Overview and Facts

Christopher G. Lamoureux

January 16, 2013

Time-Varying Volatility

The fact that stock return volatility is not constant over time was noted by Mandelbrot (1963, p. 418). He documented a temporal clustering phenomenon. Nevertheless from 1963 through 1986 the study of stock returns primarily focused on the marginal distribution of returns.

Examples

- Clark, *Econometrica* 1973.
- Blattberg and Gonedes *Journal of Business 1974*.
- Tauchen and Pitts *Econometrica* 1983.

Discussion

Generally, the lower the frequency of returns, the better fit the Gaussian density. So monthly returns are more "normal" than daily returns. This fact works against the hypothesis (Mandelbrot) that returns follow an unconditional stable distribution.

Stock returns look different at high frequencies because of the trading mechanism that generates the ticker. We have intuition that such "distortions" matter a lot for the continuous price path, but that they are integrated out at monthly, even weekly frequency.

Important early analyses of the trading mechanism and its implications for price dynamics include Niederhoffer and Osborne (1966) and Roll (1984). These papers are the forebears of the market microstructure literature.

GARCH

Engle (1982 *Econometrica*) extended heteroskedasticity-consistent variance estimation to the time-series setting. Because ARCH neatly modeled the volatility clustering that Mandelbrot had described, it became widely adapted as a model for speculative price dynmamics. Especially with the generalization of Bollerslev (1986 *Journal of Econometrics*).

$$
\begin{aligned}
r_t &= \epsilon_t \\
\epsilon_t &\sim \mathcal{N}(0, h_t) \\
h_t &= \gamma_0 + \gamma_1 \epsilon_{t-1}^2 + \gamma_2 h_{t-1}
\end{aligned}
$$

GARCH Discussion

There are several lingering misconceptions and misunderstandings about GARCH.

- GARCH is *not* a stochastic volatility model.
- Like many "time-series models" it is cast in discrete time, so its adding up properties are unclear.
- Dan Nelson's work on EGARCH addresses the above concern and ensures positive coniditional variance on each date.

The most successful extension to GARCH is the Glosten, Jagannathan, and Runkle (1994) model:

$$
\begin{aligned}
r_t &= \epsilon_t \\
\epsilon_t &\sim \mathcal{N}(0, h_t) \\
h_t &= \gamma_0 + \gamma_1 \epsilon_{t-1}^2 + \gamma_2 h_{t-1} + \gamma_3 I_{t-1} \cdot \epsilon_{t-1}^2
\end{aligned}
$$

Where I_{t-1} is 0 if $\epsilon_{t-1} \geq 0$ and 1 if $\epsilon_{t-1} \leq 0$.

Estimation

The GARCH model is a nice setting to learn MLE using the Berndt, Hall, Hall, and Hausman (1974) algorithm.

- Must work with log-likelihood.
- Construct the Score matrix.
- Step size algorithm.

Problems

We typically estimate GARCH(1,1) on a fairly long series of daily returns; (5 years).

Forecasts from GARCH tend to overstate the persistence of recent shocks. Lamoureux and Lastrapes (1991, *JBES*) use the same logic as Perron (1991) (who looked at means) to explain this.

We generally find that γ_3 in the GJR specification is positive: Black (1976); Christie (1982). Interpretation of the *leverage effect*.

Whither GARCH

- In retrospect expectations for GARCH *as a model of economic behavior* were too high.
- Do we have a better understanding of why we see volatility clustering?
- Most of the empirical analysis of time-varying volatility since 2000 is in the realm of *realized volatility*. This relies on the quadratic variation theorem, and the availability of transactions data. The issue here is that the "microstructure noise" becomes very important, but is largely viewed as a nuisance.

Realized Volatility

Consider the problem of estimating the mean and variance of stock returns given a specific time period (e.g., January 1, 2010 - December 31, 2012). Merton (1980 *JFE*) shows that using higher frequency data affords a more precise estimate of the variance but not of the mean. This point is related to the quadratic variation theorem.

