[SQUEAKING]

[RUSTLING]

[CLICKING]

PETER KEMPTHORNE: Well, today we're going to talk about volatility modeling. And this topic in finance is one that has really interested me for decades. And there are a lot of subtle details involved with understanding volatility, how to measure it, how to forecast it. And we've had some talks already with the group projects about dealing with volatility estimation and measuring volatility, in particular implied volatility.

Today we'll focus more on realized volatility and modeling that. And so to begin, here's just an overview of different ways one might define volatility. And its basic definition is the annualized standard deviation of the change in price or value of a financial security. And of course, this concept is really important for measuring risk and evaluating the value of options on underlying assets. Yep?

AUDIENCE: What do you mean by annualized?

PETER KEMPTHORNE: What that means, OK, suppose we have a sigma day that's equal to, say, 0.01%. So 1% variability per day. If we wanted to measure volatility on an annualized scale, then we could consider possibly the square root of 252 times sigma day. So it's a matter of scaling the volatility up to an annual horizon.

And the value of that rescaling is that we can talk about volatilities over different maturities. And they're measured on the same scale. And with different estimation approaches, there's historical sample volatility measures which don't require really any probability models with them.

Then we have geometric Brownian motion. That's a very familiar model to everyone here. And we can consider extensions of that to the Poisson jump diffusion model. A group delivered a nice lecture note on the Poisson jump diffusions.

And then following these cases are cases where volatility might have some time dynamics in them. With these geometric Brownian motion and Poisson jump diffusion models, we think of volatility being some constant, which is the same from one period to the next. But it turns out that there can be time dependence in volatility.

And so there are ARCH and GARCH models, which stand for Autoregressive Conditional Heteroskedastic Models. And then the GARCH are generalized versions of those. And then there are stochastic volatility models as well. I know, let's see, the Heston option pricing model uses a stochastic volatility term in its formulation. So we have those kinds of extensions. And then finally, there's implied volatility from options and derivatives.

So let's just look at computing volatility from an historical series of actual prices. If we have prices PT, then we can define the log returns by looking at the ratio of prices from one day to the next, or one period to the next, and taking the logarithm of that. This is a useful way of measuring returns as an alternative to simply the percentage change in value using the log scale or the log return, if we look at the cumulative change over successive periods. With log returns, they just add. Whereas if we were using percentage returns, we'd have to basically multiply factors together representing that.

Now, in modeling volatility, we consider models that correspond to stationary processes for our returns. So the log returns of price series are often consistent via tests with a stationary distribution of returns. And under the assumption of covariance stationarity, we can define sigma to be the variance of or the standard deviation of those returns.

Now, in this notation here, this sigma for the volatility would be corresponding to a frequency of observation that's daily or weekly or monthly. And we'd have sample estimates given by the square root of the sample variance where we traditionally use the estimator for variance, which is an unbiased estimate of the variance. So that factor divisor t minus 1 is included there.

And then if we want to annualize these values, it's common practice just to treat these returns as independent or uncorrelated over time, so that the variance of a sum is the sum of the variances. So there are roughly 252 days in a year. So we multiply by the square root of 252 for daily returns. For weekly returns, we'd do the root 52 times sigma hat and route 12 for monthly.

Now, when measuring historical volatility, we can think of a simple historical average since the beginning of our time series. So this historical average would use all available data. It's reasonable to think that volatility isn't necessarily stable over the entire time range of a price series. And so we could consider just using the last m values. So take a rolling average of the last m values of this sigma hat squared t minus j.

Now, if we define sigma hat squared t to just be a single day's volatility, that could be like the square of the daily return on a given day. Now, using equal weighted values over a window of length m might be improved upon by using an exponential moving average.

And so exponential moving averages are averages where we take a weighted average of the previous estimate of volatility and the current estimate of volatility using the time t value. So this 1 minus beta beta weight allows us to weight the past versus the present in an exponential way. And with this formula, it's recursive over t. And so it can actually apply across an entire range of historical data.

An alternative to just the regular exponential moving averages and exponential weighted moving average where we only go back maybe m periods, but we decrease the weights on past volatility estimates according to the power of beta j. So beta j or beta is a decay factor in estimating volatilities.

