Kent Osband

FINFORMATICS

Samurai GARCHkillers

In which the ubiquity of GARCH gets a swift Katana to the Bento box

Having long admired Julius Irving's abilities as a basketball player, the world was even more impressed to discover his talents as a novelist. *The World According to GARCH*, written in 1978, was an instant bestseller; and the screen version in 1982 was just as popular. Soon economics departments joined in, adapting GARCH to their purposes. It is now the favored technique for estimating the volatility of financial time series. But I think things have gotten way out of hand.

In case you got your financial quant training on Mars, let me note that GARCH stands for Generalized Autoregressive Conditional Heteroskedasticity. To translate, skedasticity refers to the volatility or wiggle of a time series. Heteroskedastic means that the wiggle itself tends to wiggle. Conditional means the wiggle of the wiggle depends on something else. Autoregressive means that the wiggle of the wiggle depends on its own past wiggle. Generalized means that the wiggle of the wiggle can depend on its own past wiggle in all kinds of wiggledy ways.

On merit economists could just as easily call these WWWWW processes, and perhaps they should have. But nowadays that would confuse things with the World-Wide Web. So GARCH it will remain. Only that's so vague — kind of like eating "flavorful" ice cream—that folks tend to add or subtract more letters for clarification. There's threshold GARCH or TARCH, power ARCH or PARCH, exponential GARCH or EGARCH, and component GARCH or CGARCH, and various combinations thereof.

Also, like their cousin ARIMA modelers, GARCHists are entitled to add a (p,q) suffix or vice-versa to indicate the number of lags of different types. As usual in econometrics, more is always better except when it's too much, but since it usually is, people tend to stock with (1,2) or (2,1) or even the lowly (1,1).

In every case, the point is to write down a tracking equation where the left hand side that has to be explained is variance or a function of variance like square root or logarithm, and the right hand side that has to explain it uses lagged values thereof mixed in with new noise. Lo and behold, if you use GARCH for estimation you will predict that the wiggle of the time series wiggles to its past wiggle in all kinds of wiggledy ways.

This discovery came as a great relief to classical finance theorists. Before GARCH, nearly everyone tried to explain nearly everything using Brownian motion, aka white noise, aka lots of independent identically distributed disturbances that bump into each other, one after another, long into the statistical night. But all that identical stuff makes the true variance per unit time constant, while the independent stuff makes Brownian motion distributed normally without kurtosis. And that in turn makes the standard deviation of the measured variance a measly $\sqrt{2/N}$ times the true variance, where N is the number of observations. So sad, so true. The measured variance of white noise is so incredibly smooth and boring that nobody could identify it with real financial time series even when they squinted really hard in poor light.

So you can easily appreciate the jubilation with which GARCH has been received in financial econometrics. But how much does it really explain? Nothing. That's right. N-O-T-H-I-N-G-A-R-C-H. It's just a way to fit empirical data. It's like using Ptolemaic epicycles to track the perceived rotation of sun and planets around the earth. Add enough circles within circles and you can approximate an ellipse centered somewhere else.

Explaining the Wiggledy Waggle

To really explain the wiggledy waggle of the wiggle you need to start with the fundamentals. Analyze how they tend to evolve over time. Discount back their values to the present and roll them into a price, the price of the asset that's expected to pay those future dividends. Then analyze the volatility and the volatility of the volatility. GARCH by itself does none of that.

But finformatics does. In a recent article I showed that the differential of a cumulant of beliefs varies with the next-higher cumulant, all the way up the infinite cumulant ladder. So far the only cumulant we've bothered investigating is the first cumulant or mean. The second cumulant is the variance. Let's look at its differential in more detail:

To review previous terminology, μ refers to the various possible values of drift, E to their forecast mean $\langle\mu\rangle$ where $\langle\cdot\rangle$ is the standard physics notation for expectation, σ to the volatility of fundamentals dx (typically viewed as the percentage change dD/D of dividends D) and dW to the "normalized" fundamentals $\frac{dx - Edt}{\sigma}$ having unit volatility and zero expected drift. Fair market value is denoted V. Drifts can switch from value ν to value μ with instantaneous probability $\lambda_{\nu\mu}$ (i.e., approximately $\lambda_{\nu\mu}dt$ in a small interval dt), with the additional convention $\lambda_{\mu\mu} = -\int_{\nu\neq\mu}\lambda_{\mu\nu}d\nu$ to indicate the (negative) instantaneous probability of leaving drift μ. As two more shorthands I will use Var to denote $\left\langle(\mu - E)^2\right\rangle$, the variance of beliefs, and $Skew$ to denote $\frac{\left\langle(\mu - E)^3\right\rangle}{Var^{3/2}}$, the skewness of beliefs. Then, drawing on results derived in previous articles, the learning processes for E and Var can be expressed as:

