Volatility bands with predictive validity*

Dimitris N. Politis
Professor of Mathematics and Economics
University of California, San Diego
La Jolla, CA 92093-0112, USA
www.math.ucsd.edu/∼politis

Abstract

The issue of volatility bands is re-visited. It is shown how the rolling geometric mean of a price series can serve as the centerline of a novel set of bands that enjoy a number of favorable properties including predictive validity.

Acknowledgment. Many thanks are due to A. Venetoulias of Quadrant Management for introducing me to Bollinger bands, to K. Thompson of Granite Portfolios for the incitement to re-visit this issue, and to two reviewers for their helpful comments.

*Paper to appear in the *Journal of Technical Analysis* in Jan./Feb. 2007.

Volatility bands with predictive validity

Abstract: The issue of volatility bands is re-visited. It is shown how the rolling geometric mean of a price series can serve as the centerline of a novel set of bands that enjoy a number of favorable properties including predictive validity.

Introduction

Consider financial time series data $P_1, \ldots, P_n$ corresponding to recordings of a stock index, stock price, foreign exchange rate, etc.; the recordings may be daily, weekly, or calculated at different (discrete) intervals. Also consider the associated percentage returns $X_1, \ldots, X_n$. As is well known, we have:

$$X_t = \frac{P_t - P_{t-1}}{P_{t-1}} \simeq \log P_t - \log P_{t-1}, \tag{1}$$

the approximation being due to a Taylor expansion under the assumption that X_t is small; here log denotes the natural logarithm.

Eq. (1) shows how/why the logarithm of a price series, i.e., the series $L_t := \log P_t$, enters as a quantity of interest. Bachelier's (1900) original implication was that the series L_t is a Gaussian random walk, i.e., Brownian motion. Under his simplified setting, the returns $\{X_t\}$ are (approximately) independent and identically distributed (i.i.d.) random variables with Gaussian $N(0, \sigma^2)$ distribution.

Of course, a lot of water has flowed under the bridge since Bachelier's pioneering work. The independence of returns was challenged first by Mandelbrot (1963) who pointed out the phenomenon of 'volatility clustering'. Then, the Gaussian hypothesis was challenged in the 1960s by Fama (1963) who noticed that the distribution of returns seemed to have fatter tails than the normal. Engle's (1982) ARCH models attempt to capture both of the above two phenomena; see Bollerslev et al. (1992) and Shephard (1996) for a review of ARCH/GARCH models and their application. More recently, Politis (2003, 2004, 2006) has developed a model-free alternative to ARCH/GARCH models based on the notion of normalizing transformations that we will find useful in this paper.

On top of an ARCH-like structure, the working assumption for many financial analysts at the moment is that the returns $\{X_t\}$ are locally stationary, i.e., approximately stationary when only a small time-window is considered, and approximately uncorrelated. As far as the first two moments are concerned, this local stationarity can be summarized as:[1]

$$EX_t \simeq 0 \text{ and } \; Var(X_t) \simeq \sigma^2. \tag{2}$$

Here E and Var denote expectation and variance respectively. Letting Cov denote covariance, the approximate uncorrelatedness can be described by:

$$Cov(X_t, X_{t-k}) \simeq 0 \text{ for all } \; k > 0. \tag{3}$$

Note, that since $X_t = L_t - L_{t-1}$, the condition $EX_t \simeq 0$ of eq. (2) is equivalent to

$$EL_t = \mu_t \text{ with } \; \mu_t \simeq \mu_{t-1}. \tag{4}$$

Thus, the mean of L_t and the variance of X_t can be thought to be approximately constant within the extent of a small time-window.

[1] The approximate constancy of the (unconditional) variance in eq. (2) does not contradict the possible presence of the conditional heteroscedasticity (ARCH) phenomenon and volatility clustering.

A simple way to deal with the slowly changing mean μ_t is the popular Moving Average employed in financial analysis to capture trends. Furthermore, the notion of volatility bands (Bollinger bands) has been found useful in applied work. It is important to note, though, that the usual volatility bands do *not* have predictive validity; see e.g. Bollinger (2001). Similarly, no claim can be made that a desired percentage of points fall within the Bollinger bands.

Nevertheless, it is not difficult to construct volatility bands that *do* have predictive validity, i.e., prescribe a range of values that—with high probability—will 'cover' the future price value P_{n+1}. We construct such predictive bands in the paper at hand; see eq. (22) in what follows. To do this, the notion of a geometric mean moving average will turn out very useful.

