CIMAT, Guanajuato, Mexico
5th Workshop on Game-Theoretic Probability
and Related Topics, 13-15 November 2014

■

A Prequential Approach to Financial Risk Management

■

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

Paper: http://arxiv.org/abs/1410.4382

AGENDA

- Financial risk measures: internal and external
- Weather forecasting
- Consistent prediction
- Applying the consistency test
- Quantile forecasting
- Risk measures involving mean values
- Estimating CVaR: an impossibility theorem
- An algorithm for quantile prediction; application to FTSE data
- A test for serial dependence

Financial Risk Management

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. Figure

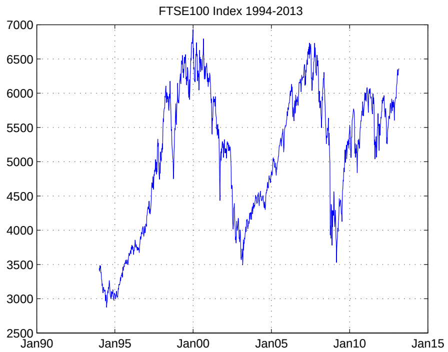


Figure 1: FTSE100 index: weekly values 1994-2013

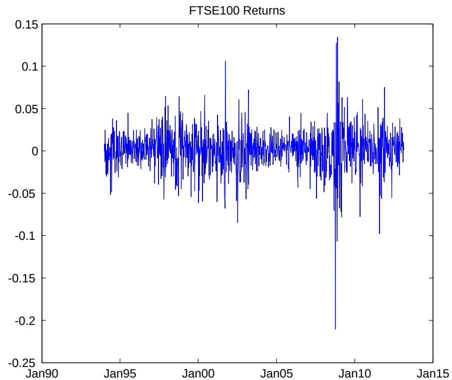


Figure 2: FTSE100 weekly return series.

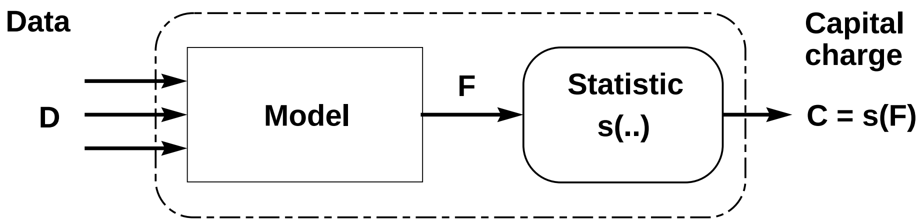
2 shows the associated series of returns $X_n = (S_n - S_{n-1})/S_{n-1}$ and demonstrates the typical stylised features found in financial price data: apparent non-stationarity and highly 'bursty' volatility. The empirical distribution has power law tails $1/x^3$ on both sides.

The problem: In risk management we're interested in computing the *conditional distribution* F_k of returns for the kth period given data up to today (the end of the $(k-1)$th period), or some statistic $\mathfrak{s}(F_k)$ such as a quantile $q_\beta(F_k)$. Each time, we are predicting a *different distribution*, even if the model is stationary. Consequently, no *direct* verification of correctness is possible.

External vs. Internal Risk Measures (Kou, Peng & Heyde MOR 2013)

External risk measures are used for regulatory purposes and imposed on all regulated institutions. Typical confidence level 99.5%, 99.75%.

How do we know if the calculations are correct? We don't—but that's not really the point. (See Cont, Deguest & Scandolo, QF 2010) Ultimate objective is to ensure banks have adequate capital cushion. This is analogous to flood barrier design (but harder).

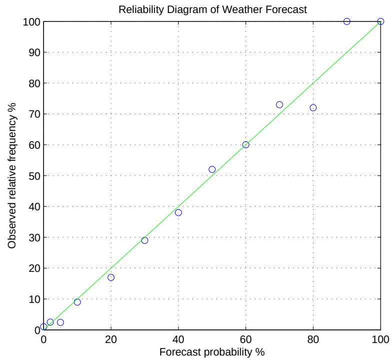


Is this a good structure? See A. Haldane 'The dog and the frisbee' 2012, Keppo, Kofman & Meng, JEDC 2010

Internal risk measures Used within banks to monitor the risks of trading books. Typical confidence level 95%. Here it is possible to compare predictions to outcomes. This talk identifies criteria for 'success'.

