[SQUEAKING]

[RUSTLING]

[CLICKING]

PETER KEMPTHORNE: All right, we'll get started. See, last time we introduced you to the definition of martingales. And what we're going to do the first part today is to see how useful and powerful martingale properties are and solving interesting questions and challenges.

So just as a reminder, if we have a stochastic process xn, which is just a sequence of random variables, generally the index n corresponds to time. Then we can consider a derived process capital M sub n, which is a function of the first n values of the stochastic process x.

And we say that this second process, this derived process, is a martingale if the expected value into the future from time n minus 1 is simply equal to the time n minus 1 value. So with the process evolving over time, we have an M sub n process. Basically, at time n minus 1 we have a value, our prediction is that the expected value of that in the future is Mn minus 1.

And so some examples of these. We went through last time a simple random walk. And with the simple random walk, we end up having realizations of this martingale that evolve from, say, a starting point 0. And what's critical with these realizations is that the future paths that have yet to be graphed will have average value equal to the last point on these realizations. So there's a notion of there being flatness in the martingale expectations.

With the example here of a simple random walk with equal probability of going up or down one step, one can graph the standard deviations up from the average 0 for the processes. And basically the realizations of the random walk will, by the central limit theorem, converge to a normal distribution at any given time point far into the future.

Now, this second example of a martingale where we have a random walk Sn but then we square it and subtract n times the sigma squared, the variance of each of the steps, this is also a martingale. And so looking at different realizations of this martingale, we can see that it's very, very different from that of a simple random walk. And importantly, there's basically of a lower limit of the process, it appears, for fixed n and an unbounded upper limit. But this process as well is a martingale.

And let's see. A different martingale can be defined, not in terms of random walks with steps that add, but with random shifts with factors x1 to xn. So if we have Mn being the product of x1 to xn, where the x's are independent random variables with mean value 1, then this particular process M sub n will be a martingale with respect to the original series.

So here's an example of our Bernoulli factors, which are with 50% probability 1.5 and with 50% probability 0.5. This would correspond to the change in wealth from betting your wealth in a lottery where you either make 50% or lose 50% with equal probability.

And here is the graph of a realization over 50 steps. Here's other realizations. And what's interesting to see is that with this particular martingale, there are opportunities to, basically, if you're playing this lottery game at each step, to be able to grow your wealth to 5 or 10 times its initial value. But with moderate or even high probability, you actually have lost almost all your wealth.

So yet another example, and this one is really rather curious to introduce, but it shows you how we can be flexible in generating or constructing martingales. If we have the yn sequence being iid random variables and the moment generating function e to the lambda yn or e to the lambda y, yn is equal to the moment generating function. Or that's e to the ty. In our notation, we're going to use a phi of lambda here or phi of t.

Then if we consider taking xn equaling e to the lambda yn divided by phi of lambda, then this xn will be a martingale. Or this xn, rather, will be factors that have expectation equaling 1. And if we take the product of these x's, we basically have M sub n equaling e to the lambda times the sum of the yn's divided by phi of lambda to the n-th power. This will also be a martingale.

And what we'll see towards the middle of this lecture is that this particular martingale can be used to solve probabilities of gambler's ruin, when you have gambles that have different probabilities of success versus failure, not necessarily equal fair bets.

All right. Well, with martingales, it turns out to be very useful to consider what we call the information sets on the subsequence of variables x1 through xn. So if we think of information progressing over time, the information up to time little n can be given by the outcomes of the x process. And we can consider expectations of random variables given this information set. And the kinds of random variables we might be interested in, z, can correspond to, say, trading strategies based on the outcomes of asset prices progressing over time.

And if we have, let's see, if this original information set is represented in terms of a martingale, then we can consider functions of the first n values given by y, and relating to this function of the first n values, which could be, say, the trading gains from applying a trading strategy with asset prices given by the x's. It will be very useful to use random variables that are called non-anticipating.

And so a non-anticipating random variable is a random variable An that is defined signed by the information set. It's non-random given the information set of the first n minus 1 x's. So if we wanted to represent the outcome of making trading decisions using the first n minus 1 values, this An could characterize how we might invest for the n-th trading period.

Now, with such a non anticipating random variable, we can consider a transformation of the original martingale, given by M's with no tildes, and consider a transformation of the Mn martingale by considering this decomposition of M tilde. So we have an initial value M0. Then we have the non anticipating variable A1 that depends only on information at time 0 times the increment of the Martingale from time 0 to time 1. And we add successive terms.

So these A's basically are random variables at the beginning of time. But after observing, say, j time periods, the Aj, let's see, will be-- or j minus 1 time periods, the Aj will be defined and non-random going from time j to j plus 1.