Before proceeding, however, let us briefly discuss the notion of predictive intervals. The issue is to predict the future price P_{n+1} on the basis of the observed data $P_1, \ldots, P_n$. Denote by $\hat{P}_{n+1}$ our predictor; this is a 'point' predictor in the sense that it is a point on the real line. Nevertheless, with data on a continuous scale, it is a mathematical *certainty* that this point predictor—however constructed—will result in some error.

So we may define the prediction error $w_{n+1} = P_{n+1} - \hat{P}_{n+1}$ and study its statistical properties. For example, a good predictor will result into $Ew_{n+1} = 0$ and $Var(w_{n+1})$ that is small. The statistical quantification of the prediction error may allow the practitioner to put a 'margin-of-error' around the point predictor, i.e., to construct a 'predictive interval' with a desired coverage level.

The notion of 'predictive interval' is analogous to the notion of 'confidence interval' in parameter estimation. The definition goes as follows: a predictive interval for P_{n+1} is an interval of the type $[A, B]$ where A, B are functions of the data $P_1, \ldots, P_n$. The probability that the future price P_{n+1} is actually found to lie in the predictive interval $[A, B]$ is called the interval's coverage level.

The coverage level is usually denoted by $(1-\alpha)100\%$ where α is chosen by the practitioner; choosing $\alpha = 0.05$ results into the popular 95% coverage level. The limits A, B must be carefully selected so that a prescribed coverage level, e.g. 95%, is indeed achieved (at least approximately) in practice; see e.g. Geisser (1993) for more details on predictive distributions and intervals.

I. Smoothing and prediction

Under eq. (2)–(4) and given our data $L_1, \ldots, L_n$, a simple nonparametric estimator of μ_{n+1} is given by a general Moving Average[2] in the log-price domain, i.e., by

$$MAL_{n,\theta,q} = \sum_{k=0}^{q-1} \theta_k L_{n-k}, \tag{5}$$

where q is the length of the Moving Average window, and θ_k some weights that sum to one, i.e., $\sum_{k=0}^{q-1} \theta_k = 1$. The simplest choice is letting $\theta_k = 1/q$ for all k, i.e., equal weights, but other weights are also possible e.g. exponential smoothing weights. Choosing q is of course a difficult problem and involves the usual tug-of-war: q should be small so that local stationarity holds for $L_{n-q+1}, \ldots, L_n$ but q should be large as well so that the averaging effect is successful.

[2]Here, and throughout the paper, all Moving Averages will be predictive in nature, i.e., only use present and past data.

To get a prediction interval for the next observation L_{n+1}, an expression for the variance of the prediction error: $L_{n+1} - MAL_{n,\theta,q}$ is required. It is easy to calculate that:

$$\sigma^2_{pred,q} := Var(L_{n+1} - MAL_{n,\theta,q}) =$$

$$Var(L_{n+1}) + Var(MAL_{n,\theta,q}) - 2Cov(L_{n+1}, MAL_{n,\theta,q}) \tag{6}$$

where

$$Var(L_{n+1}) = (n+1)\sigma^2, \tag{7}$$

$$Var(MAL_{n,\theta,q}) =$$

$$\sigma^2 \left(n - q + 1 + [\sum_{k=0}^{q-2} \theta_k]^2 + [\sum_{k=0}^{q-3} \theta_k]^2 + \cdots + [\theta_0 + \theta_1 + \theta_2]^2 + [\theta_0 + \theta_1]^2 + \theta_0^2 \right) \tag{8}$$

and $Cov(L_{n+1}, MAL_{n,\theta,q}) =$

$$\sigma^2 \left(n - q + 1 + [\sum_{k=0}^{q-2} \theta_k] + [\sum_{k=0}^{q-3} \theta_k] + \cdots + [\theta_0 + \theta_1 + \theta_2] + [\theta_0 + \theta_1] + \theta_0 \right). \tag{9}$$

For any given chosen set of weights $\theta_k,\ k = 0, 1, \ldots, q-1$, equations (6), (7), (8) and (9) give an expression for the desired quantity $\sigma^2_{pred,q}$. The n's cancel out in the final expression so the starting point of the series is immaterial.