Additionally, one could consider regression models with these volatility sigma hats indexed by t. And depending on the relationship of volatility to the past, we might use more data to increase precision of estimators, but use data closer to time t when we believe that the volatility may be changing over time.

And what's neat about these estimation methods is that we can evaluate, say, the exponential moving average estimates of volatility over different assets and different asset classes. And this approach was applied in RiskMetrics. Here there's a web link to the RiskMetrics technical document. I think I have it listed here.

So actually this RiskMetrics was created by JP Morgan back in the 1990s. Coincidentally, when I was at the Sloan School teaching statistics and financial modeling, I was part of the Financial Services Research Center that was sponsored by JP Morgan and other groups when this risk metrics document was created.

And if we look at, let's see, there's a section here of risk metrics decay factors for volatility. Basically, for different countries, one can use the exponential moving average of asset returns for foreign exchange 10 year 0 prices equity indices and other assets.

And what's interesting to see is that the decay factor is generally well above 0.9. It's more like 0.95, which suggests that volatility is changing slowly over time. And the way they chose these decay factors was to look at how predictive they were with mean squared error of volatility. So these are reasonable estimates of volatility.

And an important aspect of these measures is having benchmark values that people can agree upon. So we're not arguing that these are the best estimators, but they're a simple objective estimator that can be applied. And when we communicate about these other people, understand what we're measuring. So those are important.

Now, let's see. With geometric Brownian motion, turning now to having a stochastic process model for the prices. If we have this stochastic differential equation, the geometric Brownian motion one, we see that, and it's totally obvious, but if we divide both sides by s of t, we get mu dt plus sigma ds of t or sigma dw of t, where w of t is a standard Brownian motion. So by looking at essentially the percentage change on an infinitesimal level, we basically have a linear trend that's deterministic and then a random walk with Brownian motion steps on that scale.

And with this geometric Brownian motion, we have basically independent increments with normally distributed increments on this log scale for the Brownian motion. And with our homework exercise this past week, we actually dealt with estimating the drift parameter and the volatility parameter in geometric Brownian motion. And so if our increments of time are unit increments, then we can estimate a drift term mu star with the average log return. And we can estimate the period volatility or period variance with the sample variance using n instead of n minus 1.

Now, these essentially correspond to maximum likelihood estimates of these parameters. And what's important to note here is that the drift term mu star per period is actually equal to the mu of the geometric Brownian motion minus the square of the volatility term divided by 2. And so what's important is that when we apply this stochastic differential equation, there basically comes an extra term in the drift that's due to the squared variation in the process.

So in working with these the stochastic differential equations, it's important not to just look at first order terms, but to look at second order terms in your Taylor series approximation. And that's how that mu star gets in there to be different from mu.

Now, an interesting exercise, which hopefully was of interest to people who worked on that last week, if we vary the increments delta j over which we observe this price process, then we can talk about how accurate or how precise are our estimates of mu star and sigma squared.

And it turns out that if the delta j, the increments between observations varies, then our estimate of mu star is always the same. It corresponds to just the cumulative return over the entire period divided by the length of the period. So it's the average rate of return per period.

Interestingly, if we increase the number of increments, then our precision on the variance gets more and more precise, and the variability of that actually goes down with a factor of n. So that was a homework exercise that I think is an important one to realize.

Now, a really interesting extension of maximum likelihood estimation with normally distributed increments coming from the Brownian motion process is looking at the problem of using more information than just the close to close prices. And so what we want to do is introduce you to the Garman-Klass estimator. And let me just turn to the Garman-Klass paper. These papers are distributed in Canvas in the volatility notes section.

But if we think of a period of time with one day, going from yesterday's close to today's open to today's close, then in thinking about measuring volatility using the difference between today's close and yesterday's close, that can be very useful, but it's missing out on other information that we may have access to, such as what's the opening price today? What's the low value today? What's the high value today? And how can we use that extra information to have more precise estimates of volatility?

Anecdotally, when I was studying for a doctorate in statistics at Berkeley, I was a research assistant to Michael Klass. And I was getting interested in non-statistics topics like finance, and so I took a course from Mark Garman, who co-wrote this paper. So it's rather interesting to have had that exposure. Actually, it was a very inspiring set of classes that I took, which motivated my, I guess, lifelong interest in financial modeling.