Weather Forecasting

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.

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 $x \in \mathbb{R}$,*

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

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 1 *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*

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

Applying the consistency test

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 (1) for all $\mathbb{P} \in \mathcal{P}$.

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

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

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' of Prequential Statistics.

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* $U[0, 1]$.

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

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

Thus the quantile statistic $\mathfrak{s}(F) = q_\beta$ is $(l, b', \mathcal{P}^0)$-consistent in accordance with Definition 1, 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 (2) follows. □

Of course, if convergence holds in (2) 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

$$\frac{1}{n} \sum_{k=1}^{n} (\mathbf{1}_{(X_k \le q_k^m)} - \beta) \tag{3}$$

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

Comments

- The striking thing about Theorem 1 is that consistency of quantile forecasting is obtained under essentially *no* conditions on the mechanism generating the data.
- Theorem 1 is a 'theoretical' result in that (2) is a tail property, unaffected by any initial segment of the data. Nonetheless, it is practically relevant to compute the relative frequency (3), as we show later.
- We can supplement computation of (3) 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 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

$$\mu_k^m = \int_{\mathbb{R}} x F_k^m(dx). \tag{4}$$

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.

Martingale analysis

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

$$Y(k) = \tilde{X}(k) - \mu_k^n, \qquad S(n) = \sum_{k=1}^{n} Y(k) \tag{5}$$

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 $\mathbb{E}|S(n)| \le K$ 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 (5) 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

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

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* (1)*, 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* $\{\omega : \langle S \rangle_\infty(\omega) = \infty\}$*.*

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.

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

$$\lim_{n\to\infty} \frac{Q_n}{\langle S \rangle_n} = \frac{1}{\sigma^2} \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y^2(k) = 1 \quad \text{a.s.} \tag{6}$$

by the SLLN. In the general, martingale, case we may or may not have convergence as in (6). We do not go into this here but simply present the following definition.

Definition 2 *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* (5).

(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 2 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. □

Estimating CVaR

Let F be a distribution function on $\mathbb{R}^+$. Recall from (**??**) that the CVaR at level β can be expressed as

$$\mathrm{CVaR}_\beta(F) = \frac{1}{1-\beta}\int_\beta^1 q_\tau d\tau.$$

where q_τ is the τ-quantile of F. We saw that the empirical distribution of returns for the FTSE100 data set displayed power tails with tail index 2.35 on the left side. It is not claimed that the returns are i.i.d. samples from the same distribution, but nevertheless this fact does add credibility to the idea of considering power-tail distributions as candidates for a model.

An Impossibility Theorem

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*

$$\text{CVaR}_\beta(F) = \frac{1}{1-\beta}\left(\int_\beta^\eta q_\tau d\tau + \frac{\kappa}{\kappa-1}(1-\eta)q_\eta\right). \tag{7}$$

It will be seen in the next section that quantile estimation for financial data is something that can be achieved convincingly for significance levels out to 95% at least. Suppose we wish to compute CVaR_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 (7) and the value of q_η are known, but the result also depends on the value of κ, and $\text{CVaR}_\beta(F) \to +\infty$ as $\kappa \downarrow 1$. To place an upper bound on CVaR requires a reliable estimate for the tail index κ but by definition this is impossible to obtain.

Various expedients

(i) If the empirical return data exhibits power tails, for example the FTSE100 data where the left (=loss) tail index is $\kappa = 2.35$, then use this value beyond the last point where the quantiles can be accurately estimated.

(ii) Use methods based on extreme-value theory.

(iii) Extrapolation: given reliable estimates for q_β and q_η and assuming one is already in the tail regime at q_β one can back out the implied value of κ.