To simplify this expression, let us now—and for the remainder of the paper—focus on the simple Moving Average with equal weights $\theta_k = 1/q$, i.e.,

$$MAL_{n,q} = \frac{1}{q} \sum_{k=0}^{q-1} L_{n-k}. \tag{10}$$

In this case, eq. (8) and (9) give:

$$Var(MAL_{n,q}) = \sigma^2 \left(n - q + 1 + \frac{(q-1)(2q-1)}{6q} \right) \tag{11}$$

and

$$Cov(L_{n+1}, MAL_{n,q}) = \sigma^2 (n - \frac{q-1}{2}). \tag{12}$$

Putting it all together, we have the following formula for the prediction error variance:

$$\sigma^2_{pred,q} = Var(L_{n+1} - MAL_{n,q}) = \sigma^2 \left(1 + \frac{(q-1)(2q-1)}{6q} \right). \tag{13}$$

Note that σ^2 is unknown in the above and must be estimated from data; to this end, we propose using the sample variance of the last Q returns, i.e., let

$$\hat{\sigma}^2 = \frac{1}{Q} \sum_{k=0}^{Q-1} X^2_{n-k}. \tag{14}$$

Estimating $\sigma^2_{pred,q}$ by

$$\hat{\sigma}^2_{pred,q} = \hat{\sigma}^2 \left(1 + \frac{(q-1)(2q-1)}{6q} \right), \tag{15}$$

we are led to the approximately standardized[3] ratio

$$\frac{L_{n+1} - MAL_{n,q}}{\hat{\sigma}_{pred,q}}. \tag{16}$$

Recall that Bachelier (1900) postulated that L_t is Gaussian. As mentioned in the beginning of the paper, this belief does not hold any more. However, with Q properly chosen, it has recently been shown by Politis (2003, 2006) that the ratio of eq. (16) has an approximately standard normal distribution.

[3] A random variable W is called 'standardized' if it has mean zero and variance one.

II. Geometric vs. arithmetic mean

Let $z(\alpha)$ denote the α–quantile of the standard normal distribution, i.e., $z(\alpha)$ is such that the region $\pm z(\alpha/2)$ captures probability equal to $1-\alpha$.Then, the standard normal approximation for the ratio of eq. (16) implies that

$$[MAL_{n,q} - z(\alpha/2)\hat{\sigma}_{pred,q}, \quad MAL_{n,q} + z(\alpha/2)\hat{\sigma}_{pred,q}] \tag{17}$$

is an (approximate) $(1-\alpha)100\%$ *prediction interval* for the unobserved value L_{n+1}. A typical value for α is 0.05, i.e., 95% prediction intervals, with corresponding quantile $z(\alpha/2) = z(0.025) = 1.96$.

Let $A = z(\alpha/2)\hat{\sigma}_{pred,q}$ as a short-hand. The above prediction interval has the interpretation that the event

$$MAL_{n,q} - A \leq L_{n+1} \leq MAL_{n,q} + A \tag{18}$$

occurs with probability $1-\alpha$. Exponentiating both sides of eq. (18) the inequalities are preserved; it follows that the event

$$\exp(MAL_{n,q}) \cdot \exp(-A) \leq P_{n+1} \leq \exp(MAL_{n,q}) \cdot \exp(A) \tag{19}$$

occurs with probability $1-\alpha$, i.e., a $(1-\alpha)100\%$ *prediction interval* for the unobserved (un-logged) price P_{n+1}. Finally, note that

$$\exp(MAL_{n,q}) = \exp(\frac{1}{q}\sum_{k=0}^{q-1} L_{n-k}) = \exp\left(\sum_{k=0}^{q-1}\log(P_{n-k}^{1/q})\right) = \left(\prod_{k=0}^{q-1} P_{n-k}\right)^{1/q} \tag{20}$$

which is nothing other than the *geometric mean* of the values $P_{n-q+1}, \ldots, P_n$.

Let us denote the geometric mean by

$$GMP_{n,q} = \left(\prod_{k=0}^{q-1} P_{n-k}\right)^{1/q}. \tag{21}$$

Then, our proposed $(1-\alpha)100\%$ prediction interval for P_{n+1} is given by

$$GMP_{n,q} \cdot \exp(-z(\alpha/2)\hat{\sigma}_{pred,q}) \ \le P_{n+1} \le \ GMP_{n,q} \cdot \exp(z(\alpha/2)\hat{\sigma}_{pred,q}) \tag{22}$$

i.e., it is an interval centered around the geometric mean $GMP_{n,q}$; the latter also serves as the point predictor of P_{n+1} given our data. Note that this is an asymmetric interval due to the nonlinearity of the exponential function, and also because the upper and lower bounds are given in a multiplicative way in connection with the center value.