So let's take a look at some notation. Very simply, we'll think of just a single period, like one day, and refer to the closing price on day j, opening price on day j, the high and low on day j as well. Importantly, the notation here includes this-- well, includes the time points, but it also includes the parameter f, which is the length of time that the market is closed on a given day. So the time between the close and open of the next day.

So we have this notation. And if we consider just focusing on single periods, and in this case, the first one, if we look at the change in closing values, now, perhaps it makes sense to think of these on the log scale of the prices. The close to close squared return basically will have mean sigma squared and variance equal to 2 sigma to the fourth.

Now, this corresponds to the theory of normal random variables for returns, which comes from the Brownian motion. And the expected squared value is-- well, actually, we can think of maybe the mean return being 0 or being subtracted from all the returns.

But the variance of this sigma 0 hat squared is the variance of a chi squared random variable with 1 degree of freedom. And so in probability classes, we learn that a chi squared random variable has expectation equal to the degrees of freedom. And if it's a scaled chi squared, it's the scale factor sigma squared times that. And the variance of a chi squared is equal to twice the degrees of freedom. So that's where that formula comes from.

Now, if we look at the close to open squared return, the same arguments lead to this sigma 1 hat squared. Looking at the squared return from closed to open, normalized by the length of the period between the close and open, this will also have expectation equal to sigma squared. And its variance will also be 2 sigma to the fourth. So if the model applies throughout the time periods of the market being open as well as the market being closed, then these formulas apply.

Now, we could also do the open to close squared return and, again, if we normalize the squared change in price or log price, if we normalize that by the length of the period 1 minus f, then we get the same properties of the moments for these estimates of sigma squared. So all of these have expectation equal to sigma squared, and all of them have the same degree of variation.

Well, this second or the index one and two cases are estimates that use different time periods, different increments, disjoint increments of the process. And so these have to be independent if our Brownian motion model applies.

And so if we simply take the weighted average of these or the simple average, then the expectation is unchanged and the variance is equal to basically 1/2 squared times the sum of the variances. And that equals actually sigma to the fourth. So we're able to get a more precise estimate of the volatility using these two increments separately.

And so it's common in statistical theory of estimators to talk about the efficiency of one estimator versus another. And so if we compare this sigma star, we can divide its variance into the variance of sigma 0 squared and get a factor of 2. So we basically have-- let's see. Have I got that correct? Right. We basically have twice the efficiency in this case. So we're thinking of inverse relationships of the variance.

Now, let's see. Parkinson came up with considering the high and low values over a period and showed that this estimate sigma hat 3 squared is an unbiased estimate of sigma squared when f is equal to 0 and has efficiency equal to 5.2.

So this high and low data really has a lot of information in it compared with just the close to close returns. Our precision with estimating sigma squared basically can increase by a factor of 5 with using the high and low values instead of the close to close.

And let's see. I just posted today the problem set due next week where there's an article by Parkinson discussing issues of high and low or extreme value estimates of the variance, which I think will be interesting to read. I think it's somewhat fun to read the classic papers in these areas, especially because the older they are, the easier they are to read, I think, although sometimes that's not totally true.

Anyway, what Garman and Klass did was they considered alternatives, which included the high-low volatility estimate. And a simple average, or weighted by a and 1 minus of this high-low estimate of Parkinson and the sigma 1, which is the-- let's see. Let's go back. And the sigma 1, which is the open to close value.

This has an even higher efficiency of 6.2 and a rather striking result of their paper, as they were able to derive the best analytic scale invariant estimator. So, depending on whether you like mathematics or not, you may like this paper or not. But it's really interesting to see how a property of finding a scale invariant estimator that is best among the class of such estimators, that's actually derived by Garman and Klass.

So let me just highlight how this section five in their paper talks about best analytic scale invariant estimators. And you may be able to see from the derivation here that their arguments just are dealing with some quadratic forms and solving for properties, minimizing these quadratic forms. And so it's rather neat that they're able to achieve this result. Yeah?

AUDIENCE: Why is it necessary [INAUDIBLE]? Because that's just commenting about the prepared bearings, right?

PETER KEMPTHORNE: Well, with a higher efficient estimator, you can have as precise an estimator with a smaller or shorter time series. So instead of using, say, 20 days close to close, you could use a week of data and have estimators that are comparably precise.