$$dE = \left(\int \langle \lambda_{\nu\mu} \rangle_\nu (\mu - E) d\mu \right) dt + \frac{1}{\sigma} Var\, dW$$

$$\frac{dVar}{Var} = \frac{\int \langle \lambda_{\nu\mu} \rangle_\nu (\mu - E)^2 d\mu}{Var} dt + \frac{1}{\sigma} Skew \sqrt{Var}\, dW$$

With two regimes, we have seen that the fair value price-to-dividend ratio is affine in E.; that is, it equals $a + bE$ for suitable constants a and b. With multiple regimes, E might not always map to a single price-dividend ratio, but in general the relation should be roughly affine. This implies that

$$\begin{aligned} \frac{dV}{V} &\cong \frac{dD}{D} + \frac{bdE}{a+bE} = Edt + \sigma dW + \frac{b}{a+bE} dE \\ &= \left(E + \frac{b}{a+bE} \int \langle \lambda_{\nu\mu} \rangle_\nu (\mu - E) d\mu \right) dt + \left(\sigma + \frac{b \cdot Var}{(a+bE)\sigma} \right) dW \end{aligned}$$

Price volatility, which I will abbreviate as *Pvol*, is given by the multiplier on dW, namely $\sigma + \frac{b \cdot Var}{(a+bE)\sigma}$. Define Ψ as the fractional excess vol compared to fundamentals; that is, $\Psi \equiv \frac{Pvol}{\sigma} - 1 = \frac{b \cdot Var}{(a+bE)\sigma^2}$. Its differential is given by

$$\begin{aligned} dPvol &= Zdt + \frac{b \cdot Var}{(a+bE)\sigma} \cdot \left(\frac{dVar}{Var} - \frac{b \cdot dE}{a+bE} \right) \\ &= Z^* dt + (Pvol - \sigma) \left(\frac{1}{\sigma} \right) \left(Skew \sqrt{Var} - \frac{b \cdot Var}{a+bE} \right) dW \\ &= Z^* dt + \Psi (Skew \sqrt{Var} - \Psi \sigma^2) dW \end{aligned}$$

where Z and $Z*$ include drift terms arising from Ito's rule that I'm not bothering to write out. For an alternative formulation in terms of price variance $Pvar \equiv (Pvol)^2$,

$$\frac{dPvar}{Pvar} = 2\frac{dPvol}{Pvol} = \frac{Z^*}{Pvol} dt + \frac{Pvol - \sigma}{Pvol \cdot \sigma^2} (Skew \sqrt{Var} - Pvol \cdot \sigma + \sigma^2) dW$$

In words, the logarithm of price volatility will have some elements that are approximately white noise and some that move inversely with volatility. This is broadly speaking what the EGARCH formulation says. Interestingly. Hamilton's 1994 bible on Time Series Analysis notes (p. 672) that EGARCH is one of the best-fitting GARCH models.

The comparison to EGARCH presumes that pesky $Skew\sqrt{Var}$ doesn't get in the way. Let's work it out for the simplest possible case, namely a two-regime world with drifts $\pm\mu$ and believed probability p of drift $+\mu$ .There $Var = 4\mu^2 p(1-p)$ and $Skew = \frac{1-2p}{\sqrt{p(1-p)}}$, implying

$$Skew \sqrt{Var} = 2\mu(1-2p) = \mp \sqrt{\mu^2 - Var}$$

Hence, far from undermining the comparison, the skewness term reinforces it as it decreases in absolute value with the uncertainty in beliefs.

In addition, our derivation leaves ample scope for asymmetric responses. To begin with, the multiplier on dW is biased negative by the $-\Psi\sigma^2$ term. Hence, all else being equal, volatility will tend to be more volatile when the regime is perceived to be bad. That effect stems from the same dE passthrough effect of dividends on price despite a lower price-to-dividend base. In other words, dP is approximately the same but P is lower when news has been poor, so that $|dP/P|$ is higher.