We conclude this section with a step-by-step algorithm for the construction of the predictive interval.

Algorithm for the construction of predictive interval (22):

1. Decide on the desired coverage level, i.e., choose α; a typical choice is $\alpha = 0.05$.

2. Look up $z(\alpha/2)$ from a table on the standard normal distribution; for example, $\alpha = 0.05$ yields $z(\alpha/2) = 1.96$.

3. Choose values for q and Q; see Section IV for a discussion on this issue.

4. Compute $\hat{\sigma}^2$ from eq. (14), and $\hat{\sigma}^2_{pred,q}$ from eq. (15).

5. Compute the geometric mean $GMP_{n,q}$ from eq. (21).

6. Compute the left and right limits of the $(1-\alpha)100\%$ predictive interval as $GMP_{n,q} \cdot \exp(-z(\alpha/2)\hat{\sigma}_{pred,q})$ and $GMP_{n,q} \cdot \exp(z(\alpha/2)\hat{\sigma}_{pred,q})$ respectively.

III. An illustration

The predictive interval (22) can be computed for n as small as the $\max(q, Q)$ and as large as N (the end of our dataset), thus providing a set of volatility bands that can be used in the same way that Bollinger bands are used. By contrast to Bollinger bands, however, the predictive bands will contain approximately $(1-\alpha)100\%$ of the data points. Furthermore, letting $n = N$, is is apparent that the predictive bands (22) extend to the future of our dataset, yielding a predictive interval for the value that is one-step-ahead, i.e., P_{N+1}.

As an illustration, Figure 1 depicts daily recordings[4] of the S&P500 index from August 30, 1979 to August 30, 1991 with 95% predictive bands superimposed (using $q = Q = 10$); the extreme values associated with the crash of October 1987 are very prominent in the plot. Because of the density of the plot, Figure 2 focuses on shorter (4-month) periods.

In Figure 2 (a) is is apparent that the predictive bands miss (as expected) the largest downward movement of the 1987 crash, but rebound to widen and capture all the subsequent points; because of the predictive nature of the bands, they would have been of some help to analysts during those days. Figure 2 (b) shows the end of the dataset which is during a more normal trading atmosphere. The ability of the bands to capture about 95% of the points is visually apparent, as is the fact that the bands extend one-step-ahead into the future.

Figure 3 is the same as Figure 2 but using Bollinger bands[5] instead with

[4]The plot is re-normalized to the value of the S&P500 index on August 30, 1979; thus the starting value of one.

[5]Recall that in Bollinger bands the centerline is a standard Moving Average on the price series, i.e., $MAP_{n,q} = (1/q)\sum_{k=0}^{q-1} P_{n-k}$ in our notation. The bands are obtained as