AUDIENCE: [INAUDIBLE]

PETER KEMPTHORNE: Yeah, yeah. That's right. And so there are all these extra assumptions that complicate things. And in many of these developments, they're simplifying assumptions like the mean is 0. But if you were trying to model time varying volatility, then rather than looking at, say, close to close volatility and how that changes over time, if you use these other estimators, the Garman-Klass, for example, you could maybe use a much shorter time period and be able to track or model changes in volatility over time with that. Well, let's see.

So then finally, there's this composite estimator, which has efficiency 8.4. And finally, there are a number of extensions of the Garman-Klass that are listed here. And the one that is actually recommended these days is actually this Yang Zhang model, which has a drift independent volatility estimation based on the high-low close. So the fact that it's drift independent deals with that assumption that was just mentioned.

Now, let's see. There's a case study of using R to calculate these different volatility estimates. And let's see here. So here's the case study, which goes through calculating different measures of volatility, using close to close Parkinson's all the way down to Yang and Zhang in section 1.6 and then comparing those for the S&P 500 Index.

And we'll go through these in a moment. But following just looking at these individually, we consider time series modeling of these different volatility statistics. And so in terms of understanding realized volatility, a useful analysis is to be able to predict how our realized volatility is going to change into the future. So there may be some predictability there.

And what we'll see is that in the time series analysis of these volatilities, time series models such as autoregressive integrated moving average models can be applied and can be very useful. And in particular, there's a role to be played with seasonal ARIMA models. So let's take a look at that.

And just as a preview, let's see, one of the problems in the next problem set is to look at this same set of analysis for the S&P 500, look at the same analysis for the stock ARKK, which is Cathie Wood's ETF, and just see how different the results are. And what's nice is that the R markdown files that create this case study can be applied to any other stocks you want to analyze, if you so choose.

For the problem set, I posted the two versions of this study for S&P 500 and ARKK. So you don't need to do R at all. But if you set up a Posit Cloud account with RStudio Cloud, you should be able to just upload the file, the R markdown file, and rerun the programs and do so with another stock, if you so choose.

So in this note, the first section just goes through and looks at the S&P 500 index. It actually is going through a couple of days ago. It's a good time to have an equity exposure.

And if we look at close to close volatility, this is a graph of the 21 day variance scaled up to an annualized basis by multiplying by root 252. And what's important is to see that volatility generally is well below 25%. Often it's around 12% to 15% when it's low, but it does have spikes. And during COVID, there was the huge spike up. And more recently, there was some elevated volatility that suggests that volatility is not constant.

So anyway, this is the close to close volatility. If we look at the high-low volatility of Parkinson, then we just are averaging the log of the ratio of the high to low and scaling it up. And so we get this plot, which is very, very similar.

In principle, this estimator is more precise, at least if our underlying Brownian motion model or geometric Brownian motion model is true, but it gives us different values. And then the Garman-Klass model is here.

So what's useful just in this case study is partly just seeing what the formulas are for the volatility estimates and then being able to compare those. And then finally, this Yang-Zhang model is an extension of Garman-Klass. And as it says here, this is currently the preferred version of volatility estimates.

Now, in looking at these different volatility estimates, we can create a panel series of all the volatility measures. This shows, not surprisingly, there's peaks at the same times with all of these. What's more interesting is to look how they move simultaneously. And if you look at too much data, it's hard to see exactly what's going on. So there's a graph here of looking at just the last year, almost two years of data, and seeing how these different volatility estimates vary.

So what this highlights is how a choice of volatility measure matters. There can be significant differences between them. And what would you anticipate is a good property of a volatility estimator? That's a very general question, but intuitively, what would make one estimator better than another?

And there's no right or wrong answer to this, but I think it's useful to think about what you would like to have in an estimator, and whether these estimators might be consistent with that or not. Any thoughts?

Well, one of the problems, I guess I'll call it a problem, with the close to close estimator, is that if you have a rolling window of close to close variance estimates, then if you have a spike in returns, positive or negative, that spike is going to remain in that window until it falls off. And so there can be elevated volatilities due in the close to close series, volatility series. That may not be as persistent with the other estimators.

