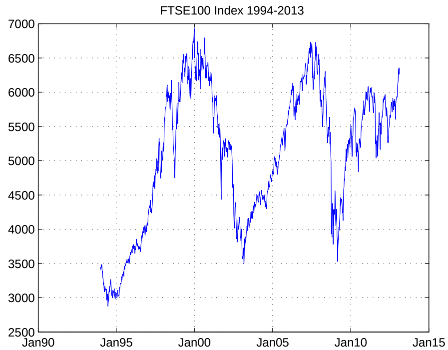


Figure 3: FTSE100 index: weekly values 1994-2013

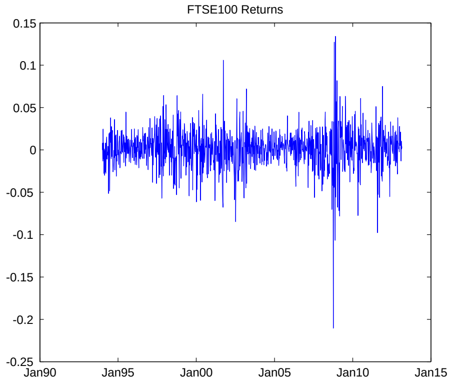


Figure 4: FTSE100 weekly return series.

Computing the quantile forecast

1. *Econometrics.*

   – Choose a model (say, GARCH(1,1))

   – Estimate parameters by ML for some window of data.

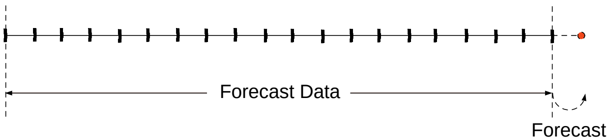
   – Compute conditional 1-week ahead distribution with estimated parameters

   – Find 10% upper quantile.

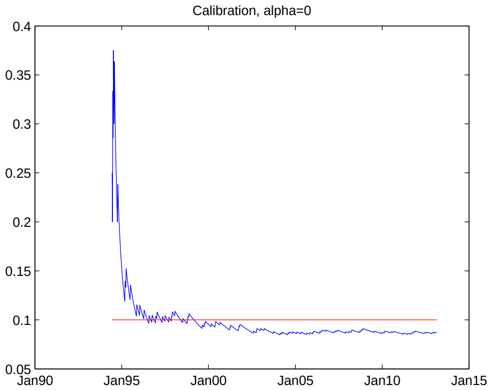
2. *Data-driven algorithm*

   – Find the 2nd largest of the most recent 20 return values (estmates 10% quantile).

   – Use this as the forecast.



... not bad, but slightly miscalibrated.

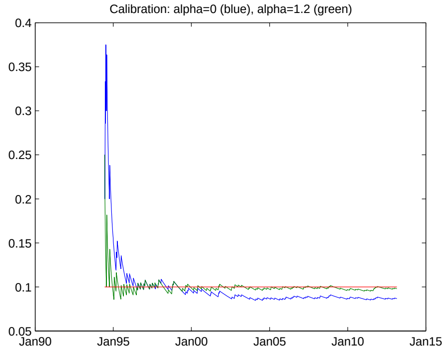


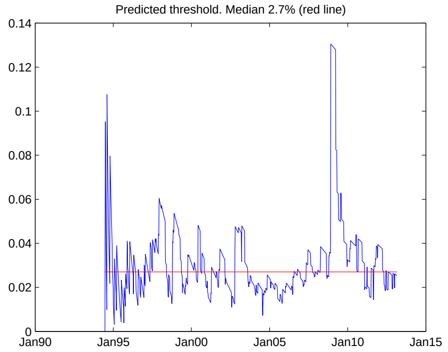
Remedy: take 1-week ahead forecast $\tilde{f}_{n+1}$ given data up to week n as

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

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 α is a parameter.

Result—almost perfect calibration. Lower graph shows the sequence of thesh-olds produced by the algorithm.





Testing the LIL

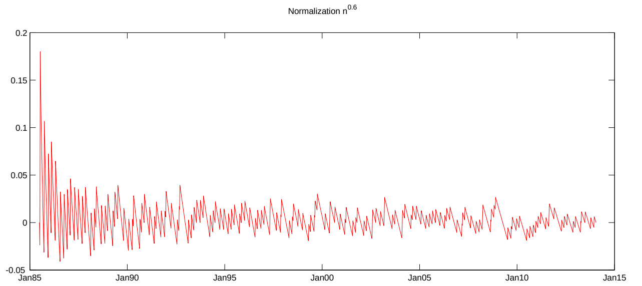


Figure 5: Long data series, normalization $n^{0.6}$.

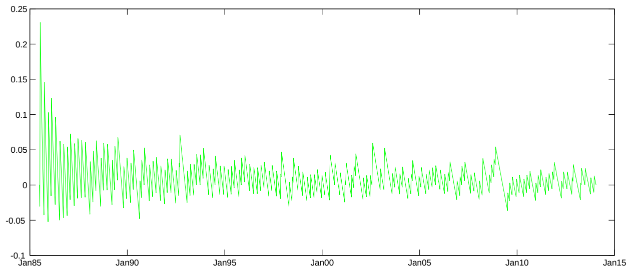


Figure 6: Long data series, normalization $n^{0.5}$.

Running Performance

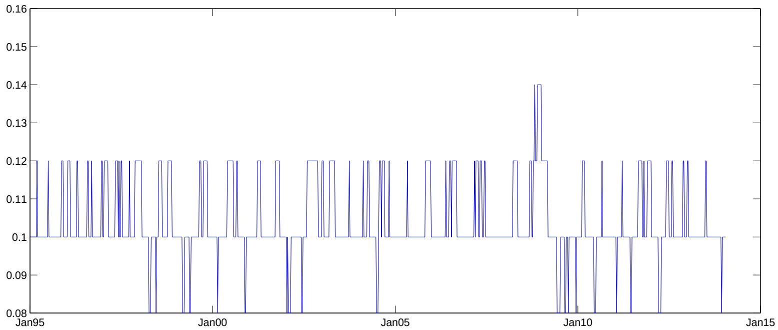


Figure 7: Running 50-week performance of feedback algorithm

Statistics

0.08: 42

0.10: 744

0.12: 207

0.14: 7