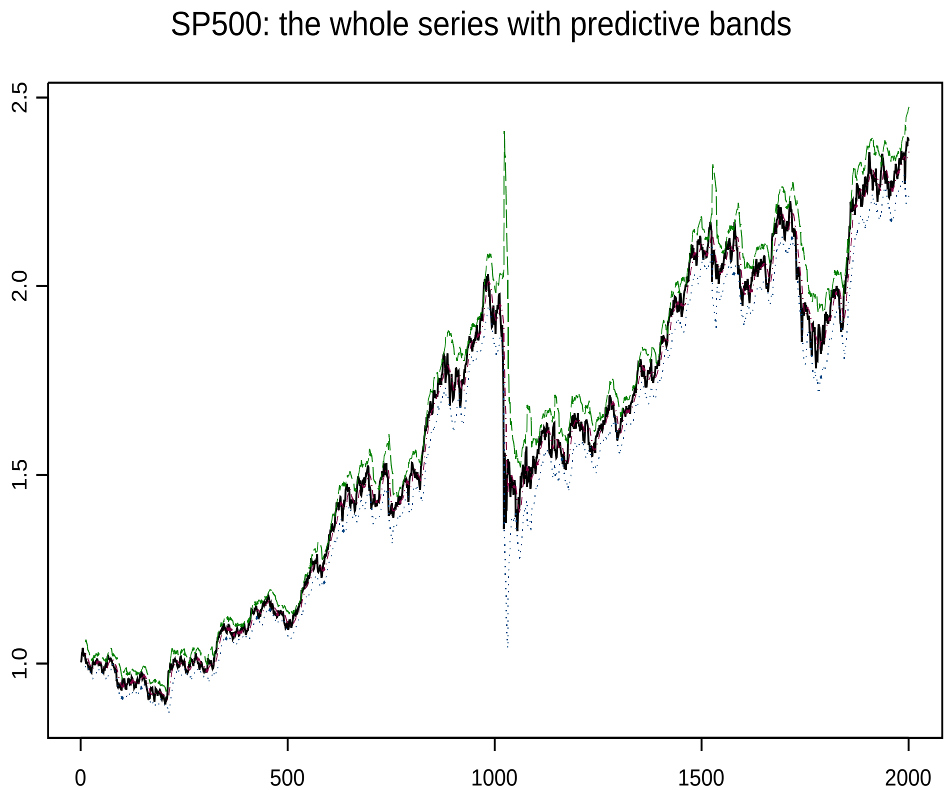

Figure 1: Daily S&P500 index spanning the period 8-30-1979 to 8-30-1991 with 95% predictive bands superimposed (using $q = Q = 10$); the extreme negative value associated with the crash of October 1987 is very prominent. [Black=data, red=centerline, green=upper band, blue=lower band].

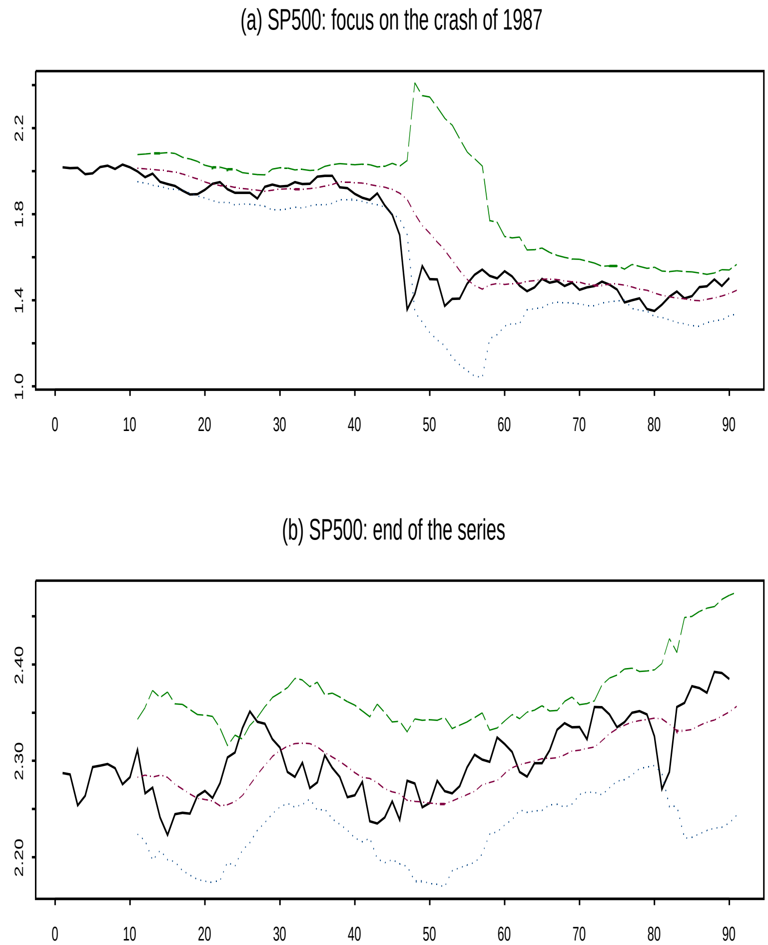

Figure 2: As in Figure 1 but focusing in on shorter (4-month) periods: (a) two months before and two months after the crash of 1987; (b) last 4 months of the series.

the same parameters, i.e., $q = Q = 10$. The Bollinger bands are on average somewhat narrower than our predictive bands of Figure 2; since the latter are designed to capture about 95% of the points, it follows that the Bollinger bands will be able to capture a smaller percentage of points. Indeed, looking over the whole of the S&P500 dataset of Figure 1, the Bollinger bands manage to capture 90.1% of the data points as compared to 94.4% of the predictive bands.

Note that bands that are narrower than they should be may lead to incorrect decision-making. Furthermore, is it apparent that the geometric mean (centerline of Figure 2) tracks (actually: predicts) the data much better than the arithmetic mean (centerline of Figure 3). In addition, the width of Bollinger bands appears to vary with no cause in sight; for example, the Bollinger bands start shrinking just before the crash of 1987 which must have been quite misleading at the time. Lastly, note that Bollinger bands are not predictive in nature, i.e., do not give a prediction interval for the one-step-ahead observation as our predictive bands do.