So when there are spikes in volatility, often those spikes occur when there's some great uncertainty in the market that gets resolved. And the volatility often drops a fair amount after that. And so I would think that those volatility estimates that basically drop faster following the spikes could be better. And certainly if we're trying to use these volatility estimates to predict future realized volatility, that property could be useful.

Now, what's interesting in section two of this case study is to just consider some really, really basic time series modeling of the different volatility measures. So what we are modeling is different volatility measures over a 21 day period that's rolling through our time series.

And so one of the most basic forecasting models there is is a model called the naive model. And with a naive model, we have a time series yt. And we basically have that y hat t is equal to yt minus 1. And if we think of yt is equal to y hat t plus an error term, then our errors are equal to yt minus y hat t, which is equal to yt minus yt minus 1.

And this specification corresponds to a random walk model. So there's basically a random walk model for yt. And the random walk model is consistent with efficient markets where our best estimate of what's going to happen is where things are now.

So this naive model basically has an autocorrelation function, which has a very high-- or our data has a high first order autocorrelation. And if we calculate first differences, which are the residuals from this model, then we get this residual diagnostic information, the time series of residuals, the autocorrelation function of the residuals, and the histogram of residuals.

Now, would you say that these residuals are consistent with white noise that may be Gaussian or normal? I see some going mm-hmm and others going hmm. Well, one thing that's important is that the autocorrelation function of residuals, if the residuals are consistent with white noise, there should be no significant autocorrelation at any lag.

And so there are some spikes of autocorrelation that break beyond the critical levels. And so those are suggesting that this is not white noise. And if we look at the distribution of the residuals, they're unimodal, but they certainly aren't well described by a normal distribution.

So let's see. So we can plot forecasts of this naive method. And then let's see. We can also fit an ARIMA model to this close to close volatility series. And with the ARIMA model, actually there's this package in R called FPP2.

And this was created by the authors of a book, *Forecasting Principles In Practice,* that's in the references to this. They have an automated ARIMA modeling function, which will take an input, a time series, as output, what is judged to be a decent autoregressive integrated moving average model. And so this ARIMA 310 corresponds to ARIMA p equals 3, d equals 1, and q equals 0, where p is the order of the autoregression. d is the order of differencing. Yes?

AUDIENCE: How does it [INAUDIBLE]?

PETER KEMPTHORNE: Yes, it uses a criterion that's called the corrected Akaike Information Criterion. And so there's this AICC, which is equal to the corrected AIC. And the corrected AIC is due to the fact that with shorter series, there's a bias in the AIC that is systematic. And so this correct for that bias.

Now, when we do this with any fitted model, what's really important to check is how our estimated coefficients compare with their standard errors. When the signal estimate to noise standard error ratio is bigger than 2, we have some confidence that that's an important factor in the model. And in this case, the AR3 has a signal to noise ratio of almost 3, and the AR2 has a ratio that's 1 and 1/2.

Now, if we go through, let's see, the second section does the same analysis of the Yang-Zhang volatility estimate. And I guess importantly here, we have an ARIMA model now for this Yang-Zhang model, which is an ARIMA 111.

And the coefficient estimates are many multiples of the standard error. So these are highly significant. And this is our residual analysis check, which suggests that, for the most part, the residuals are consistent with white noise. The autocorrelations tend to be close to 0. However, there's this large spike at 21. And here the residuals.

Now, why do you think there's a large spike at 21? Well, we're computing rolling statistics. And so the window is of length 21. So this corresponds to cases entering and exiting the windows.

And so interestingly, we can accommodate that by fitting a seasonal ARIMA model. And so with the seasonal ARIMA model I'll jump to the conclusion here at the very end, which is for the Yang-Zhang volatility estimator. This is its simple autocorrelation function before the model. And then when we fit the seasonal ARIMA model automatically, it gives us a 213 and a 201.

And let's see. When we introduced you to time series models before, let's see, if we have an ARIMA model, which is p, d, q, we basically have 1 minus phi 1 b minus phi 2 b squared minus phi p b to the p power times yt turns out to be equal to 1 minus theta 1b minus theta 2b squared minus theta, or sorry, 1 plus here, plus theta qb to the q power times our error term.

