Forecasting Financial Volatility Using High-Frequency Data

Martin Martens

Good volatility forecasts are important for risk management, portfolio management and option pricing. Until recently it appeared that for example daily or monthly volatility forecasts from the popular GARCH model performed poorly. This has all changed with the introduction of high-frequency data into measuring volatility, proving that GARCH models do provide good volatility forecasts. The high-frequency data can also be utilised to produce volatility forecasts that are in turn superior to the GARCH forecasts. This is undoubtedly the most exciting development in volatility forecasting since the introduction of GARCH models in 1986.

Measuring volatility

The problem with volatility forecast evaluation is that even ex-post the volatility is unobservable. Ideally we would like to observe the continuous path of volatility to compute, for example, the ex-post daily volatility,

$$(1) \qquad \int_0^1 \sigma_u^2 du$$

In practice this is impossible. Until recently ex-post daily volatility was measured by the squared daily return. This is an unbiased but also extremely noisy estimator of the true ex-post volatility. An extreme example is a volatile day where the closing price is identical to the previous closing price. In that case the daily return is zero, and hence the squared daily return is zero. Obviously the true volatility is non-zero.

Using the daily squared return for volatility forecast evaluation is therefore not the proper way to go. In the popular regression,

$$(2) \qquad r_t^2 = \alpha + \beta \cdot h_t + e_t$$

where r_t is the daily return on day t, and is h_t the volatility forecast based on information up to day t - 1, the parameter β and the regression R^2 will be downward biased. It is therefore not surprising that a typical R^2 is as low as 4%.

Andersen and Bollerslev (1998) suggest that such an apparent poor forecast performance of all popular volatility models is due to a poor ex-post measure for volatility, not due to poor volatility forecasts. They suggest to use high-frequency data to measure ex-post volatility much more accurately. For the 24-hour foreign exchange market they take the sum of 288 squared 5-minute returns, $r_{t,i}$,

$$(3) \qquad \sum_{i=1}^{288} r_{t,i}^2$$

Both theoretically and empirically Andersen and Bollerslev show the ex-post volatility measure in equation (3) is far superior to the daily squared return. Utilizing this measure on the left-hand side in equation (2) instead of the daily squared return gives a completely different picture into the adequacy of popular volatility forecasting models. The same daily GARCH(1,1) volatility forecasts that initially yielded a regression R^2 less than 5%, now have an R^2 in excess of 40% for the DEM/USD and YEN/USD exchange rates.

Theoretically the highest possible frequency in equation (3) will provide the most accurate volatility measures. Practically, however, one has to consider market microstructure effects such as bid-ask bounce. The 5-minute frequency is often considered to be a reasonable compromise.

Whereas for the 24-hour foreign exchange market the

	S&P 500 index		YEN/USD		Sweet, Crude Oil	
	r_t^2	v_t	r_t^2	v_t	r_t^2	v_t
Mean	1.217	1.201	0.670	0.708	4.139	4.549
Std dev.	3.225	1.775	2.387	1.125	7.908	3.985
Min	0.000	0.063	0.000	0.033	0.000	0.485
Max	59.379	33.327	68.277	27.911	125.882	45.507
ρ_1	0.235	0.479	0.103	0.549	0.077	0.452
ρ_5	0.151	0.389	0.060	0.218	0.095	0.280
ρ_{10}	0.095	0.270	0.000	0.140	0.033	0.241

Table 1: Sample characteristics ex-post volatility measures

expression in equation (3) can be used, for other markets this is often not the case. The US stock market, for example, only trades from 9:30 through 16:00 New York time. For individual stocks Andersen, Bollerslev, Diebold and Ebens (2001) suggest to use the sum of all the squared 5-minute returns from 9:30 to 16:00, ignoring the overnight volatility from 16:00 through 9:30. Martens (2002), for the S&P 500 index, suggests the use of the sum of all the squared 5-minute returns from 9:30 to 16:00 as well as the squared close-to-open return. In the same study I use overnight futures trading to also cover the overnight period. The measure for ex-post volatility is then

$$v_t = \sum_{n=1}^{N} (r_{t,n}^N)^2 + \sum_{d=1}^{D} (r_{t,d}^D)^2 \quad (4)$$

where $r_{t,n}^N$ is the 30-minute return on day t in intranight interval n, and $r_{t,d}^D$ is the 5-minute return on day t in intraday interval d. During the night the 30-minute frequency is preferred due to thin trading. The measure in equation (4) is superior to measures that ignore the overnight information as in Andersen et al.

