



BANK OF ENGLAND

Staff Working Paper No. 673

Borderline: judging the adequacy of return distribution estimation techniques in initial margin models

Melanie Houllier and David Murphy

September 2017

Staff Working Papers describe research in progress by the author(s) and are published to elicit comments and to further debate. Any views expressed are solely those of the author(s) and so cannot be taken to represent those of the Bank of England or to state Bank of England policy. This paper should therefore not be reported as representing the views of the Bank of England or members of the Monetary Policy Committee, Financial Policy Committee or Prudential Regulation Committee.



BANK OF ENGLAND

Staff Working Paper No. 673

Borderline: judging the adequacy of return distribution estimation techniques in initial margin models

Melanie Houllier⁽¹⁾ and David Murphy⁽²⁾

Abstract

The advent of mandatory central clearing for certain types of over-the-counter derivatives and margin requirements for others means that margin is the most important mitigation mechanism for many counterparty credit risks. Initial margin requirements are typically calculated using risk-based margin models, and these models must be tested to ensure that they are prudent. However, two different margin models can calculate substantially different levels of margin yet both pass the usual tests. This paper presents a new approach to parameter selection based on the statistical properties of the worst loss over a margin period of risk estimated by the margin model under test. This measure is related to risk estimated at a fixed confidence interval yet leads to a more powerful test which is better able to justify the choice of parameters used in margin models. The test proposed is used on a variety of volatility estimation techniques applied to a long history of returns of the S&P 500 index. Well known techniques, including exponentially weighted moving average volatility estimation and generalised autoregressive conditional heteroskedasticity approaches are considered, and novel approaches derived from signal processing are also analysed. In each case a range of model parameters which give rise to acceptable risk estimates is identified.

Key words: Conditional volatility, filtered volatility, GARCH(1,1), initial margin model, model backtesting, volatility estimation.

JEL classification: G13, C52, C12.

(1) The London Institute for Banking and Finance. Email: mhoullier@libf.ac.uk. Work conducted while employed at the Bank of England.

(2) Prudential Policy Directorate, Bank of England. Email: david.murphy@bankofengland.co.uk

The views expressed in this paper are those of the authors, and not necessarily those of the Bank of England. The authors would like to thank Paul Bedford, Fernando Cerezetti, Michalis Vasios and the anonymous referees for comments on previous versions of this paper.

Information on the Bank's working paper series can be found at www.bankofengland.co.uk/research/Pages/workingpapers/default.aspx

Publications and Design Team, Bank of England, Threadneedle Street, London, EC2R 8AH
Telephone +44 (0)20 7601 4030 email publications@bankofengland.co.uk

1 Introduction

The collection of margin to reduce counterparty credit risk is a key feature of the post-crisis financial market reforms. This is because policy makers wanted to reduce the direct interconnectedness between financial institutions. Thus:

- the clearing of standardised over-the-counter ('OTC') derivatives between large market participants, with the associated clearing house margin requirements, was mandated by the G-20 [19]; and
- margin requirements are being introduced for bilateral OTC derivatives between many market participants in a revision to the Basel Accord [6].

All of this means that some financial institutions hold tens of billions of pounds of initial margin against their exposures to their derivatives counterparties.¹

The question of how initial margin is calculated is therefore commercially important. Often a *risk-based initial margin model* is used: this typically uses some representation of the portfolio's risk – some set of risk factors – and some information about how those risks have behaved in the past, to determine margin. For instance, a Value-at-Risk-based (VAR-based) margin model might determine an initial margin requirement based on the 99th percentile of the estimated loss distribution of the portfolio in question over an assumed liquidation horizon. The model thus targets a *confidence interval*, determining margin based on portfolio value changes which are as or less probable than this threshold.

1.1 Margin model design

The design of an initial margin model entails making a number of decisions. These can be summarised as:

1. How should the risk of a portfolio be represented?
2. What history of risk factors is to be selected, and how is it to be used?
3. What algorithm should be used to determine the portfolio return distribution?²
4. How are the parameters of that algorithm to be selected?

Thus the design of a simple margin model for portfolios of equity index futures might be based on the decisions:

1. The risk factors are positions in index futures of various maturities;
2. A ten year history of the prices of these futures will be used to calculate returns;
3. A historical simulation ('HS') VAR model will be used;
4. With a ten year window and a 99% confidence interval.

The initial margin model is *both* the algorithm (HS VAR) *and* its calibration (ten year window, 99% confidence interval).

Given the sums at stake, and the potential consequences for the margin taker if there is too little margin, it is vital that robust margin models are used. One obvious requirement is that if a model purports to calculate margin to some degree of confidence, it actually does so. Thus *backtesting* of the calculated margin for a portfolio against the losses it would have experienced over some history of market movements is an important element of model validation, as for instance Berkowitz et al. [8] and Campbell [10] discuss. However, as shown in Gurrola Perez [22],

¹ For instance LCH.Clearnet Group Limited's 2015 consolidated financial statement states that the total margin liability of members at 31st December 2015 was €110 billion.

² There are sometimes two parts to this question: how is the return distribution for a single risk factor calculated, and how are these combined when more than one risk factor is relevant. The latter is the *portfolio margining* question.

the same algorithm with widely different parameters and giving different margin requirements can pass backtesting. Clearly backtesting alone often does not provide enough discriminating power to justify all of the features of an initial margin model nor to fix its parameters within narrow ranges. Figure 1 illustrates this phenomenon, showing the amount of margin which two different filtered historical simulation VAR models would demand for a position in the S&P 500 index. Both models pass the standard (Kupiec [29]) test, yet the amount of margin they demand on the same date can differ substantially. Should the model which has on average higher margin requirements but with lower margin variability (and hence lower liquidity burdens on margin posters) be preferred to the more reactive model?

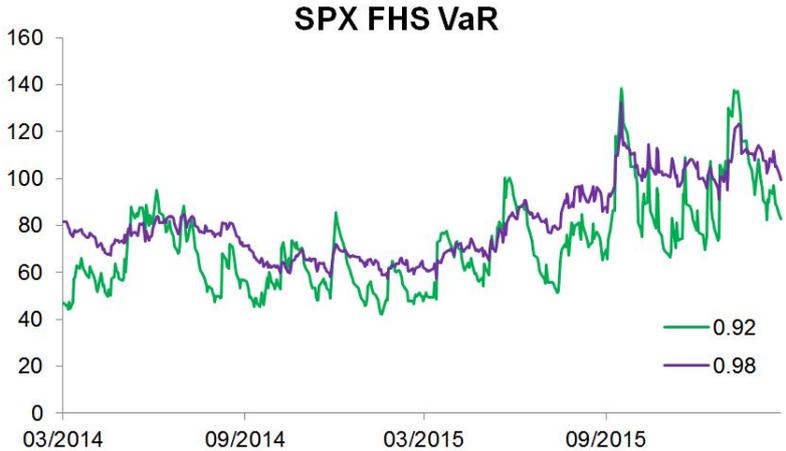


Figure 1: The VAR calculated by two different filtered historical simulation models (one more variable with a decay constant of 0.92; the other less so with a decay constant of 0.98) for a position in the S&P 500 index.

1.2 Sensitivity analysis

The importance of selecting good model parameters is recognised in derivatives policy. For instance, European regulation requires that central counterparties (“CCPs”) carry out *sensitivity analysis* to test model parameters stating:³

Sensitivity analysis shall be performed on a number of actual and representative clearing member portfolios. The representative portfolios shall be chosen based on their sensitivity to the material risk factors and correlations to which the CCP is exposed. Such sensitivity testing and analysis shall be designed to test the key parameters and assumptions of the initial margin model at a number of confidence intervals to determine the sensitivity of the system to errors in the calibration of such parameters and assumptions.

The work presented here is a contribution to the study of initial margin model design and sensitivity analysis. A new technique for comparing models is proposed based on the worst loss that would be experienced liquidating a defaulter’s portfolio over some margin period of risk. This method is used to compare various algorithms used in initial margin modelling, and to select parameters for them which can be justified statistically. Our techniques provides a new and discriminating way of selecting acceptable algorithms and parameters than conventional backtesting approaches: it maps a borderline which encompasses a narrower selection of good models than most techniques currently in use.

³ The quotation is from the EMIR Regulatory Technical Standards [18].

The rest of the paper is structured as follows. In the remainder of this section related work is summarised. Section 2 discusses the model design and sensitivity testing problem. Section 3 shows how the statistical properties of empirical losses can be used for sensitivity testing, and section 4 applies this approach to some volatility estimation algorithms. Both popular techniques such as EWMA volatility estimation and less common approaches such as half-kernel estimators are considered. Section 5 sets out some nuances and extensions to the sensitivity analysis technique proposed, and section 6 concludes.