So we have this kind of mathematical equation for an ARIMA model, where we assume that the epsilon t series is white noise, which means that its expectation is always 0 and the variances are constant, and the correlation between any different epsilons is equal to 0 if t is not equal to t prime.

The power of autoregressive integrated moving average models is that these autoregressive terms, as well as moving average terms, can account for all of the correlations, autocorrelations in the original series. And our residuals, which are estimates of this error sequence, should be consistent with white noise.

Now, what ends up happening with-- actually, I should maybe put here yt prime, where y prime is maybe equal to 1 minus b to the power d yt. So if we have an ordered d model, we take the d-th order difference and put that into an ARMA model.

And if we have a seasonal ARIMA model, we actually add a capital P, capital D, capital Q. And here we have the frequency of the data. And in this example, I considered a 21 period frequency just to correspond to the window size.

And what this model does is it has another polynomial of-- so this is b to the 2 times 21. So we basically end up having backshift operator polynomials that work with the seasonal frequencies of the time series.

So this is a polynomial of order q of order p in B to the 12th-- sorry, B to the 21. And then we would also multiply this on this side, perhaps, by a 1 minus capital theta 1 B to the 21, or plus, plus theta 2 B to the 21 squared. So times 2.

And so this kind of notation leads us to these coefficient estimates. So there's the simple order autoregressive terms and moving average terms. Order two, autoregressive, order three, moving average. But then for the seasonal autoregressive terms, we have these terms here. And for the seasonal moving average, we have the last term here. And what's generally important in fitting models is when you consider the last terms to enter into a model, such as in this case, the seasonal moving average term, its estimate is many multiples of its standard error. So that's a very important factor in the model.

As an anecdote, let me just say-- let's see. If we take a series yt and we look at a seasonal difference operator and a first order difference operator, this turns out to be equal to the first difference operator multiplying the seasonal difference operator. And it equals basically 1 minus B plus B to the 21. Sorry, minus B to the 21 plus B to the 22 times yt.

And so with seasonal differences in autoregressive moving averages, whether we take first differences non-seasonally and then seasonal differences subsequently, or reverse the order, we end up getting the same transformation of the data. Well, actually with economic time series, where we have a seasonal frequency of, say, months, then basically the seasonal differences correspond to year over year changes, whereas the simple differences are just period by period differences.

Anyway, with this final example here, this is how the residual analysis looks. And it actually does yield residuals that do look a lot like white noise, although there is still this persistent spike just beyond the window width for the computation. But the distribution of the residuals looks more Gaussian, perhaps, than, well, much more than the cases before.

So with that, in analyzing different time series models, there are these accuracy statistics that can be computed for any forecasting model. And so you're probably familiar with root mean squared error or mean squared error. But we also have the MAPE, which is the Mean Absolute Percentage Error. And I actually find that pretty useful for judging how effective the models are.

And what's really nice about the FPP2 package is that one can just ask it to plot the time series and the forecast, and it generates a point forecast in a solid blue line and prediction intervals about that forecast, which assume Gaussian or normal distributions for the error terms. But one can see how those evolve into the future. And it's really consistent for prediction intervals to get wider the further out we forecast. And that's happening here.

What's rather interesting about this forecast using the Yang-Zhang measure and forecasting that into the future is that it actually, after being at a lower level, is predicting that it might go back to more of a mean level of volatility. So that's an interesting property.

Well, let's go back to the notes here. And let's see. There are extensions of geometric Brownian motion that we have seen before. One of them is a Poisson jump diffusion model, where the stochastic differential equation has the Brownian motion part as the first two terms, but then has a jump process as a last term where the jump process can be modeled as a Poisson process.

So a jump occurs at some rate that defines the Poisson process. And the rate is lambda. And if an event in the Poisson process occurs, then there's a jump in the stock, which is given by this factor here. So one can actually model data with this Poisson jump diffusion model.

I ended up working with David Pickard and Arshad Zakaria on fitting this Poisson jump diffusion model. And it was an interesting application in statistics of using the EM algorithm. Basically, the jumps are treated as latent or hidden values in our data. And those hidden values occur according to Poisson probabilities, like the probability of one jump or two jumps, can be computed using the Poisson parameter. And one can apply this EM algorithm to obtain maximum likelihood estimates of all the parameters.