Similar comments apply to Bollinger bands using $q = Q = 20$ as Bollinger (2001) recommends; see Figure 4 for an illustration. The $q = Q = 20$ Bollinger bands manage to capture only 87% of the points of the S&P500 dataset; in general, as either q or Q increases, the coverage of Bollinger bands decreases/deteriorates.

$MAP_{n,q} \pm B$ where $B = 2S_Q$, and $S_Q^2 = (1/Q)\sum_{k=0}^{Q-1} P_{n-k}^2 - (MAP_{n,Q})^2$, i.e., a rolling sample variance of the price series.

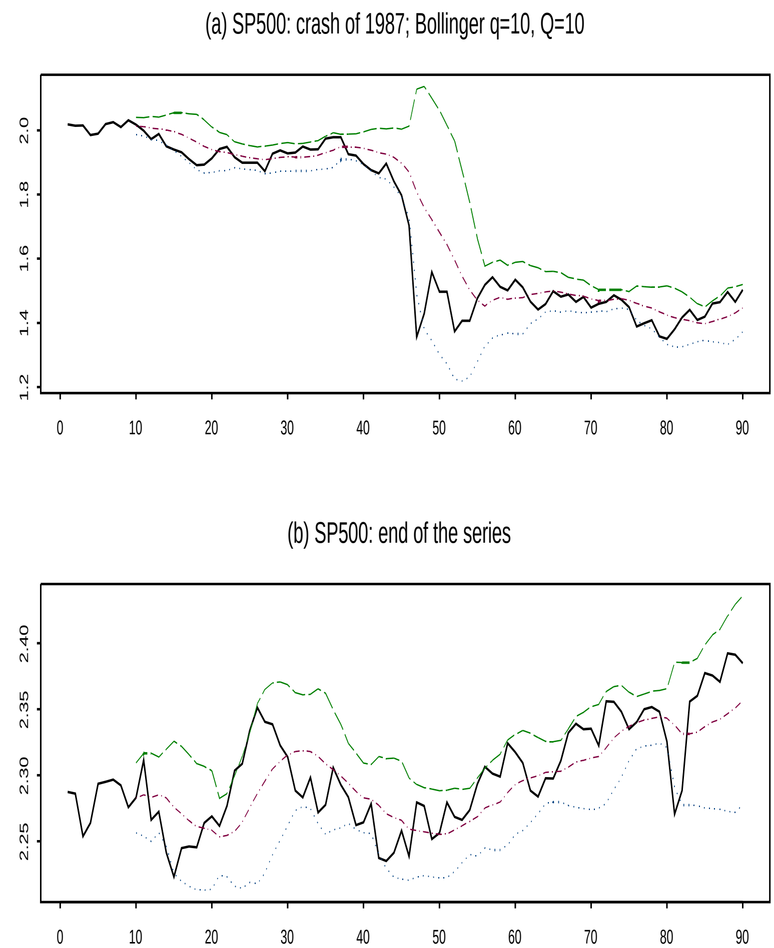

Figure 3: As in Figure 2 but using Bollinger bands with $q = Q = 10$.