1.3 Related work

The need for good algorithm and parameter selection in risk modelling has been recognised since such models were first used by banks. Kupiec's early and influential paper [29] recognised the difficulty of this, saying:

It does not appear possible for a bank or its supervisor to reliably verify the accuracy of an institution's internal model loss exposure estimates using standard statistical techniques.

Two broad schools of model testing approach have appeared in response to this observation. Initially the focus was on occasions when losses in excess of the risk predicted by the model at a fixed confidence were observed. The frequency of these *exceptions* form the basis of the test proposed by Kupiec, while later tests also use information on the time between exceptions or their size. The tests proposed by Christoffersen [11], the 'mixed Kupiec' test of Haas [24] and Pelletier & Wei's 'Geometric VaR' [37] are examples of the former, while Lopez's work [33] uses the latter.

Subsequent work proposed tests which focussed on the extent to which the model's forecast return distribution could be said to be accurate, given the empirical quantiles observed. Crnkovic and Drachman [12] and Diebold et al. [14] both suggest approaches of this type, and Berkowitz [7] extended this idea to provide a pass/fail test.

More recently the understanding of risk model testing has been informed by a closer study of data issues and small sample biases: see for instance the work of Escanciano and collaborators [16] and Danielsson & Zhou [13].

2 The model testing problem

The target confidence interval for initial margin models is typically high: for instance, European regulation sets a minimum level at 99% for cleared exchange traded derivatives and 99.5% for cleared OTC derivatives. Margin should be adequate to cover all but the most unlikely moves in portfolio value, making the margin taker fairly safe from counterparty credit risk. However, a high confidence interval can give problems with backtesting as 1 in 100 or 1 in 200 events, by definition, do not happen often. Therefore in some naïve sense, in standard backtesting the model's performance on most days does not matter, and this means that the discriminating power of standard backtesting is often poor.⁴

Some of the approaches to backtesting discussed in section 1.3 ameliorate this problem. But there is another alternative: a risk-based margin model must predict (at least some properties

⁴ The fact that backtests can be performed on many different portfolios helps here: portfolios which depend sensitively on a particular part of the return distribution, or on particular properties of the returns (such as their autocorrelation) should be selected for testing, amongst others. However, the behaviour of financial times series is so rich that identifying all of the relevant behaviours is difficult, and many portfolios will be required, with the associated problem that as the number of portfolios backtested rises, false positives become more likely.

of) the portfolio return distribution each day, so these predictions of a quantity closely related to the risk estimate can be used to evaluate the model. Therefore in this paper the accuracy of a risk model's prediction of (conditional) volatility will be the primary test object.

2.1 Using volatility estimates directly

A simple approach to formalising a test based on volatility estimates is to compare the realised squared return each day to the volatility estimated on the prior day. In particular suppose we have some series of returns (either of single risk factors or of portfolios) $\{r_t\}$. Then, given a prediction of volatility at time t , $\tilde{\sigma}_t$, it should be the case that the observed squared return r_t^2 is proportional to $\tilde{\sigma}_t^2$, so a simple test of the accuracy of the volatility estimator is to look at the coefficient of determination of this relationship. However, this R^2 will typically be low simply because even if the expectation of r_t^2 is $\tilde{\sigma}_t^2$, there is a lot of variation around this average, so this approach will not be very discriminating.⁵

2.2 The margin period of risk

Another complication arises because it is not just one day returns that are typically of interest. Instead margin is usually calculated over some *margin period of risk* ('MPOR') which is longer than a day: two days is the regulatory minimum MPOR for exchange traded derivatives in the house account, five days for cleared OTC derivatives, and ten days for bilateral OTCs in Europe, for instance. Thus potentially ten day returns are of interest, and there are ten times less of them, compounding the problems with test discrimination for one day returns.⁶

In the next section sensitivity analysis technique based on volatility estimates which does not suffer from the two problems identified above will be presented.

3 Using the worst loss over the margin period of risk

Suppose we have a time series of observations of the value of some portfolio x_0, x_1, \dots, x_n at the close of each day. For an m day MPOR, the worst loss assuming default at t , $WL_m(t)$ is defined by

$$WL_m(t) = x_t - \min_{0 \leq u \leq m} x_{t+u}$$

This can be thought of as the worst loss incurred liquidating the portfolio given the last successful margin call was based on prices at t , and a subsequent close out after a default occurred at the close on the worst of the m days in the margin period of risk after t . As such, predicting the worst loss over an appropriate MPOR is the key task of an initial margin model, and hence the accuracy of this prediction is a natural thing to test.

3.1 The worst loss for conditionally lognormal distributions

It is common in risk factor modelling to assume conditional log-normality.⁷ That is, the incremental change in a risk factor x from time t is assumed to be driven by a Brownian motion

⁵ Using 1,000 days of S&P 500 data, the *best* R^2 obtained for an EWMA volatility estimate using a range of decay constants from 0.94 to 0.99 was 0.15. This just shows that the squared return series is too noisy to be used directly to select the parameter(s) of a volatility estimation technique, an observation that goes back at least to Andersen & Bollerslev [2], and which as Poon & Grainger [38] point out, directly relates to the difficulty of estimating conditional kurtosis for fat tailed distributions.

⁶ Non-overlapping returns are preferred to over-lapping ones in the analysis of multi-day MPORs due to the difficulty of interpreting an exception which comes from a single day but which manifests in a number of overlapping MPOR-length returns if overlaps are used.

⁷ For one day returns on some asset classes, even conditional log normality does not capture all of the features of the observed kurtosis: see for instance [31]. However, as margin period of risk lengthens, returns become more

with a volatility σ_t which varies only slowly with t and hence can be assumed constant over the MPOR. Under this assumption, the dependence of $WL_m(t)$ on σ_t has been investigated in the literature. The standard source is Aitsahlia & Lai [1], who consider the related problem of pricing discrete lookback options. They start with a process

$$dx = \mu x dt + \sigma_t x dZ$$

where μ is a constant drift, dZ is a Brownian motion, and σ_t is the conditional volatility at time t . Under the assumption that volatility is constant over the lookback period, it is straightforward from their results to derive the probability density of $WL_m(t)$ over an m day period. For a driftless process it is:

$$\sum_{v=1}^m \alpha_{m-v}(\sigma_t) f_v(x, \sigma_t) \quad (1)$$

where α and f are defined recursively by the convolutions

$$\begin{aligned} \alpha_0 &= 1 & f_1(x, \sigma_t) &= \psi(x, \sigma_t) \\ \alpha_k &= \int_{-\infty}^0 g_k(x, \sigma_t) dx \quad k \geq 1 & f_n(x, \sigma_t) &= \int_0^{\infty} f_{n-1}(y) \psi(x-y, \sigma_t) dy \quad 2 \leq n \leq m \end{aligned}$$

the auxiliary functions g and ψ are defined by

$$\begin{aligned} g_1(x, \sigma_t) &= \psi(x, \sigma_t) \\ g_n(x, \sigma_t) &= \int_{-\infty}^0 g_{n-1}(y, \sigma_t) \psi(x-y, \sigma_t) dy \quad 2 \leq n \leq m \\ \psi(x, \sigma_t) &= \phi(0, \sqrt{m}\sigma_t, x + \frac{m\sigma_t^2}{2}) \end{aligned}$$

and $\phi(\bar{x}, SD, x)$ is the standard normal density function.

Equation (1) allows us to infer how probable a given worst loss k is under these assumptions. In particular the probability of seeing a worst loss no bigger than k , conditional on volatility being $\tilde{\sigma}_t$ over the MPOR can be calculated by integration:

$$\Pr(WL_m < k | \sigma = \tilde{\sigma}_t) = \sum_{v=1}^m \alpha_{m-v}(\tilde{\sigma}_t) \int_0^k f_v(x, \tilde{\sigma}_t) dx \quad (2)$$

It is easiest to illustrate this if we express the worst loss as a multiple of the daily conditional volatility. The MPOR is fixed at ten days and the subscript m is dropped going forward. The cumulative probability of a given worst loss as a fraction of the conditional volatility WL/σ_t is illustrated in Figure 2. The level of conditional volatility σ_t has a minor impact on the shape of this cumulative distribution: the illustration is for a daily volatility of 1%.

3.2 Using the probability of worst losses to compare volatility estimates

Suppose we observe on consecutive days events which we estimate as five, three, four, six and two standard deviation occurrences. We might reasonably conclude that either unusual things are happening a lot or that our estimate of the standard deviation is wrong. If the pattern of these five days repeats for many subsequent periods, the latter conclusion becomes more and more likely, assuming that events on consecutive days are independent. This insight is key to using the cumulative conditional probabilities (2) to test the quality of the prediction $\tilde{\sigma}_t$.