(iv) Cont *et al.* suggest modifying the definition of CVaR_β to

$$\frac{1}{\eta - \beta} \int_\beta^\eta q_\tau d\tau, \quad \text{for some } \eta < 1.$$

This removes the tail problem, at the expense of introducing an arbitrary parameter η.

(v) Kou, Peng & Heyde propose replacing CVaR by CMVaR, the conditional *median* loss beyond VaR. Clearly, $\mathrm{CMVaR}_\beta = \mathrm{VaR}_{(1+\beta)/2}$, so computation reduces to VaR estimation.

Claim: (v) is the winning suggestion: it brings in no unjustifiable assumptions while providing a realistic estimate of the 'loss beyond VaR'.

An algorithm for quantile forecasting

30 years of weekly values S_n of the FTSE100 stock index 1984-2012.

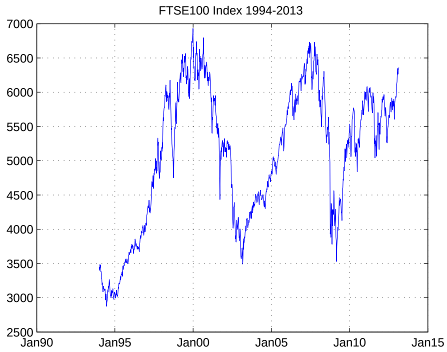


Figure 3: FTSE100 index: weekly values 1994-2013

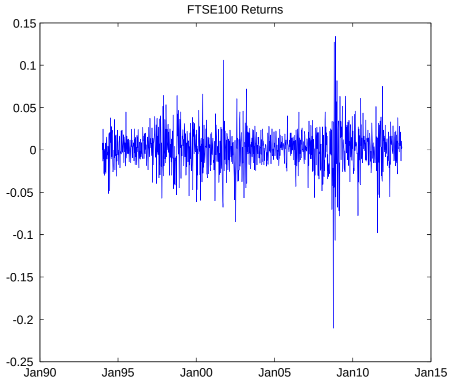


Figure 4: FTSE100 weekly return series.

Computing the quantile forecast

1. *Econometrics.*

   – Choose a model (say, GARCH(1,1))

   – Estimate parameters by ML for some window of data.

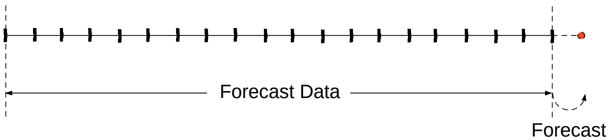
   – Compute conditional 1-week ahead distribution with estimated parameters

   – Find 10% upper quantile.

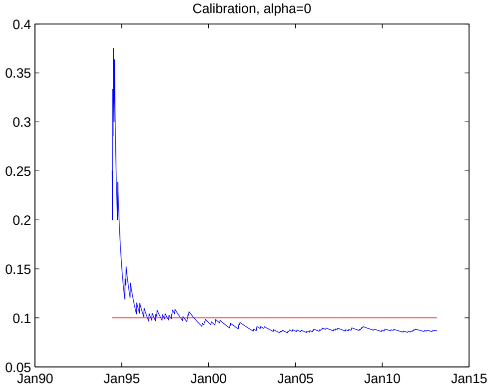
2. *Data-driven algorithm*

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

   – Use this as the forecast.



... not bad, but slightly miscalibrated.

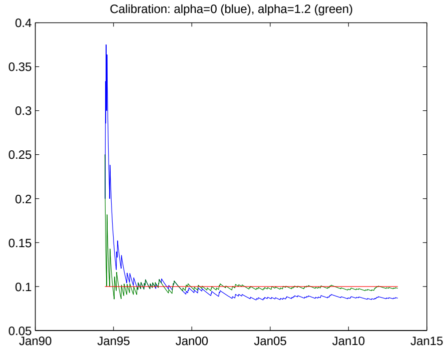


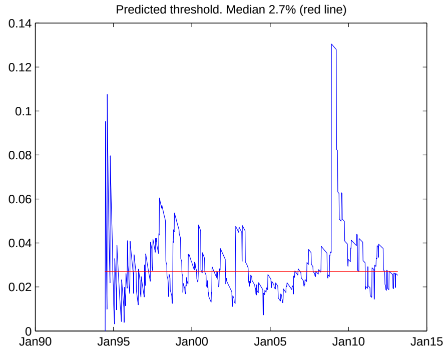
Remedy: take 1-week ahead forecast $\tilde{f}_{n+1}$ given data up to week n as