Data

For this article I constructed the volatility measures in equation (4) for three of the largest futures contracts in the world, the S&P 500 index-futures contract, the YEN/USD futures contract, and the Sweet, Crude Oil futures contract. In all three cases overnight futures trading is available. All three series run until the end of 2000, and each series starts from the first day overnight trading was introduced. Overnight futures trading is especially crucial for the YEN/USD and Oil futures data, as in excess of half the volatility originates outside U.S. trading hours. Using the squared close-to-open return for the overnight volatility would run into similar noise problems as using the squared close-to-close return for daily volatility. NYMEX introduced overnight trading for Sweet, Crude Oil futures in June 1993. The CME introduced overnight futures trading for the S&P 500 in January 1994 and for the YEN/USD in January 1996. In the context of this article it is interesting to look at the sample characteristics of realized volatility as measured by equation (4) as well as the daily squared return. Remember that the latter was used as realized volatility until very recently. The sample characteristics are provided in Table 1. There are some obvious differences between the squared return and the realized volatility based on high-frequency data. First, the squared return series are more volatile and have a larger min-max range. Second, the autocorrelation in volatility is much more present in the v_t series. The reason is that these series have much less noise than the squared return. In fact, Andersen and Bollerslev argue that ex-post volatility has become more or less observable.

Forecasting volatility

The realized volatilities based on high-frequency data have particular characteristics that suggest a good forecasting model. Andersen, Bollerslev, Diebold and Labys (2001) show that the time series v_t exhibits the long memory characteristic. This means that a shock to the volatility stays very long in the system. In particular the autocorrelations decrease to zero only in a hyperbolic way. In comparison, the autocorrelations of a unit root process, I(1), are close to one as the system has a unit root. The autocorrelations of an I(0) process, if present, will go to zero very quickly. A long memory process is a fractionally integrated process, I(d), with $0 < d < 1$. Andersen et al. (2002) suggest the following procedure to forecast volatility. First, transform the realized volatilities, v_t, by taking the square root and then the logarithm,

$$y_t = 0.5 * \ln(v_t) \quad (5)$$

The rationale for this is that the volatility is positively skewed whereas the log volatility follows approximately a normal distribution. In addition the impact of outliers is reduced.

Next, estimate an ARFIMA(p, d, 0) model,

(6) $A(L)(1 - L)^d(y_t - \mu) = \varepsilon_t$

where L is the lag operator, d is the long memory parameter, and $A(L)$ captures any remaining short-run dynamics through an AR(p) model. First, the parameter d is be estimated using the Geweke and Porter-Hudak (1993) estimator. Subsequently the series y_t is filtered to obtain an I(0) process,

(7) $x_t = y_t - dy_{t-1} + d(d-1)y_{t-2} / 2! - d(d-1)^* (d-2)y_{t-3} / 3! + \ldots$

Hereby it is necessary to cut off the sequence, as otherwise no data remain to estimate the short-run dynamics. The cut off here is 100 lags.

For the new series, x_t, an AR(p) model is estimated, which can be translated into $A(L)$. This completes the model for y_t, which can be rewritten into an AR(∞) representation,

(8) $$(y_t - \mu) = \sum_{k=1}^{\infty} \pi_k (y_{t-k} - \mu) + \varepsilon_t$$

where for μ simply the sample mean is used, and the parameters π_k are computed from the estimated d and the AR parameters. For example, if d = 0.4, and p = 1 with the AR(1) coefficient estimated at a_1 = –0.05. Then $\pi_1 = d + a_1$ = 0.4 – 0.05 = 0.35, $\pi_2 = -d(d-1) / 2! + d * a_1$ = 0.12 – 0.02 = 0.10, etceteras.

Equation (8) then provides a forecast for y_t, given data from 1 … t - 1. Finally, a forecast for the volatility (variance) v_t is produced using equation (5),

(9) $$v_t = e^{2(y_t + \frac{1}{2}\mathrm{var}(y_t))}$$

where the adjustment by var(y_t), the variance of y_t, is standard when y_t is normally distributed. After all E[e^y] π $e^{E(y)}$ where E[.] is the expectation operator.