The data needed here are illustrated in Figure 3. A time series of risk factors is used to generate worst losses in each non-overlapping MPOR; a model estimates the conditional volatility

conditionally log-normal. Moreover, many margin models use conditional log-normality to some extent. Hence the assumption is not wholly unreasonable.

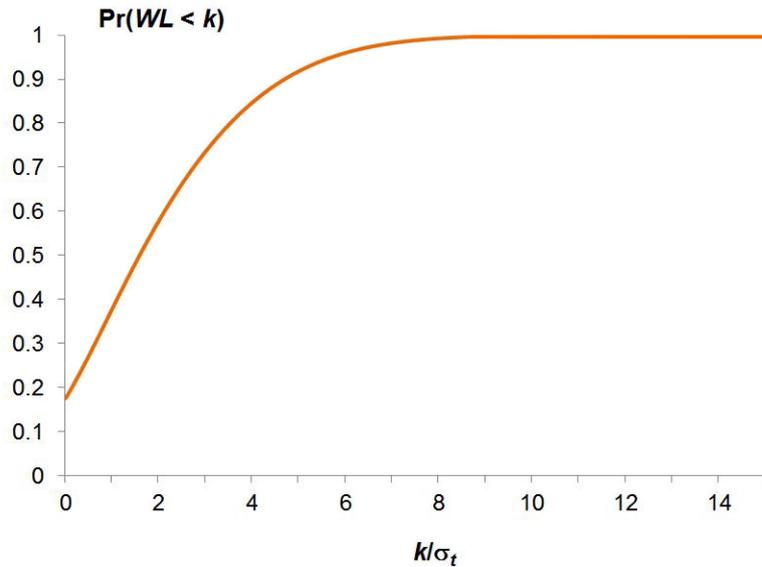


Figure 2: The cumulative probability of seeing a given worst loss for a ten day MPOR, measured as a multiple of conditional volatility. A worst loss of zero happens about 17% of the time, and the 99th percentile worst loss is approximately 7.4 times the conditional volatility.

for the relevant period (based on data up to the start of it); then the cumulative probability of each worst loss conditional on that volatility is calculated. That is, a (likely different) process is estimated for each window, and this process gives a conditional volatility estimate which is used for a MPOR-length period following the window. It is convenient to index these results by the time the window ends t , so $\tilde{\sigma}_t$ is the conditional volatility estimate using the window stretching back from t , and $WL(t)$ is the worst loss for the MPOR $t + 1, \dots, t + MPOR$. Given a sequence of these worst losses $WL(t), WL(t'), \dots$ and conditional volatility estimates $\tilde{\sigma}_t, \tilde{\sigma}_{t'}, \dots$, the collection of all the cumulative probabilities should be uniformly distributed if the volatility estimates are unbiased.

A simple way to test this is to choose a binning scheme for the cumulative probabilities $[0, b_1), [b_1, b_2), \dots [b_n, 1]$ and to use a χ^2 test versus the expected uniform distribution.⁸

Date of estimate	Volatility estimate	Worst loss in subsequent MPOR	Cumulative probability
13/01/1986	0.98%	3.81	0.648
20/01/1986	0.94%	0	0.176
	⋮		
07/03/2016	1.28%	22.8	0.380

Figure 3: The worst loss for non-overlapping ten day MPORs and conditional cumulative probabilities of observing them given a particular estimate of conditional volatility.

Figure 4 illustrates the empirical distribution of cumulative probabilities from the data in Figure 3 and the expected uniform distribution. A big first bucket is used to allow for the fact that $WL = 0$ occurs with expected probability c.17%. It can be seen that for the volatility estimation technique used (in this case, an exponentially weighted moving average ('EWMA'))

⁸ For the standard χ^2 test, the bin widths b_i should be chosen so that the expected population of each bin is at least 5: otherwise slight modifications, such as the use of Fisher's exact test, are needed. See also section 5.2 for a discussion of a more sophisticated approach.

estimate with a decay constant of 0.98), the empirical distribution of cumulative probabilities is close to the expected one.

In contrast, Figures 5 and 6 show the distributions for volatility estimates which have been scaled up and down by 30%.

Clearly if volatility is over-estimated, an excess of low cumulative probability events is observed, while if it is underestimated, an excess of high cumulative probability events are seen. Hence the departure of the empirical distribution from the uniformity expected can be used to test the performance of the volatility estimation technique being considered.

3.3 Example: ARCH volatility

The proposed approach works even when conditional volatility varies significantly over time. So see this, consider an ARCH(1) process: in this well-known model of Engle's [15], the assumption is that detrended daily returns of a risk factor x still follow a discretised version of $dx = \sigma_t x dZ$ but the current volatility σ_t depends on the previous return according to

$$\sigma_t^2 = \omega + \alpha r_{t-1}^2 \quad (3)$$

The returns in this model demonstrate autocorrelation, but this is entirely driven by the volatility process: Z is still white noise. Thus, if the ARCH parameters ω and α are correctly estimated, the probabilities of seeing two worst losses $WL(t), WL(t')$ at two different dates t, t' conditional respectively on σ_t and $\sigma_{t'}$ will be independent of each other, and the theory above will apply. Therefore the worst loss test can be applied to volatility estimates backed out from equation (3). Figure 7 illustrates this: here an ARCH process has been simulated with $\omega = 10^{-5}$ and $\alpha = 0.9$, and ARCH volatility estimates with a range of α s around the true value are tested. Higher values of the test statistic are worse, and models above the dotted line are rejected at 99%. Of course, the test proposed is not the optimal way to calibrate an ARCH model, but the results presented do at least demonstrate that it can pick out model parameters accurately even in the presence of volatility clustering (and thus fat tails).

4 Testing volatility estimation techniques using worst loss

This section presents some of the results obtained using worst loss tests on various techniques used in margin modelling. In all cases we use returns from the S&P 500 index from 3rd January 1984 to 24th March 2016 and calculate margin over a ten day MPOR. This period gives a total of 761 non-overlapping ten day periods.

4.1 Exponentially weighted moving average volatility estimates

Exponentially weighted moving average volatility estimators are widely used in risk models: see Gijbels et al. [20] for an account of their use. In EWMA the current volatility estimate $\tilde{\sigma}_t$ is updated based on the previous estimate $\tilde{\sigma}_{t-1}$ and the previous return r_{t-1} using the recurrence

$$\tilde{\sigma}_t^2 = \lambda \cdot \tilde{\sigma}_{t-1}^2 + (1 - \lambda) \cdot r_{t-1}^2$$

The key issue in designing volatility estimation models of this class is selecting an appropriate decay constant, λ : if λ is too low, the volatility estimate over-reacts to changes in conditions; while if it is too high, it reacts too slowly. Figure 8 show the results from a worst loss based sensitivity analysis of EWMA models with various λ s.

These results suggest that the hypothesis 'the S&P 500 is well-described by a locally-constant volatility which can be estimated using an EWMA technique' is only true for EWMA models with a narrow range of decay constants around 0.98: models with other λ s fails the test, including (at least for the window length studied) the unweighted volatility estimator.

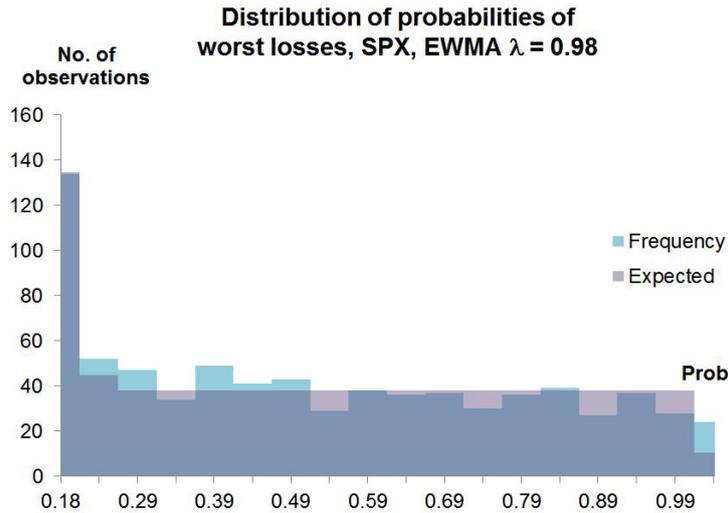


Figure 4: The distribution of conditional cumulative probabilities for ten day worst losses given a particular method for estimating conditional volatility.