And in applying this, an interesting property that occurred was that there were quite frequent-- or the rate of Poisson events was quite high, like every couple of days. And what the model was doing was capturing heavier tailed distributions than a normal distribution.

So instead of a single normal model for the return, one has a mixture of a normal with no jumps, a normal with one jump, two jumps, weighted according to Poisson rates. And so the implication of the model fitting was that one didn't have rare events necessarily that occurred, but there was evidence suggesting a jump process that related to trying to capture the excess variability of the stock price returns.

Now, that leads to another model extending Brownian motion. And this is the Laplace distribution, where if instead of observing a time series over equal increments, so between 0, and one can think of just equal increments of time. So t1, t2, and so forth up to tn, say, from time 0 to capital T. It may be that we observe the Brownian motion process at random times t1, and then here's t2, here's t3, where the increments of time themselves are also random.

And it turns out that if the times at which we observe have a time increment that's exponentially distributed, then the increments of the Brownian motion follow a Laplace distribution. So heavier tailed distributions than normal, and this is the density of a Laplace distribution, it depends on the exponential of minus the magnitude of the deviation of x from mu as opposed to the square of that deviation.

A Gaussian or normal distribution would have a power 2 there with the Laplace distribution as the absolute. And with this distribution, we have a mean mu, which is at the center of the distribution. The variance turns out to depend on this scale factor B, which is twice B squared.

So let's just take a look at what Laplace models look like. They basically are bilateral exponential distributions. And if we consider fitting the daily returns on the S&P 500, this is data for a four year period. This peaked distribution with exponentially decaying density is fitting the data really quite well. And more importantly, much better really than a Gaussian curve does. So here, we have the fit of the Gaussian bell curve in red, comparing it to the Laplace curve in blue.

And an interesting question is, how do we judge whether one model or another is a good model? And a useful tool is to construct a q-q plot, which is a quantile-quantile plot. So we consider points where the vertical value corresponds to the magnitude of our data, and the horizontal coordinate corresponds to the theoretical value for the smallest from a sample of Laplace, the second smallest from a given Laplace distribution.

And so we basically plot observed quantiles versus theoretical quantiles for the fitted model. So this fits quite close to a line, although it's not perfect. At least if it were perfect, it would even snake around the straight line where the sample quantiles equal the theoretical quantiles. If we do the same q-q plot for a Gaussian or normal model, this is what our q-q plot looks like. And so one can see that this second q-q plot is not fitting the data very well at all. So this would favor using the Laplace instead of the Gaussian as a model for independent daily returns.

Now, the independence of our period by period returns is one that may not be satisfied by the data. And Robert Engle proposed looking at time series of returns and considering the possibility that the variability of these returns given by the variability of the epsilon t terms, that these, in fact, have time varying volatility sigma t, where the sigma t squared evolves according to a moving average of past squared errors. So an autoregressive conditionally heteroskedastic model is what he proposed.

And in looking at this model, it's interesting to note that if one can re-express the ARCH model in terms of this squared error term and have this model in the second equation being consistent with the first, and this second model corresponds to there being an autoregressive model consistent with the squared returns. And so this Lagrange multiplier test can fit the autoregressive model to these epsilon hat squareds and test whether there is autoregressive structure or not.

And if we look at the autocorrelations of the S&P 500 returns squared, you can see that the first almost 10 autocorrelations are significantly different from 0 in magnitude. So there is strong time dependence in the log returns.

And so how does one fit this model? Well, the go to typically in statistical modeling is to consider a likelihood approach where we consider a density for our data specified by the model. So if the error is epsilon t are independent Gaussian distributions with different variances, then this likelihood function defines the joint density of our data given underlying parameters.

And what becomes tricky with this likelihood specification is that there, in fact, are constraints on our parameters. So with the ARCH model, we can't have a negative alpha, because with the Gaussian squared epsilon, that could be arbitrarily large negative, or epsilon could be-- sorry. If the alphas were negative, then the epsilon squared could be arbitrarily large positive and we'd get a negative variance. So we can't have that. Also we have this property of the sum of the autoregressive parameters is less than 1. And this corresponds to the process not being explosive in its movements.