For all 3 series (S&P, YEN/USD and Oil) the above procedure is followed initially for the first 500 data to produce a forecast for day 501. Then every day a new observation is added to the in-sample period, the model is re-estimated and a new forecast is produced. The model also allows to produce forecasts at any desired horizon. For example, to forecast two days ahead the one day ahead forecast is used in addition with the in-sample data in equation (8).

To appreciate the quality of volatility forecasts from the high-frequency data, I also produce forecasts from the standard daily GARCH(1,1) model,

(10) $$\begin{aligned} r_t &= c + u_t \\ u_t \mid \Psi_{t-1} &\sim N(0, \sigma_t^2) \\ \sigma_t^2 &= \omega_0 + \omega_1 \cdot u_{t-1}^2 + \omega_2 \cdot \sigma_{t-1}^2 \end{aligned}$$

Results

The forecast performance is evaluated using an amendment to equation (2), replacing the daily squared return by the sum of intraday squared returns:

(11) $v_t = \alpha + \beta\, h_t + e_t$

Table 2 reports the regression R^2 from equation (11), where h_t is either the GARCH volatility forecast or the Long Memory (LM) forecast. It is obvious that in all cases the long memory model produces superior forecasts to that provided by the popular GARCH model. The general level of forecastibility for the YEN/USD is a bit lower than for the S&P and Sweet, Crude Oil. This is due to the fact that the out-of-sample period, from 1998 to 2000, includes the Asian currency crisis.

	S&P 500 index		YEN/USD		Sweet, Crude Oil	
Horizon	GARCH	LM	GARCH	LM	GARCH	LM
1 day	0.213	0.286	0.219	0.295	0.098	0.183
5 days	0.318	0.462	0.237	0.286	0.176	0.318
10 days	0.328	0.481	0.239	0.268	0.191	0.327
20 days	0.370	0.427	0.271	0.283	0.229	0.375

Table 2: Regression R^2 from regressing realized volatility on the volatility forecast from either the daily GARCH model or a Long Memory (LM) model for realized volatility.

Conclusion

Recent developments in the volatility forecasting literature suggest the use of high-frequency data to both measure and predict volatility. The high-frequency data, say 5-minute returns, are simply aggregrated into one volatility measure per day. Obviously it would be possible to use all transactions during a day and specifically modelling the features of such data, but this dramatically increases the complexity of the analysis with no certainty of any improvement. A long memory model for the daily realized volatilities yields superior forecasts compared to more common volatility models based on daily return data, such as a GARCH model and historical volatility including Riskmetrics^{TM} which uses exponentially declining weights. In my current research I compare the long memory volatility forecasts with option implied volatilities. The completed work on the S&P 500 and YEN/USD futures and futures options suggest that long memory forecasts can compete with implied volatilities and sometimes even outperform implied volatilities. Both long memory volatility forecasts and implied volatilities have incremental information the other does not have, suggesting a combination of the two as a superior volatility forecast.

Whereas currently a lot of time and effort is required to produce the realized volatilities in equations (3) and (4), I predict it is a matter of time high-frequency data will be more commonly used to measure and predict volatility. GARCH models, which treat volatility as a latent variable, will cease to exist. It is time to turn on-line data feeds directly into stored daily volatility measures.

Finally it is important to point out that the same principles apply to the covariance between two assets. This is important for portfolio management. For an example of the characteristics of covariances computed from high-frequency data, see Andersen, Bollerslev, Diebold and Ebens (2001). The main challenge ahead for the multivariate case is to guarantee that the variance-covariance matrix is positive-definite.

Personalia

Martin Martens is UHD at the Econometric Institute and Finance Department at Erasmus University Rotterdam.
E-mail: `mmartens@few.eur.nl`

References

Andersen, T.G., T. Bollerslev, 1998. *International Economic Review* 39, 885-905.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and H. Ebens, 2001. The distribution of realized stock return volatility. *Journal of Financial Economics* 61, 43-76.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2001. The Distribution of Realized Exchange Rate Volatility. *Journal of the American Statistical Association* 96, 42-55.

Andersen, T.G., T. Bollerslev, F.X. Diebold, and P. Labys, 2002. Modeling and Forecasting Realized Volatility. Working paper University of Pennsylvania.

Geweke, J. and S. Porter-Hudak, 1993. The Estimation and Application of Long Memory Time Series Models. *Journal of Time Series Analysis* 4, 221-238.

Martens, M., 2002. Measuring and Forecasting S&P 500 Index-Futures Volatility using High-Frequency Data. Forthcoming *Journal of Futures Markets* 22.