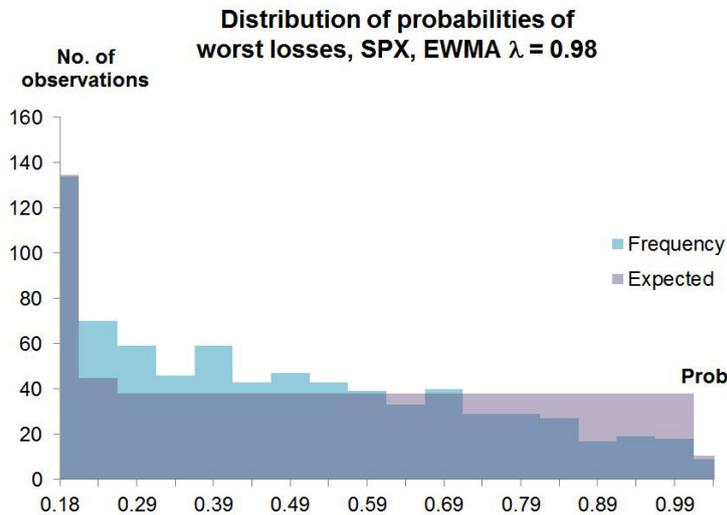


Figure 5: The distribution of conditional cumulative probabilities of ten day worst losses when volatility estimates are scaled up by 30%.

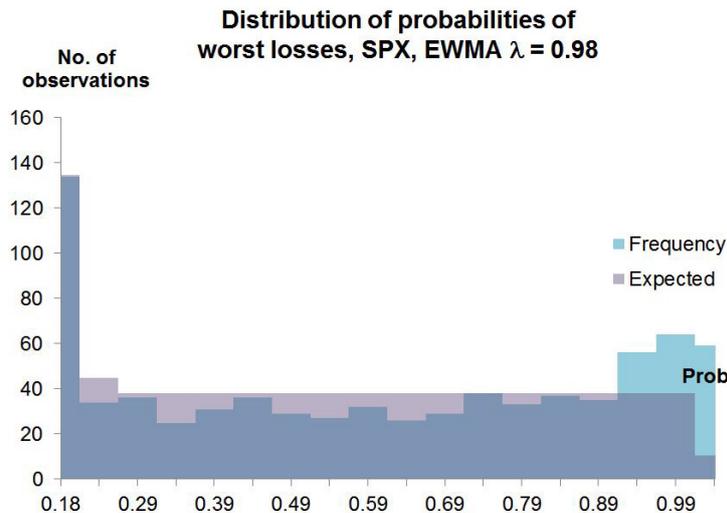


Figure 6: The distribution of conditional cumulative probabilities of ten day worst losses when volatility estimates are scaled down by 30%.

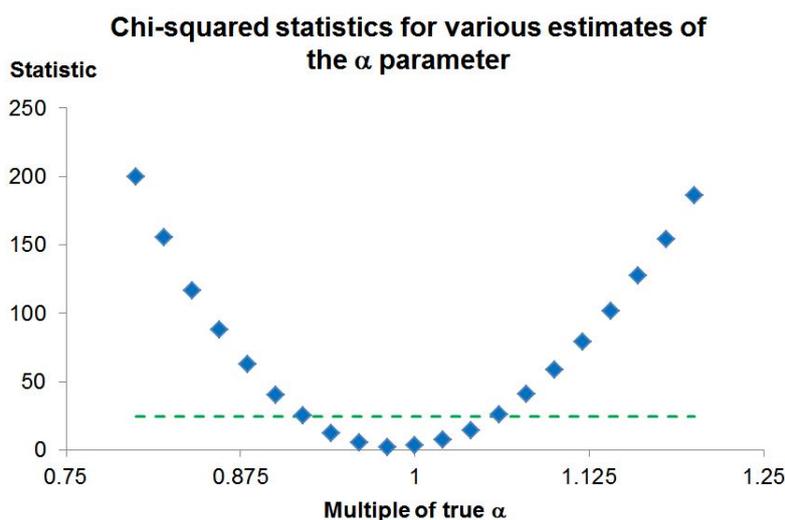


Figure 7: The χ^2 statistic for a range of hypothesised α parameters around the true value for an ARCH process. The dashed green line shows the critical value of the statistic, so models above the line are 99% likely to be wrong.

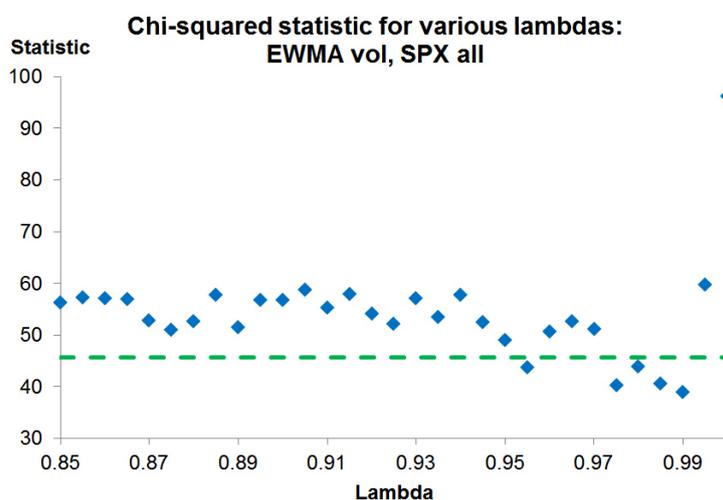


Figure 8: The χ^2 statistic for EWMA volatility estimates on S&P 500 index returns as a function of the decay constant λ .

4.2 Blended volatility estimates

EWMA models with decay constants close to 1 react relatively slowly to new data. Another way to produce a volatility estimate which reacts slowly is to keep a reactive decay constant but to 'blend in' some long term average volatility (such a ten year unweighted volatility estimate) σ_{LT} , by estimating variance at t as

$$\frac{1}{2} \left(\tilde{\sigma}_t^2 + \sigma_{LT}^2 \right)$$

where $\tilde{\sigma}_t$ is an EWMA volatility estimate.

Figure 9 shows the sensitivity analysis for the blended volatility estimates technique.⁹ The

⁹ There is no theoretical justification for this volatility estimator, but it is of practical interest. There is a connection here with the requirement in the EMIR regulatory technical standards that 'the data used for calculating historical volatility capture a full range of market conditions, including periods of stress' [18]. This requirement is sometimes met by 'blending in' a stressed volatility with a current volatility estimate.

impact of blending in a long term average volatility is clearly to tame the reactivity of the smaller decay constants: the resulting volatility estimates are acceptable, or close-to-acceptable, for a wide range of λ s.

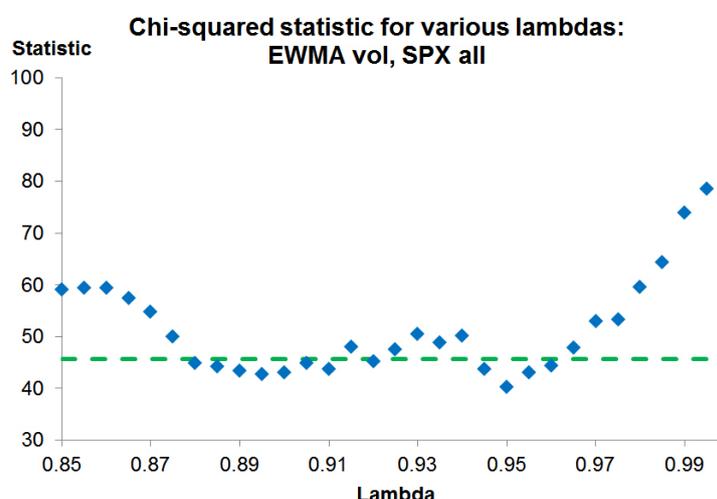


Figure 9: The χ^2 statistic for a blend of an EWMA volatility estimate and long term average volatility.

4.3 The normal half-kernel estimator

Exponential smoothing is a common model-free means of forecasting a future realization of a time series, but it is not the only one. As Stărică [40] points out, the general question is to select some (likely declining) weight function w_i such that the w -weighted variance estimate

$$\frac{\sum_{i=1}^m w_i r_{t-i}^2}{\sum_{i=1}^m w_i}$$

is optimal. Clearly EWMA with decay constant λ is an example of this approach with $w_i = (1 - \lambda)^i$. The function w is known as a half-kernel.

The literature suggests that the normal half kernel (where the weights are defined by negative half of the normal PDF) is often of interest: see Wand & Jones [41] for a discussion of this and other half kernels and Figure 10 for a comparison of the normal half-kernel weighting function to that used in the EWMA approach.

The probability of worst loss approach was used to test a number of normal half kernel volatility estimators with different widths. Figure 11 shows the performance of these volatility estimators as a function of the width parameter. For comparability with other approaches, width is measured in ‘half life’, i.e. the number of days before the weighting function falls to 50% of its peak value.