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

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

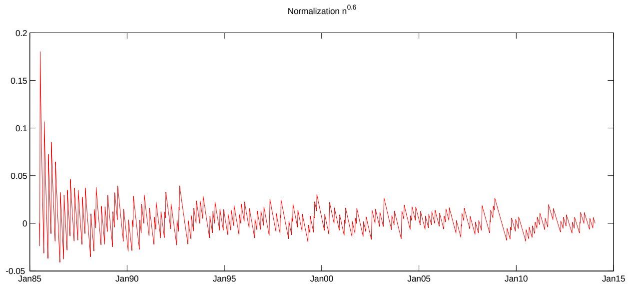


Figure 5: Long data series, normalization $n^{0.6}$.

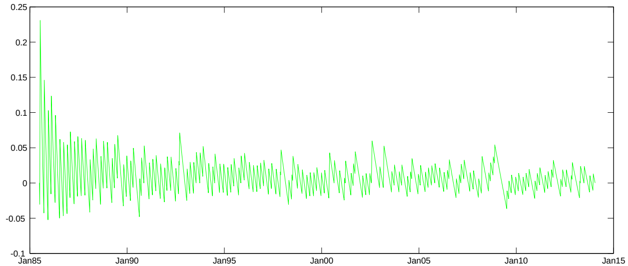


Figure 6: Long data series, normalization $n^{0.5}$.

Running Performance

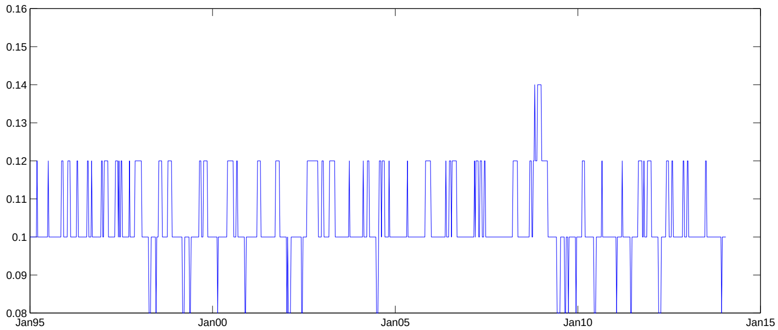


Figure 7: Running 50-week performance of feedback algorithm

Statistics

0.08: 42

0.10: 744

0.12: 207

0.14: 7

A test for serial dependence

Given our prediction algorithm and the data return sequence X_k we generate a sequence $\mathfrak{a} = (a_0, a_1, \ldots)$ of binary r.v. $a_k = \mathbf{1}_{(X_k \le q_k^m)}$. The above tests give confidence that that $\mathfrak{a}$ is consistent with a model in which $\mathbb{P}[a_k = 1] = \beta$. We now want to test the "i" in i.i.d., the hypothesis being

$$\mathfrak{H}_0 : \text{The } a_k \text{ are i.i.d. with } \mathbb{P}[a_k = 1] = \beta.$$

A possible set of alternatives is

$$\mathfrak{H}_{\beta,q} : \begin{array}{l} \mathfrak{a} \text{ is a sample from a 2-state Markov chain} \\ \text{with stationary distribution } \mathbb{P}[a_k = 1] = \beta. \end{array}$$