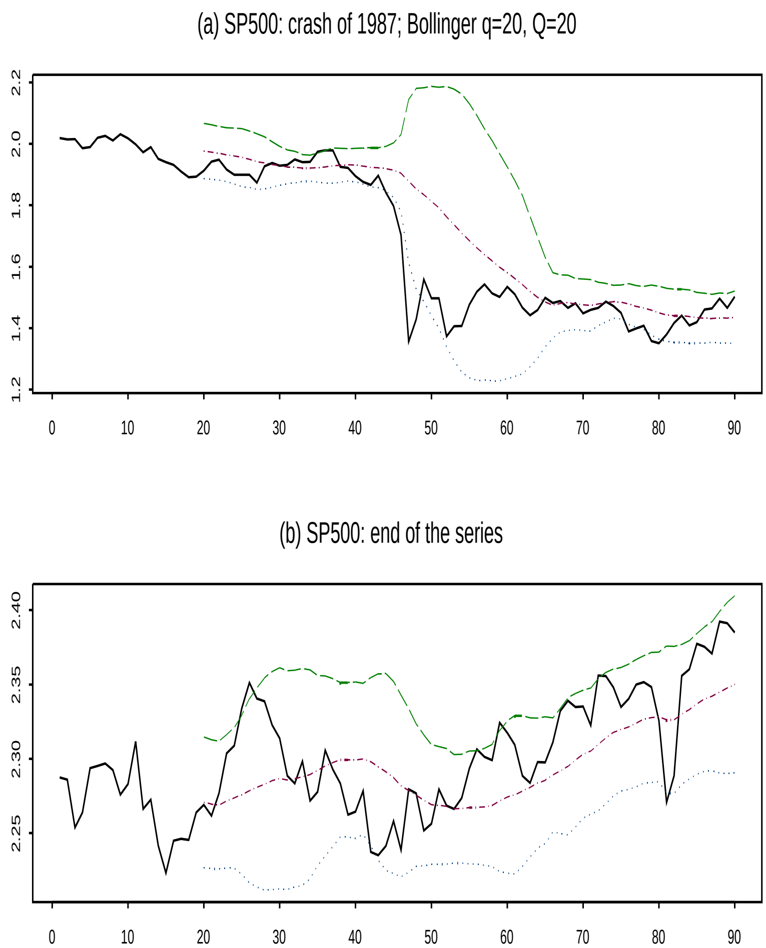

Figure 4: As in Figure 2 but using Bollinger bands with $q = Q = 20$.

$q \setminus Q =$	10	13	16	19	22	25	28	31	34	37	40
10	0.113	0.114	0.115	0.115	0.116	0.117	0.117	0.118	0.118	0.119	0.119
20	0.155	0.157	0.158	0.158	0.159	0.160	0.161	0.161	0.162	0.162	0.163
30	0.189	0.191	0.192	0.193	0.194	0.194	0.195	0.196	0.196	0.197	0.198
40	0.218	0.220	0.221	0.222	0.223	0.224	0.225	0.226	0.226	0.227	0.227
50	0.244	0.246	0.247	0.249	0.250	0.251	0.252	0.253	0.253	0.254	0.254
60	0.267	0.270	0.272	0.273	0.274	0.275	0.276	0.277	0.278	0.279	0.279
70	0.289	0.292	0.294	0.295	0.297	0.298	0.299	0.300	0.301	0.302	0.302
80	0.310	0.313	0.315	0.316	0.318	0.319	0.320	0.322	0.322	0.323	0.324
90	0.329	0.332	0.334	0.336	0.338	0.339	0.341	0.342	0.343	0.344	0.344

Table I. Empirically found *widths* of the 95% predictive bands with different combinations of q and Q for the S&P500 dataset of Figure 1.

$q \setminus Q =$	10	13	16	19	22	25	28	31	34	37	40
10	0.056	0.062	0.070	0.072	0.074	0.075	0.079	0.080	0.084	0.086	0.091
20	0.057	0.059	0.058	0.059	0.061	0.062	0.068	0.075	0.078	0.084	0.086
30	0.062	0.058	0.058	0.059	0.058	0.059	0.063	0.063	0.069	0.076	0.079
40	0.071	0.067	0.058	0.056	0.056	0.060	0.058	0.061	0.060	0.063	0.065
50	0.082	0.076	0.073	0.064	0.061	0.063	0.063	0.063	0.062	0.064	0.064
60	0.091	0.081	0.078	0.076	0.067	0.065	0.067	0.069	0.067	0.065	0.062
70	0.093	0.082	0.084	0.075	0.069	0.065	0.072	0.067	0.069	0.066	0.067
80	0.098	0.084	0.089	0.078	0.072	0.065	0.070	0.069	0.070	0.070	0.069
90	0.097	0.085	0.090	0.087	0.076	0.071	0.068	0.070	0.074	0.076	0.073

Table II. Empirically found *non*–coverages of 95% predictive bands with different combinations of q and Q for the S&P500 dataset of Figure 1.

IV. Remarks on usage and performance

To implement the predictive bands, the parameters q and Q must be chosen, i.e., the window size of the Moving Average for the centerline, and the window size for the rolling estimate of variance. As mentioned before, choosing q involves a difficult trade-off of bias and variance, and is best left to the practitioner; for details on optimally choice of a smoothing parameter see Wand and Jones (1994), Hart (1997), or Fan and Yao (2005) and the references therein. Note though that on top of the usual bias-variance trade-off in obtaining a good point predictor, some consideration on the width of the prediction interval is also in order. The interval's width increases in length as q increases. To see this, note that $\hat{\sigma}^2_{pred,q}$ is an increasing function of q; as a matter of fact, $\hat{\sigma}^2_{pred,q}$ is asymptotic to a linear function of q, i.e., $\hat{\sigma}^2_{pred,q} \simeq \hat{\sigma}^2(1/2 + q/3)$ for large q.