The best performing normal half kernel is one with a half life of 23 days, roughly corresponding to a lambda of 0.97 (in the sense that $0.97^{23} \approx 0.5$), again suggesting that weighting schemes which fall to half strength over roughly 20-60 days tend to produce acceptable volatility estimates for the S&P 500 index. There are obviously many more half kernels that could be evaluated at this point, and this might be a fruitful area of further work.

4.4 Volatility estimation via signal processing

The philosophy of the half kernel approach is that it is not *a priori* known what the right weighting scheme for calculating volatility from squared returns is, so one should be selected based on

Normal half kernel and EWMA weights

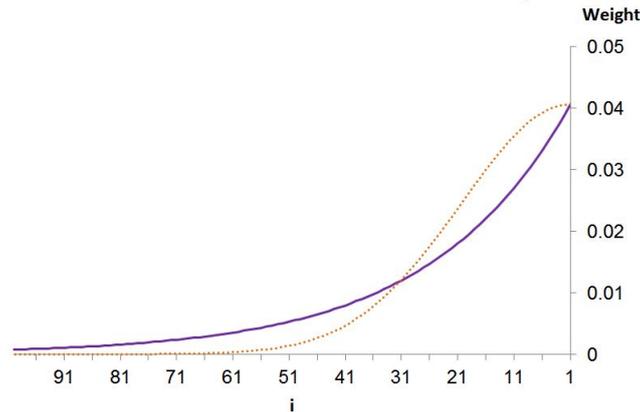


Figure 10: The weight functions for a normal half kernel (orange dotted line) and an EWMA approach (blue solid line)

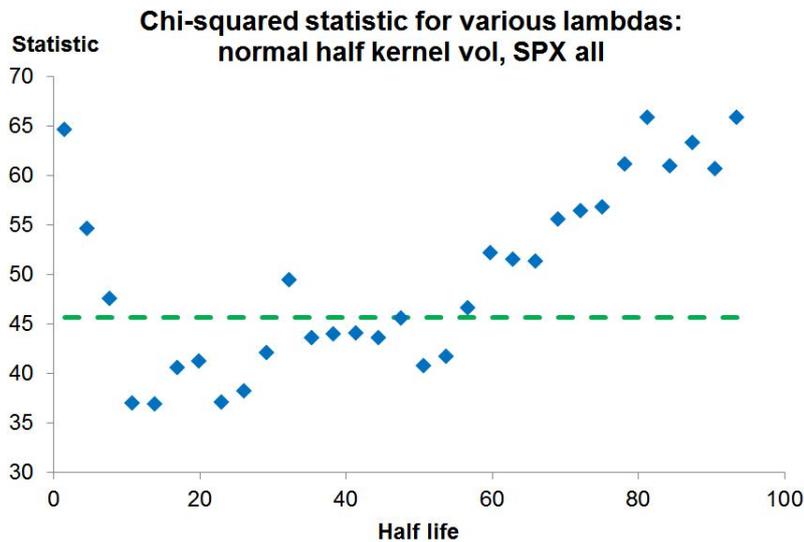


Figure 11: The χ^2 statistic for normal half kernel volatility estimates of various half lives.

performance. This problem specification suggests that we treat volatility estimation as a signal processing problem: the squared returns are the input, and volatility estimates are the output. This stance is particularly productive as there is a large literature on signal processing which can potentially be drawn upon: see for instance Proakis & Manolakis [39] for an introduction to this literature.

A key insight in signal processing is that it is sometimes helpful to work in the frequency domain. Thus many signal processing techniques first transform the input series into an equivalent representation in the frequency domain using a Fourier transform, manipulate this representation, then transform back. For instance a low pass filter retains the low frequency components of a signal while discarding the high frequency ones. If the signal is $\{r_n\}, n \in 0 \dots N - 1$, the simplest low-pass filtering approach would:

1. Calculate the (complex) coefficients of the k^{th} frequency representation

$$X_k = \sum_{n=0}^{N-1} r_n \cdot \exp\left(\frac{-2\pi i k n}{N}\right)$$

for each $k \in 0 \dots N - 1$

2. Cut-off the frequencies above some threshold v , applying a filter F by defining:

$$Y_k = F_k X_k \quad \text{where} \quad F_k = \begin{cases} 1 & \text{if } k \leq v \\ 0 & \text{otherwise} \end{cases} \quad (4)$$

3. Rebuild the filtered return series r'_i using the inverse Fourier transform on the Y_k s

$$r'_n = \frac{1}{N} \sum_{k=0}^{N-1} Y_k \cdot \exp\left(\frac{2\pi i k n}{N}\right)$$

4. Estimate volatility using the low pass filtered returns $\{r'_n\}$.

For a well-chosen v , this approach would filter out high frequency noise but retain the information in lower frequency variation in squared returns. More sophisticated versions would use a more gradual cut-off than the step function used in equation (4).

Techniques based on Fourier analysis have been used in volatility estimation by various authors including Mancino and co-authors [34, 35] and Barucci & Renò [5]. They offer significant promise in being able to offer a highly-customisable (and thus optimisable) framework. In order to assess their potential in our setting, we first compared various low pass filters.

The Hann half-window¹⁰ was found to perform well using 512 days of returns. This filter is defined by two parameters, the centre point v (i.e. the point where the filter is attenuating by 50%) and the width, w :

$$F_k = \begin{cases} 1 & \text{if } k \leq v - w \\ \frac{1}{2} \left(1 - \cos \pi \frac{k+3w-v}{2w}\right) & \text{if } v - w < k < v + w \\ 0 & \text{if } k \geq v + w \end{cases}$$

Figure 12 illustrates the Hann windowing function with a centre point of 450 and a width of 40 applied to a window of 512 days.

The data length was fixed and various centre points and widths for the Hann half window were then tested. The optimal approach was found to be a low pass filter with a high centre point and low width, so that only the highest frequencies are attenuated. With 512 days of data, the highest possible frequency is number 511, and the optimal filter was found to have centre point 509 and width 3. The performance of this approach is shown in Figure 13.

Using a normal half-kernel volatility estimator without filtering, a range of half lives between 28 and 58 days were found to be acceptable. If Figure 11 is compared to Figure 13, it can be seen that the low pass filter increases the acceptable half life: once the high frequency variation in returns is attenuated, longer term volatility estimates perform better.

Estimating volatility in forward windows (i.e. those whose first day is not the day after the estimation day) is more difficult than estimating it for spot windows, as we have thus far. The low pass technique works well here: it turns out that no EWMA model is acceptable for predicting ten day worst losses ten days forward, for instance, but several low-pass normal half kernel estimates are, including the best model for the MPOR starting tomorrow (filter centre point 509 and width 3, half kernel half life 49 days). This performance illustrates that filtering can be a useful technique in volatility estimation.

¹⁰ Window functions are traditionally applied before the Fourier transform, to improve properties of the filter such as spectral leakage or dynamic range, whereas here the window is being applied after the transform.

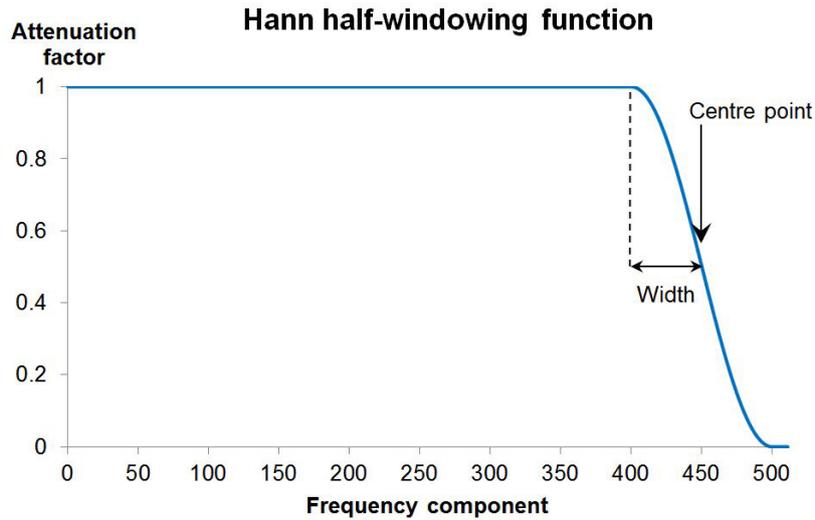


Figure 12: Above: the Hann half-windowing function with centre point 450 and width 40 for a low pass filter. Below: a time series of returns before (in blue) and after (in orange) the application of the filter.