Under $\mathfrak{H}_{\beta,q}$ the transition probabilities are

$$\begin{aligned} \mathbb{P}[a_0 = 1] &= \beta \\ \mathbb{P}[a_k = 1 | a_{k-1} = 0] &= q \\ \mathbb{P}[a_k = 1 | a_{k-1} = 1] &= q'. \end{aligned}$$

The stationary distribution is β if

$$\begin{aligned}\beta = \mathbb{P}[a_1 = 1] &= \mathbb{P}[a_1 = 1\mathbb{P}[a_1 = 1|a_0 = 0](1-\beta) + \mathbb{P}[a_1 = 1|a_0 = 1]\beta \\ &= q(1-\beta) + q'\beta.\end{aligned}$$

q and q' are related, for given β, by

$$q' = 1 - \frac{1-\beta}{\beta} q,$$

so $\mathfrak{H}_{\beta,q}$ is a 1-parameter family indexed by $q \in [0,1]$ (when $\beta \geq \frac{1}{2}$). The i.i.d. case is $q = q' = \beta$. The log likelihood ratio $\mathrm{LLR}^{\mathrm{n}}_{\mathrm{q}}(\mathfrak{a}) = d\mathbb{P}_{\beta,q}/d\mathbb{P}_0$ is given by

$$\mathrm{LLR}^{\mathrm{n}}_{\mathrm{q}}(\mathfrak{a}) = \mathrm{const} + n_1 \log(1-q) + n_2 \log(1-qf) + (n - n_1 - n_2)\log(q),$$

where $f = (1-\beta)/\beta$ and n_1, n_2 are the numbers of 00, 11 pairs respectively in $\mathfrak{a}$. We denote $\bar{n}_i = n_i/n$, $i = 1, 2$.

Proposition Suppose $\beta \geq \frac{1}{2}$. Then

(i) The maximum likelihood estimate of q is

$$\hat{q}_\beta(\mathfrak{a}) = \frac{1}{2f}\left(1 - \bar{n}_2 + f(1 - \bar{n}_1) - \sqrt{(f - c_1)^2 + 4f(c_1 - c_2)}\right)$$

where $c_1 = 1 - f\bar{n}_1 - \bar{n}_2$, $c_2 = 1 - \bar{n}_1 - \bar{n}_2$.

(ii) The estimator is consistent: under $\mathfrak{H}_{\beta,q}$, as $n \to \infty$

$$\begin{aligned} \bar{n}_1 \to n_1^* &= (1 - q)(1 - \beta) \\ \bar{n}_2 \to n_2^* &= \beta - (1 - \beta)q, \end{aligned}$$

and $\hat{q}_\beta(n_1^*, n_2^*) = q$.

The proof is based on the fact that $Y_k = (a_{k-1}, a_k)$ is an irreducible recurrent 4-state Markov chain.

Note: under $\mathfrak{H}_0$ we have $n_1^* = (1 - \beta)^2, n_2^* = \beta^2$.

Key results for FTSE100 data set, 1500 weeks

90% quantile

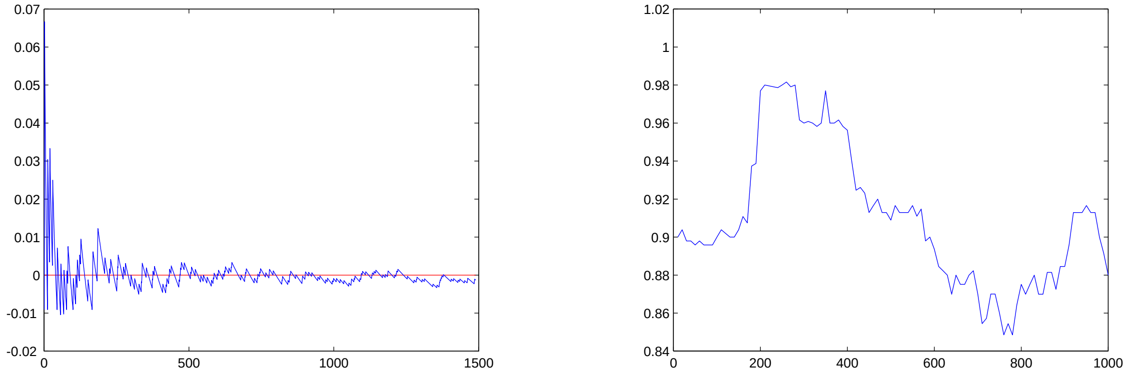
Prob	Data length 250		Data length 500		Data length 1000	
1%	0.7038	1.0000	0.7785	1.0000	0.8201	0.9672
5%	0.7676	1.0000	0.8103	0.9758	0.8418	0.9538
10%	0.7926	1.0000	0.8272	0.9652	0.8519	0.9450
50%	0.8643	0.9437	0.8728	0.9281	0.8823	0.9200