And then let's see. An extension of the ARCH models was done by Bollerslev. And this basically adds to our ARCH model terms, like an order p ARCH model. It adds an order q set of terms on lags of the volatility. And if we consider the special case of the GARCH 1, 1 model, this is its form, which is very simple and parsimonious.

What's rather surprising is that the empirical specification of these models is such that the GARCH 1, 1 model is often chosen or identified as the best GARCH order case. So we end up having a very simple model with just a few parameters to characterize the volatility.

These notes talk about how the GARCH model induces an ARMA model. So this is some time series results that are gone through here. And there's a long run variance of the GARCH model that is given here. And with a GARCH 1, 1 model, there basically is a long run volatility sigma star, which has this particular formula.

So one can find the long run variance of a GARCH model with this formula. This formula also leads to updates of volatility being a multiple of the long run variance, plus a deviation from that long run variance. So we have a maximum likelihood estimation for these models.

Now, I wanted to share with you a case study using ARCH and GARCH models with foreign exchange returns. And so this note uses data collected from the Federal Reserve on foreign exchange values. Section 1.2 actually fits Brownian motion using different time scales. So not assuming time dependence. It actually is interesting to think about different time scales.

But if we go to section 1.3 and look at time dependence in squared returns, this is the dollar pound returns. This shows the autocorrelation and partial autocorrelation of the returns. So there's strong partial autocorrelations of the actual series. And if we look at the squared returns, the actual series has high autocorrelations over time. But the partial autocorrelations have very high spikes at low lags. So this suggests an autoregressive model in the squared returns.

And the note fits an ARCH 1 and ARCH 2. So one lag, or sorry, one autoregressive term, two autoregressive terms or 10, and then fits the GARCH 1, 1. And the results are highlighted here, just showing you the different coefficient estimates for these different models. They're all statistically significant in terms of the parameters.

What's of interest, though, is just to look at the time series plot of the fitted volatilities using ARCH of different orders. So here's the ARCH 1 model. Here's the ARCH 2 model. A property of these ARCH models is that it never goes below 0. There's a lower bound. Here's the ARCH 10 model, which looks like it's maybe tracking time varying volatility well. And then finally, we can compare that with the GARCH 1, 1 model.

So instead of having 10 autoregressive parameters in ARCH, the generalized ARCH model is capturing time varying volatility in an appealing way, whether that's really valid or not is of interest. But it certainly is an alternative to these arch models that is quite powerful. Well, here's just a graph of all three together.

This note actually does a final section of looking at what happens if our error distribution is not Gaussian, but is a t distribution. There are packages, rugarch, which allow us to specify parameters and fits using t models, t distribution models for the errors.

And what's very curious is in judging the histogram of standardized residuals, the t distribution seems to do a decent job. And so here's with Gaussian residuals, I guess. And then here's the Gaussian residuals and q-q plot associated with the standardized Gaussian assumption. Then finally, there's the q-q plot of standardized residuals with a t distribution assumed. So what's significant is just how the quantile-quantile plot suggests that this model is really fitting the data much, much better in terms of characterizing the variation.

And the choice of which order model to use for the degrees of freedom, there's this negative log likelihood function of models for different t distributions as we vary the degrees of freedom for the t. And the t distribution has a heavier tail than a Gaussian distribution. At least if the degrees of freedom is really high, it's virtually equivalent to a Gaussian. But this is suggesting that there's possibly a 9 degrees of freedom for that.

So that's a very interesting finding. And it actually arises with some mathematical analysis that t distributions are appropriate if we think of spherical symmetry in the number of dimensions equal to the degree of freedom. And so if you imagine a joint distribution that is spherically symmetric in really high dimensions, when you project that down onto coordinate axes, you end up getting a Gaussian distribution. But if their symmetry is just of, say, 9 dimensions, then you end up having a t distribution for what those projections are.

So this suggests that in terms of time periods over which the distribution of volatility is stable, could be close to 9 dimensions. So perhaps corresponding to 9 periods of data, I believe.

All right. Well, we'll finish there. There's a lot of interesting material for you to check out, if you like. And I hope you can enjoy reading some of the articles of people who've studied volatility and see what their work has contributed. And it's a really interesting area for final papers, modeling volatility, predicting volatility. This is really scratching the surface of the topic. So there's a lot of interesting problems you could study. All right. Thanks a lot.