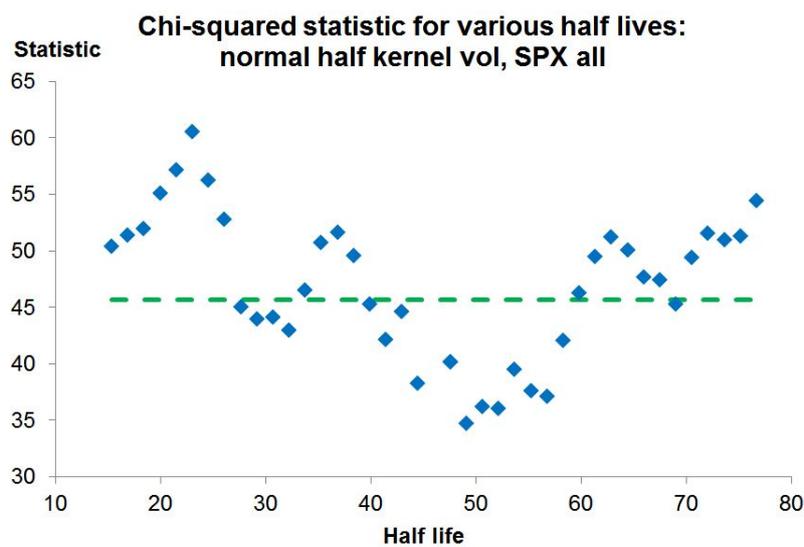
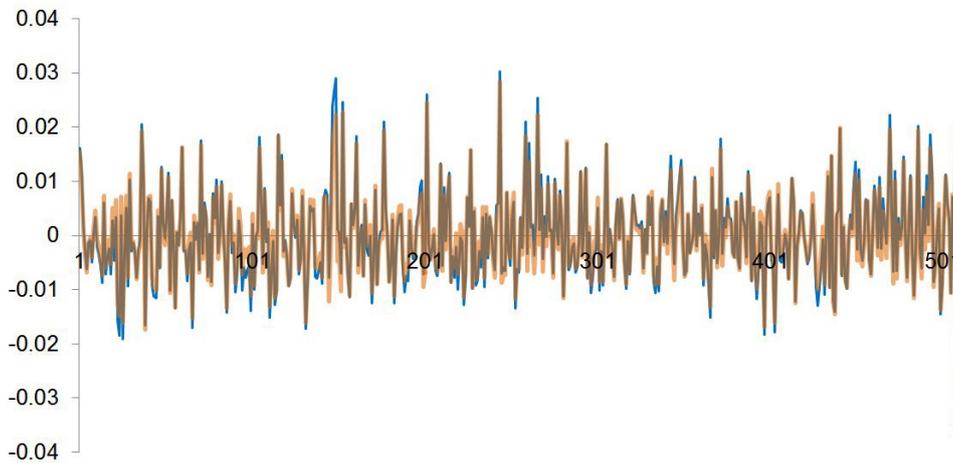


Figure 13: The χ^2 statistic for various normal half kernel volatility estimators using returns filtered with a low pass Hann half window.

4.5 Historical and filtered historical simulation models

Historical simulation techniques are popular in risk modelling. In these approaches (as discussed by Jorion [28]) a time series of returns for the portfolio at hand is calculated by revaluing the portfolio using the actual changes in risk factors experienced during each of the last N periods for some window N . The distribution of value changes is then sampled at the chosen confidence interval to calculate VaR.

A problem arises from the fact that it is difficult to choose a satisfactory N : smaller values make the model more responsive to current conditions but more likely to under-estimate risk in placid periods, while larger values are more likely to ‘remember’ past stress, but are slower to respond.

The more sophisticated filtered historical simulation approach introduced by Hull & White [27] and Barone-Adesi and collaborators [3, 4] and surveyed in Gurrola Perez & Murphy [23] are widely used in margin models. These models address the window size problem using *volatility updating*. Here the historical sample used to estimate VaR is rescaled to be approximately stationary in volatility. This is done by:

- Calculating, for each period in the same $0 \leq n \leq N - 1$, an estimate of the current volatility $\tilde{\sigma}_n$;
- ‘Devolatilising’ the returns by dividing each r_n by $\tilde{\sigma}_n$ to obtain a series of ‘residuals’; then
- Rescaled or ‘revolatilising’ the residuals by multiplying them by the volatility estimate for the current period.

Thus in FHS ‘filtered’ is used in a different sense to signal processing: the historical returns are rescaled to account for the current level of conditional volatility. Typically the estimator used is EWMA volatility calculated using some chosen decay constant λ .

Note that a key feature of both FHS and HS models is that the distribution of returns in the next period is assumed to be specified by some finite set. This observation is the key to calculating the conditional worst loss probabilities implicit in a given HS or FHS model. Suppose that the series of one day returns for a prediction from t is $\{r_n^t\}$ for some window of length N , $n \in 0 \dots N - 1$, and the level of the risk factor at t is x_t . The probability of an observed m -day worst loss starting at t , WL_t can be estimated by the following procedure:

- Pick m returns randomly from $\{r_n^t\}$: r_a^t, r_b^t, \dots say.¹¹
- Calculate an m day path using these returns assuming a starting point of 100:

$$100, 100 \cdot (1 + r_a^t), 100 \cdot (1 + r_a^t) \cdot (1 + r_b^t), \dots$$

- Calculate the worst loss on this simulated path.
- Repeat many times, then ‘bin’ the set of simulated worst losses so that the number of occurrences in a bin an estimate of the probability of observing a worst loss falling between the bin boundaries.
- Determine which bin the observed relative worst loss $100 \cdot WL_t / x_t$ falls in to, and hence how probable this worst loss is, assuming that the returns for the m days forward from t are distributed identically to the model returns r_n^t .

¹¹ The choice of random returns rather than a contiguous series $r_u^t, r_{u+1}^t, \dots, r_{u+m}^t$ is equivalent to the assumption that there is no auto-correlation in the series. This is more reasonable for FHS models, where the claim is that the standardised residuals are white noise, than for HS models.

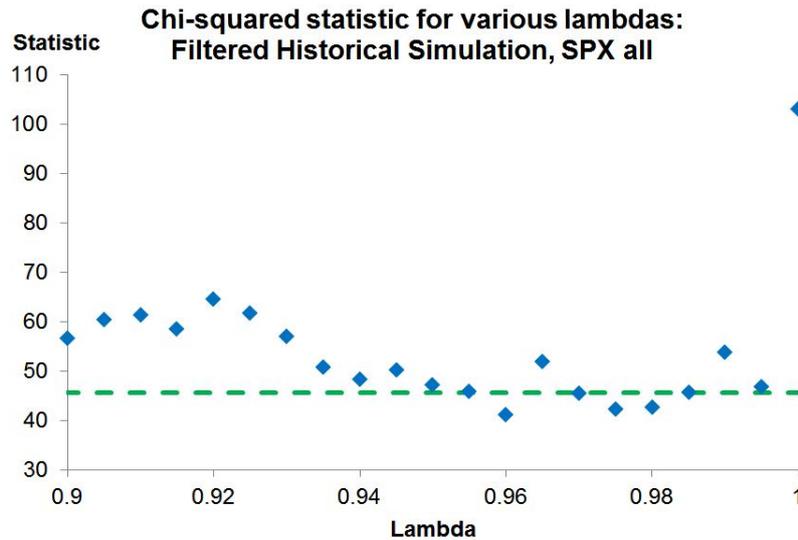


Figure 14: The χ^2 statistic for FHS VaR models using a 512 day data window and various decay constants.

Figure 14 illustrates the performance of a variety of FHS models using the sensitivity analysis technique proposed. Values of the decay constant λ around 0.98 again tend to be preferred. While low ($\lambda < 0.96$) decay constants are still not acceptable, the use of filtered returns improves the performance of these models significantly, suggesting that the residuals here are not pure white noise. The figure also shows that the HS ($\lambda = 1$) model is rejected for this window length.

4.6 GARCH models

Generalised autoregressive conditional heteroskedasticity ('GARCH') models as described by Bollerslev [9] and extended in various ways, for instance by Glosten et al. [21], are popular conditional volatility modelling approaches in the literature. In the simplest version of these models, GARCH(1,1), volatility evolves as

$$\tilde{\sigma}_t^2 = \omega + \alpha r_{t-1}^2 + \beta \tilde{\sigma}_{t-1}^2$$

There are a range of epistemological positions that can be taken regarding models like these. At one extreme, the modeller views the chosen process as a true description of the returns process, and hence their job is to find the 'right' model parameters ω , α and β . Given a long enough history, estimates of these parameters will be accurate in this paradigm, so they should not change materially on recalibration. At the other extreme, the view is that the model's local dynamics are 'close enough' in pertinent respects, and hence the model's predictions may be useful even if the 'true' process is not the modelled one. In this setting there is no reason to expect that best fit parameters will not drift over time, and hence that periodic recalibration will be necessary. Indeed, it will not be surprising here if recalibration results in substantial swings in model parameters. The distinction between the two views is well articulated by Stărică [40]: studying GARCH(1,1) models of S&P 500 returns, he describes the two positions as GARCH(1,1) with a particular set of parameters 'is the true data generating process' and it is 'a local stationary approximation of the data' (and thus by implication the best locally stationary approximation of the data for a different window might well have different parameters).