For X an Itô process:

$$X_t = X_0 + \int_0^t \sigma_s d\omega_s + \int_0^t \mu_s ds$$

$$[X]_t = \int_0^t \sigma_s^2 ds$$

Realized Volatility

With high frequency data, this would seem to yield a nice estimator of the variance in any time period. The fly in the ointment is that we do not actually observe X. A nice paper on how we might think of integrating out the "microstructure noise" using high frequency data is Zhang, Mykland, and Aït-Sahalia (*JASA* 2005).

ZMA note that a standard solution to integrating over the microstructure noise is to sample the data at a lower frequency (e.g., 5 minutes). But this throws away valuable information. They go through 5 estimators to motivate their optimal estimator.

We observe:

$$Y_{t_i} = X_{t_i} + \epsilon_{t_i}$$

ZMA Estimator 1

Ignore the problem. Then:

$$\sum_{t_i, t_{i+1} \in [0,T]} \left(Y_{t_{i+1}} - Y_{t_i} \right)^2 = 2nE(\epsilon^2) + O_p(n^{\frac{1}{2}})$$

If we sample the observed price every second, then for one day, n=23,400. And this estimator has virtually no relationship to the quadratic variation in X. (Note that it diverges to ∞ in n.)

ZMA Estimator 2

This is a "sparse" estimator. ZMA provide the example of a researcher who has the 23,400 observations of Y, but throws away 299 of every 300. (Keeping data sampled at a 5-minute interval, so $n_{\text{sparse}} = 78$.)

ZMA show that this is biased (the expected value is $[X_t] + 2n_{\text{sparse}}E(\epsilon^2)$, and it has a large variance arising from both the noise and discretization.

ZMA Estimator 3

Instead of choosing an arbitrary n_{sparse} choose an optimal n^*_{sparse} that minimizes a mean square error of $[Y]^{\text{sparse}}$. As ZMA discuss the intuition is that the smaller is $E(\epsilon^2)$, the more frequently one should sample. Intuitively, as the frequency is lower, the bias due to the noise is reduced, but the inefficiency due to discretization becomes larger.
(Of course, 5 minutes is the recommended sampling interval from several practical exercises on estimating realized volatilities on financial data.)

ZMA Estimator 4

This estimator might use n^*_{sparse}, and does not throw away data. So we would use every 5-minute interval in the day. This remains a biased estimator. (In fact for the manner I describe the bias is the same as for Estimator 3.)

ZMA Estimator 5

ZMA show that the best estimator of $[X]$ given Y is constructed using two time scales, *all* and *average*. Based on what we have seen already, by using the entire sample, we can get an estimate of $[Y]$. So the intuition here is to form $[Y]$ using an average time scale, and then subtract $[Y]$ constructed from the full time scale.

Hull and White (1987)

As noted, we see empirically that implied volatilities move around over time.

This is enough to reject the Black and Scholes model.

Hull and White (*JF* 1987) is the first model of options on a stock with stochastic volatility. They use the Black and Scholes structure by assuming that the volatility risk is not priced.

Thus the absence-of-arbitrage value of a European call option is the integral of the Black-Scholes formula taken over the volatility process over the option's remaining life.

Implied volatilities will therefore be (potentially) meaningless (Jensen's Inequality).

But, at-the-money options are close-to-linear in volatility. (Feinstein's 1987 Yale thesis; Lamoureux and Lastrapes 1993).

Heston 1993

Heston (1993) has several important extensions to Hull and White.

1. Volatility risk is priced.
2. Volatility follows a Bessel process.
3. Volatility and the stock price may be correlated. (This may also be true in HW, but it's a bit informal.)

Under Heston's model the evolution of stock returns under the actual probability measure P is:

$$dS_t = \mu S_t dt + \sqrt{v} S_t dz_t^P$$

$$dv_t = \kappa(\theta - v_t)dt + \sigma\sqrt{v_t}d\omega_t^P$$

The instantaneous correlation between the two Brownian motions (dz_t^P and $d\omega_t^P$) is ρdt.

Warnings

Note that Heston's model is cast in continuous time. The volatility process is the same process used by Cox, Ingersoll, and Ross in their seminal term structure model. It has the advantages of being positive and exhibiting mean reversion and heteroskedasticity.

It is popular because since Feller, integrals are known. However, discretization is not trivial (Broadie and Kaya *OR* 2006). See Lamoureux and Paseka (2009) for formalities.