For the same series of absolute shocks at the same price the negative shocks will tend to be more damaging if bad regimes are less common than good ones. The demonstration would require more analysis of the dt terms than I am prepared to enter here. A previous article, "Iceberg Makers", probed the general phenomenon. Basically, when bad regimes are less common, bad news tends to come as more of a surprise and require more adjustment. Again, empirically these effects are widely noted in the GARCH literature, though the explanation provided there is Ptolemy's.

The SAMURAI Alternative

Hence, GARCH is probably best viewed as an empirical approximation to the random regime-switching process described by finformatics. As I mentioned earlier, it's not a bad approximation. Only it's not all that useful, other than as a subject for journal articles, like making sculptures of manikins when real nudes aren't available.

Here's the rub. Simpler approximations do nearly as well as GARCH and often better. I'm going to present you readers a few approximations here, and invite you to test for yourselves.

The simplest approximation is an exponential moving average or *EMA*. In theory, *EMA*s are enormously complicated, because they sum up an infinite number of lagged observations with weight $\lambda(1-\lambda)^{n-1}$ on the n$^{\text{th}}$ closest observation. In practice, *EMA*s are a snap to update recursively: just calculate the new *EMA* score as the old score plus λ times the difference between the new observation and the old score. Using the operator L to denote lags, we can thus compute an *EMA* Y as

$$Y = \lambda(x - LY) + LY = \lambda x + (1-\lambda) LY$$

By setting $\lambda = \frac{2}{N+1}$ we obtain an *EMA* of duration and variance equivalent to a rectangular (equal-weighted) window of length N. Since people find rectangular windows easier to grasp than geometrically declining infinite staircases, let's label this *EMA*(x:N), and apply some lag operator algebra to express this more compactly as

$$Y \equiv EMA(x;N) \Leftrightarrow Y = \frac{\frac{2}{N+1} x}{1 - \left(1 - \frac{2}{N+1}\right) L} = \frac{2x}{N+1-(N-1)L}$$

Then if you want to track the variance of individual US stock returns and aren't overly fastidious, I recommend $EMA(x^2;30)$, where x denotes the daily return.

Actually, since I'm getting to that time of night when all moral scruples loosen, I'm going to reveal something really déclassé, namely that I use $EMA(x^2; 30)$ myself. Oops. My wife was reading over my shoulder and broke into tears, since the doormen at finer derivative bars in New York will no longer let us in. Too late dear, our only options going forward will be vanilla...

But to restore the family dignity, let me tell you an alternative approach that's exactly twice as sophisticated as EMA. Namely, average a short-term EMA with a longer-term EMA. With individual US stocks for I have found that a simple average of $EMA(x^2; 10)$ and $EMA(x^2; 65)$ often performs as well as fancier GARCH methods

When you look under the hood, the similarity is evident. Using lag operators we see that:

$$\theta EMA(x^2; M) + (1-\theta)EMA(x^2; N) = \frac{2\theta x^2}{M+1-(M-1)L} + \frac{2(1-\theta)x^2}{N+1-(N-1)L}$$

$$= \frac{[2(\theta M + (1-\theta)N + 1)(1-L) + 4L]x^2}{(M+1-(M-1)L)(N+1-(N-1)L)}$$

This is basically a GARCH(2,2) specification with several parameter restrictions including zero intercept. The only surprise is that it performs as well as it does. On reflection even that should not come totally as a surprise. Lagged values of estimated variances are so highly autocorrelated that GARCH parameters are rarely well identified. To avoid wild results, GARCHmetricians often impose additional ad hoc restraints.

Still not fancy enough for you? Then try the following super-duper-whooper method. Fix M and N at some reasonable values: 10 and 65, 10 and 260? Then let θ adjust to minimize the mean squared error of estimation. Simple rearrangement shows that this is equivalent to regressing $L^{-1}x^2 - EMA(x^2; N)$ against $EMA(x^2; M) - EMA(x^2; N)$ and using the resulting beta estimate as θ. To avoid cheating we should use only data available at the time, which means the estimated θ will evolve over time. In practice you need to be careful not to let θ evolve too fast or the estimates will "rattle" too much around the optimum. But this is a standard problem when updating estimates. GARCHmetricians typically avoid this only by cheating. They use data from the whole period before assigning the best parameters to the past.

This method creates Self-Adjusting Mixtures Using Recursive Artificial Intelligence, so naturally I abbreviate this as SAMURAI. So now at last you know how this article got its title. If finformatics can't kill GARCH, SAMURAI will!