The 'true description' paradigm can be tested by taking a long history of returns, fitting a GARCH(1,1) model, then testing the out of sample performance of this model at volatility

prediction. It is found that the hypothesis that the GARCH(1,1) predictions of worst losses are correct is rejected at 99%: the critical statistic is over 100 vs. a critical value of about 45.¹²

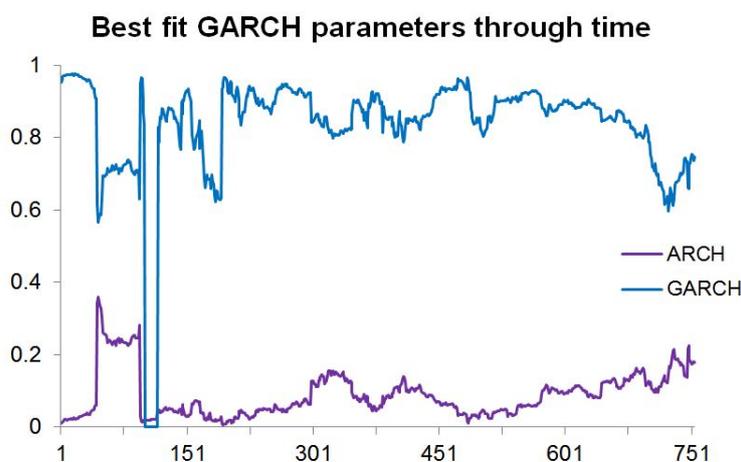


Figure 15: The best fit α (ARCH) and β (GARCH) parameters for each data window in our history of S&P 500 returns.

The ‘locally true’ paradigm requires a recalibration strategy. A window of 512 returns is selected, and a standard GARCH(1,1) model is fitted in each window using a quasi maximum likelihood estimator with variance targeting.¹³ This gives a single volatility prediction for the ten day period at the end of the period, which we can use together with the worst loss in this period. The window is then rolled on ten days and the model is re-fitted. Figure 15 illustrates the fitted parameters as a function of the starting point of the window.

The hypothesis that the volatility estimates generated by this procedure are correct is accepted at 99%. Nevertheless, the unstable model parameters suggest that the recalibration burden of this approach can be substantial.¹⁴

Models with more parameters can sometimes give a better account of returns. As an example, consider Glosten et al.’s GJR-GARCH(1,1) [21]. This model contains an additional parameter, γ , which allows volatility to increase for large negative returns:

$$\tilde{\sigma}_t^2 = \omega + (\alpha + 1_{t-1}\gamma)r_{t-1}^2 + \beta\tilde{\sigma}_{t-1}^2 \quad \text{where} \quad 1_{t-1} = \begin{cases} 0 & \text{if } r_{t-1} > 0 \\ 1 & \text{otherwise} \end{cases}$$

The additional parameter γ sometimes allows GJR-GARCH(1,1) to out-perform GARCH(1,1) models on skewed return series, as Liu & Hung report [30]. However, it makes model estimation more difficult and thus does not necessarily improve the problem with instability of fitted parameters. Indeed, the issue was worse for the data used here, with negative gammas occasionally being returned during the same period of stressed conditions that caused unstable GARCH(1,1) parameter fits. The resulting GJR GARCH volatility estimates also (just) failed

¹² This rejection of GARH predictions is not supported by some other studies based on different data sets: see Andersen & Bollerslev [2] or Hansen & Lunde [26].

¹³ Variance targeting constrains the fit so that the asymptotic variance of the process is set to the unconditional variance. This obviates the need to estimate ω and so reduces the complexity of the fitting problem.

¹⁴ The big fall in the GARCH parameter on day 101 of Figure 15 corresponds to the October 1987 crash. This is worrying as it tends to suggest that just when we need model parameter (and hence margin) stability – when the market is crashing – we do not have it.

the worst loss test. Thus this is an interesting counterexample where the additional freedom of another parameter does not improve volatility estimation.

These results are not intended as a general critique of GJR-GARCH(1,1): it may be that this technique, or the related threshold GARCH models described by Li and Lam [32] could perform well for different windows or return horizons. Rather it illustrates the practical difficulty that better explanatory power most of the time is not useful in a technique used in a margin model if it comes with occasional calibration problems. Margin posters are unlikely to agree to post margin calculated mostly using a complex model, but sometimes using something simpler because the complex model has a bad calibration day.

5 Nuances and extensions

This section describes some complexities which should be recognised before using the probability of worst loss approach for sensitivity analysis in practice.

5.1 Data requirements

Techniques which discriminate between risk models, and especially those that discriminate between the identical algorithms with similar parameters, tend to require a lot of data. Indeed, Danielsson & Zhou state ([13], page 5):

A sample size of half a century of daily observations is needed for the empirical estimators [of VAR] to achieve their asymptotic properties.

Our approach is not quite this constraining, but it is nearly so: even with 30 years worth of data, Figure 8 is not smooth in the model parameter, indicating some noise in the estimates of the χ^2 statistic. In this case it might be helpful to reduce the number of probability buckets. The lack of smoothness also suggests that in practice there will not be many risk factors where enough data is available to use fine bucketing with confidence.

5.2 A better test for uniformity

The χ^2 approach to normality testing, while appealingly simple, has relatively low power. A more powerful approach would be to use Berkowitz's probability integral transforms [7, 25]. Here the assumed-uniform distribution is transformed using the inverse cumulative normal distribution, giving data z_t . A first order autoregressive process

$$z_t - \mu = \rho(z_{t-1} - \mu) + \epsilon_t$$

is then fitted. If the origination distribution was uniform, then μ and ρ should be close to zero and the variance of the ϵ s should be close to 1. Likelihood ratios can be used to test this. We hope to explore the probability integral transform approach in further work.

5.3 Risk factor differences

There is no reason to believe that the same volatility estimator will be optimal, or even acceptable, for all the risk factors in a large portfolio. To study this, the various EWMA estimators were re-tested using the USD/JPY spot rate as the risk factor instead of the S&P 500 index. Figure 16 illustrates the results. There are some differences: $\lambda = 0.93$ is acceptable for USD/JPY, but not for the S&P 500, for instance. Clearly there is some danger that as the number of risk factors grow, there will be no model that is acceptable for all of them.¹⁵

¹⁵ Once the estimation of covariances is included in the problem, this problem clearly becomes worse.

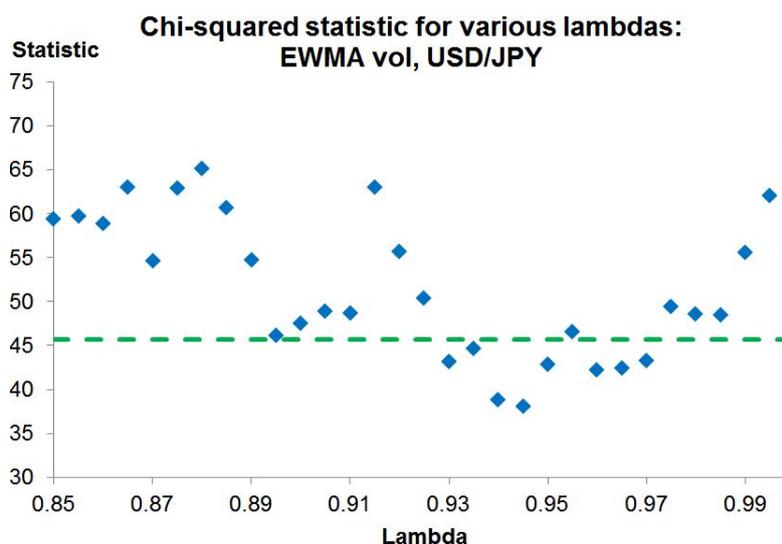


Figure 16: The χ^2 statistic for various EWMA volatility estimators applied to USD/JPY returns.

6 Conclusions

A key part of initial margin model validation will always be backtesting model margin requirements at the target confidence interval. However, this is not enough for optimal model design or to perform sensitivity analysis: other techniques will be needed to verify the choice of model parameters. One approach to this problem has been presented which focusses on the statistical accuracy of the models' predictions of worst loss over the margin period of risk. Techniques like this can also help in providing early warning of a model which is providing accurate high quantile risk estimates for a limited period of time, but which will not prove to be robust as market conditions evolve.