Tables I and II show the practical effect of choice of the parameters q and Q on the properties, i.e., width and empirical coverage level, of our 95% predictive bands. In particular, Table I shows the empirically found *widths* of the 95% predictive bands with different combinations of q and Q from the S&P500 dataset of Figure 1. The increase of the width as q increases is very prominent, and should dissuade the practitioner from using large qs; interestingly, there is also a slight increase in the width as Q increases.

We now focus attention on the choice of Q supposing that q has already been chosen. Recall that for 95% predictive bands, approximately 95% of the data points should be covered by the bands. This is theoretical though, and in practice some deviations are expected. It is of interest to see the effect of the choice of q and Q on the actual coverage of the bands. Table II shows the actual *non*–coverage of our 95% predictive bands with different combinations of q and Q.

Ideally, all the entries of Table II should be close to 0.05. Inspecting the table, one can try to find the entry closest to 0.05 in each row; this may give an indication on how to choose Q in relation to the already chosen q to *calibrate* the bands for better coverage accuracy. From Table II the simple relation $Q \simeq 7 + 0.3q$ is suggested that can be used in connection with the S&P500 dataset. However, a different dataset may (and most probably will) require a different relation of Q to q for proper calibration.

To summarize, choosing q is difficult trade-off problem that requires expertise on the part of the practitioner; from the point of view of predictive bands, we simply warn against unnecessarily large values of q since they result into bands that are too wide. Choosing Q, however, seems less critical and can be related to the (previously chosen) value of q via the proposed notion of coverage level calibration.

Summary

A method is given to construct volatility bands that are at the same time predictive bands having a pre-specified level of predictive coverage. The bands are easy to construct with basic spread-sheet calculations, and can be used where-ever Bollinger bands are used; notably, the latter lack any predictive validity. Finally, a discussion is given on choosing the band parameters q and Q, i.e., the window sizes for the Geometric Moving Average and the rolling estimate of variance.

References

[1] Bachelier, L. (1900). *Theory of Speculation*. Reprinted in *The Random Character of Stock Market Prices*, P.H. Cootner (Ed.), Cambridge, Mass.: MIT Press, pp. 17-78, 1964.

[2] Bollerslev, T., Chou, R. and Kroner, K. (1992). ARCH modelling in finance: a review of theory and empirical evidence, *J. Econometrics*, 52, 5-60.

[3] Bollinger, J. (2001). *Bollinger on Bollinger Bands*, McGraw-Hill, New York.

[4] Engle, R. (1982). Autoregressive conditional heteroscedasticity with estimates of the variance of UK inflation, *Econometrica*, 50, 987-1008.

[5] Fama, E.F. (1965). The behaviour of stock market prices, *J. Business*, 38, 34-105.

[6] Fan, J. and Yao, Q. (2005). *Nonlinear Time Series: Nonparametric and Parametric Methods, (2nd Ed.)*, Springer, New York.

[7] Hart, J.D. (1997). *Nonparametric Smoothing and Lack-Of-Fit Tests*, Springer, New York.

[8] Mandelbrot, B. (1963). The variation of certain speculative prices, *J. Business*, 36, 394-419.

[9] Politis, D.N. (2003), A normalizing and variance-stabilizing transformation for financial time series, in *Recent Advances and Trends in Nonparametric Statistics*, (M.G. Akritas and D.N. Politis, Eds.), Elsevier (North Holland), pp. 335-347.

[10] Politis, D. N. (2004). A heavy-tailed distribution for ARCH residuals with application to volatility prediction, *Annals Econ. Finance*, vol. 5, pp. 283-298.

[11] Politis, D.N. (2006), Can the stock market be linearized? Preprint available from `http://repositories.cdlib.org/ucsdecon/2006-03.`

[12] Shephard, N. (1996). Statistical aspects of ARCH and stochastic volatility. in *Time Series Models in Econometrics, Finance and Other Fields*, D.R. Cox, David V. Hinkley and Ole E. Barndorff-Nielsen (eds.), London: Chapman & Hall, pp. 1-67.

[13] Wand, M.P., and Jones, M.C. (1994). *Kernel Smoothing*, Chapman and Hall, New York.