Table 1: Confidence intervals for estimator $\hat{q}_{0.9,0}$.

$$\begin{aligned}\bar{n}_1(1500) &= 0.0100 \\ \bar{n}_2(1500) &= 0.8120 \\ \hat{q}_{0.9}(\bar{n}_1, \bar{n}_2) &= 0.8980.\end{aligned}$$

Theoretical values $((1-\beta)^2, \beta^2) = 0.0100, 0.8100.$

Left panel shows consistency test as before (but centred at 0, not $(1-\beta)$).
Right panel shows $\hat{q}$ estimates using data
$(a_0, \ldots, a_{500}), (a_1, \ldots, a_{501}), \ldots, (a_{1000}, \ldots, a_{1500})$



SLLN

Rolling $\hat{q}$, $n = 500$

Figure 8: 90% threshold

95% quantile

We repeat the tests for the 95% threshold, replacing the previous 90%. The prediction algorithm is the same except that our predicted quantile is now the largest of the previous 20 returns rather than the 2nd largest. Feedback is used in the same way.

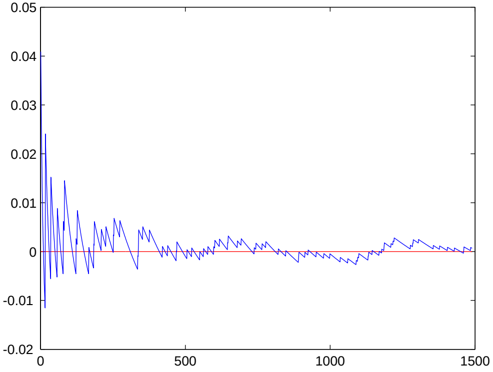
Prob	Data length 250		Data length 500		Data length 1000	
1%	0.6080	1.0000	0.7854	1.0000	0.8516	1.0000
5%	0.7600	1.0000	0.8398	1.0000	0.8800	1.0000
10%	0.8012	1.0000	0.8648	1.0000	0.8940	1.0000
50%	0.9133	1.0000	0.9249	1.0000	0.9308	0.9732

Table 2: Confidence intervals for estimator $\hat{q}_{0.95,0.95}$.

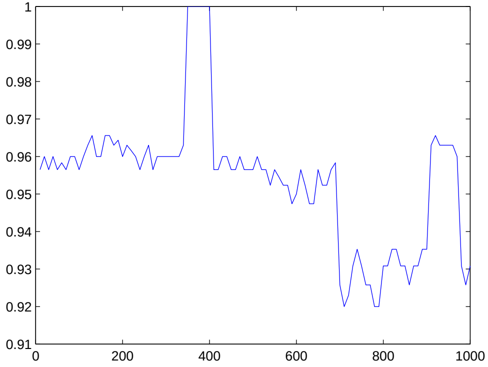
$$\begin{aligned}\bar{n}_1(1500) &= 0.0027\\ \bar{n}_2(1500) &= 0.9007\\ \hat{q}_{0.95}(\bar{n}_1,\bar{n}_2) &= 0.9481.\end{aligned}$$

Theoretical values $((1-\beta)^2, \beta^2) = 0.0025, 0.9025.$

Same tests again ...



SLLN



Rolling $\hat{q}$, $n = 500$.

Figure 9: 95% threshold