Various algorithms have been tested using the technique proposed, and acceptable parameterisations have been presented. The algorithms range from the well-known and commonly used, such as EWMA volatility estimation, to the techniques inspired by the idea that volatility estimation can be thought of as a signal processing problem. These latter approaches are potentially interesting in that they open up a large stock of filtering techniques which can assist in separating out uninformative high frequency noise in returns from valuable information about volatility trends.

One interesting feature of the analysis presented is that most models tested had *some* acceptable range of parameters, and that for decay-constant-like parameters, the acceptable range tended to include models with a half life of between 20 and 60 days. This suggests that persistence on this time scale is a somewhat model-independent feature of the S&P 500.

It can sometimes be found that the performance of a model is sufficiently insensitive to parameters choices that a wide range of them are acceptable: the blended volatility model illustrated in Figure 9 is an example of this. When this happens it is helpful to have an additional criterion for preferring one parameter setting over another. The obvious choice is *procyclicality*: margin stability is valuable, and so less procyclical models are to be preferred over more reactive ones. Murphy et al. [36] discuss measures of margin model procyclicality which can be used here. In other cases when a model only performs acceptably for a narrow range of parameters the choice is easy. Neither outcome however absolves the model user from continued diligence: it is important to perform sensitivity analysis regularly to ensure that model design and parameters remain appropriate.

References

- [1] F. Aitsahlia and T. Lai, *Random walk duality and the valuation of discrete lookback options*, Applied Mathematical Finance Volume 5, 1998
- [2] T. Andersen, T. Bollerslev, *Answering the Skeptics: Yes, Standard Volatility Models Do Provide Accurate Forecasts*. International economic review, 1998 available at econ.duke.edu/~boller/Published_Papers/ier_98.pdf
- [3] G. Barone-Adesi, K. Giannopoulos, *Non-parametric VaR techniques: myths and realities*. Economic Notes, Volume 30, Banca Monte dei Paschi di Siena, Siena, 2001
- [4] G. Barone-Adesi, K. Giannopoulos, L. Vosper *VaR without Correlations for Portfolios for Portfolios of Derivative Securities*. Journal of Futures Markets, Volume 19, 1999
- [5] E. Barucci, R. Renò, *On measuring volatility and the GARCH forecasting performance*. Journal of International Financial Markets, Institutions and Money, Volume 12, 2002
- [6] Basle Committee on Banking Supervision, *Margin requirements for non-centrally cleared derivatives*. BCBS 261, BIS 2013
- [7] J. Berkowitz, *Testing Density Forecasts with Applications to Risk Management*. Journal of Business and Economics Statistics, Volume 19, 2001
- [8] J. Berkowitz, P. Christoffersen, D. Pelletier, *Evaluating Value-at-Risk Models with Desk-Level Data*. Management Science, Volume 57 Issue 12, 2011
- [9] T. Bollerslev, *Generalized autoregressive conditional heteroskedasticity*. Journal of Econometrics, Volume 31, 1986
- [10] S. Campbell, *A Review of Backtesting and Backtesting Procedures*. Finance and Economics Discussion Series Divisions of Research and Statistics and Monetary Affairs, Board of Governors of the Federal Reserve System, 2005
- [11] P. Christoffersen, *Evaluating Interval Forecasts*. International Economic Review Volume 39, 1998
- [12] C. Crnkovic, J. Drachman, *"Quality Control" in VaR: Understanding and Applying Value-at-Risk*. Risk publications, 1997
- [13] J. Daniélsson, Chen Zhou, *Why risk is so hard to measure*. Systemic Risk Centre discussion paper No. 36, London School of Economics, 2015
- [14] F. Diebold, T. Gunther, A. Tay, *Evaluating density forecasts with applications to financial risk management*. International Economic review, Volume 39, 1998
- [15] R. Engle, *Autoregressive Conditional Heteroskedasticity with Estimates of United Kingdom Inflation*, Econometrica, Volume 50, 1982
- [16] J. Escanciano, Pei Pei, *Pitfalls in backtesting historical simulation VAR models*. Journal of Banking & Finance, Volume 36, Issue 8, August 2012
- [17] European Union, *Regulation (EU) No. 648/2012 on OTC derivatives, central counterparties and trade repositories*, 2012, commonly known as 'EMIR'
- [18] European Union, *Commission Delegated Regulation (EU) No. 153/2013 of 19 December 2012 supplementing Regulation (EU) No. 648/2012 of the European Parliament and of the Council with regard to regulatory technical standards on requirements for central counterparties*, commonly known as the 'EMIR RTS'



- [19] G-20 Leaders Statement: The Pittsburgh Summit, September 24-25, 2009 available at www.g20.utoronto.ca/2009/2009communique0925.html
- [20] I. Gijbels, A. Pope, M. Wand, *Understanding exponential smoothing via kernel regression*. Journal of the Royal Statistical Society B, Volume 51, 1999
- [21] L. Glosten, R. Jagannathan, D. Runkle, *On The Relation between The Expected Value and The Volatility of Nominal Excess Return on stocks*. Journal of Finance, Volume 48, 1993
- [22] P. Gurrola Perez, *Calibrating a new generation of initial margin models under the new regulatory framework*. Proceedings of the Third Annual LSE Conference on Systemic Risk, 2015
- [23] P. Gurrola Perez, D. Murphy, *Filtered historical simulation Value-at-Risk models and their competitors*. Bank of England Working Paper No. 525, March 2015
- [24] M. Haas, *New Methods in Backtesting*. Financial Engineering Note, Research Center Caesar, Bonn, 2001 available at <http://www.ime.usp.br/~rvicente/risco/haas.pdf>
- [25] A. Hamerle, K. Plank *A note on the Berkowitz test with discrete distributions*. Journal of Risk Model Validation, Volume 3, Number 2, 2009
- [26] P. Hansen, A. Lunde, *A forecast comparison of volatility models: does anything beat a GARCH(1,1)?* Journal of Applied Econometrics Volume 20, Number 7, December 2005
- [27] J. Hull, and A. White, *Incorporating Volatility Updating into the Historical Simulation Method for Value at Risk*. Journal of Risk, Volume 1, Number 1, 1998
- [28] P. Jorion, *Value at risk*, IRWIN, 1997 (3rd edition, McGraw-Hill 2006)
- [29] P. Kupiec, *Techniques for verifying the accuracy of risk measurement models*. The Journal of Derivatives, Volume 3, Number 2, 1995
- [30] Hung-Chun Liu, Jui-Cheng Hung, *Forecasting S&P-100 stock index volatility: The role of volatility asymmetry and distributional assumption in GARCH models*, Expert Systems with Applications, Volume 37, 2010
- [31] R. Liesenfeld, R. Jung, *Stochastic volatility models: conditional normality versus heavy-tailed distributions*, Journal of Applied Econometrics, Volume 15, 2000
- [32] W. Li, K. Lam, *Modelling Asymmetry in Stock Returns by a Threshold Autoregressive Conditional Heteroscedastic Model*, Journal of the Royal Statistical Society, Volume 44, Number 3, 1995
- [33] J. Lopez, *Evaluating the predictive accuracy of variance models*. Journal of Forecasting, Volume 20, 2001
- [34] P. Malliavin, M. Mancino, *Fourier series method for measurement of multivariate volatilities*. Finance and Stochastics, Volume 6, 2002
- [35] M. Mancino, M. Recchioni, *Fourier Spot Volatility Estimator: Asymptotic Normality and Efficiency with Liquid and Illiquid High-Frequency Data*. PLOS ONE, DOI: 10.1371/journal.pone.0139041, 2015
- [36] D. Murphy, M. Vasios, N. Vause, *An investigation into the procyclicality of risk-based initial margin models*. Bank of England Financial Stability Paper Number 29, 2014
- [37] D. Pelletier, Wei Wei, *The Geometric-VAR Backtesting Method*. Journal of financial econometrics, July 2015.

- [38] S. Poon, C. Granger, *Forecasting volatility in financial markets: A review*. Journal of economic literature, Volume 41, Number 2, 2003.
- [39] J. Proakis, D. Manolakis, *Digital Signal Processing*. 4th Edition, Prentice Hall, 2006
- [40] C. Stărică, *Is GARCH(1,1) as good a model as the Nobel prize accolades would imply?* Chalmers University of Technology, 2003 available at SSRN <http://ssrn.com/abstract=637322>
- [41] M. Wand, M. Jones, *Kernel Smoothing*. Chapman and Hall, London, 1995