And if you were to look at this M tilde sub n definition, we basically have terms that are given by increments of the martingale process, where each of those increments has expectation equal to 0. And so if we take the expectation of those from time 0, then by conditional expectation properties, the A's are fixed given the information up to time or j minus 1, and then the variation of the martingale process from time j minus 1 to j has expectation 0.

Anyway, the martingale transform theorem says that this M tilde n process is a martingale also with respect to the information sets fn. And we often call this collection of information sets a filtration. And so when we think of different processes, different realizations of a process up to time n, then whatever the history is prior to n, that is the information set we have available. And as we observe successive values of the original Mn process, we're basically adding more information about the realizations. And that's how it progresses.

Now, with martingales, an important concept for us is a stopping time random variable, which we'll call tau. And so what the stopping time variable is, is a random variable that characterizes whether, let's see, or I guess at any time n, we know whether the stopping time has occurred or not. So the event of the stopping time being less than or equal to n steps, that has to be known in the information set up to time n.

So the notation here is a little bit tricky. We think of this set or event being the underlying probability distribution states where the tau variable is less than or equal to n. And at any time n, we know whether this set has occurred or not. So the stopping time is a random variable, and it could be infinite. So it takes values on the set S varying from 0 up to possibly infinity if the process never stops.

Now, a way of trying to help explain this is with a stopping time random variable, we have different omegas in the probability space, and the omegas can be thought of as indexing all possible paths. So with different realizations, we have a different omega. And so with these paths, our expectation or our event E, which is the set of paths for which the stopping time evaluated on that path is less than or equal to n, this event is known to either have occurred or not to occur. So we have the indicator of the path being in this set E is either 0 or 1.

And so we can work with this random variable, the indicator of whether the stopping time has occurred or not. And with this setup, we can think of the process that is governed by a stopping time tau. We can represent the process at the stopped time tau as the sum of the realizations xn times a simple indicator of whether the stopping time equals n or not. So we basically have our stopped process being a sum of weights times the xn's, where those weights are actually 0 or 1.

Now, in the mathematics of dealing with stopping times, the potential for infinite values of the stopping time makes things complicated. And what we can do is consider the minimum of n and the stopping time tau denoting that by n caret tau, so the minimum of n and tau. And so if we have a finite value n, this truncated stopping time is going to be a measurable random variable with respect to the information set script FM, information set FM for M greater than or equal to n.

So if we imagine basically this time increasing, and we think of the minimum of n and tau, then we will be considering processes that maybe stop for a given tau, for a given realization. Or if they don't really stop until later, we just consider the value at the n point.

Now, this theorem can be proven with this argument. And what you can see is that we define our non-anticipating variables to be the indicator of tau, the stopping time being greater than or equal to k.

And so substituting that in to the formula for the decomposition, we end up getting that this stopped martingale at time, the minimum of n or tau is actually equal to the sum of these non-anticipating variables times the increments of the martingale. So this is easily a martingale transform and satisfies that martingale transform theorem. So this stopped martingale at the minimum of n and tau is also a martingale.

What we then want to do is consider taking-- let's see. I'm not sure if we do it here shortly or not, but we'll basically let little n grow large and argue that in the limiting case, we basically have the stopped martingale without the minimum of n in tau.

Well, let's look at the simple random walk we had before where the steps are plus 1 or negative 1. And we consider Sn being the sum of the steps. This can be used to characterize the problem of playing against an opponent and betting $1 with each iteration of the game. And you either win or lose that dollar each iteration of the game.

And we can consider having a hit level for the sum Sn being equal to plus A or minus B. And so basically, this reflects whether we beat our opponent or not, where our opponent has a, well, let's see, if we have a bankroll B and our opponent has a bankroll A, do we win the entire bankroll of our opponent and hit A first?

So we have this problem, and we want to know what's the likelihood of basically player one losing. And the SN is a martingale. And tau, the time at which the sum Sn equals plus A or minus B is a stopping time. And so this is a martingale by the stopping time theorem.

And with such a martingale we have that the expected value of S at the stopped time is equal to the expected value initially. And that is equal to 0. Starting out at time 0, Sn is equal to 0. And so if we take the limit as n goes to infinity, this converges to the random variable S tau with probability 1. And this is proven partly by showing that the stopping time can't be infinite. But if we assume that these results are valid, then we have 0 is equal to the expected value of S tau.

And if we write out what S tau is, at the stopping time tau, it's either going to equal A times the S tau hitting A or minus B times the stopped value equaling minus B. And if we take expectations of this equation, we basically have 0 is equal to A times the probability that we hit the level plus A minus B times the probability that we hit minus B. And those probabilities are complements of each other. So the chance that we hit minus B at the stop time is 1 minus the probability that we hit A.

So using these relationships, we get that the probability that we hit A first is B over A plus B. Yes?

AUDIENCE: So on the previous slide, can I just check, I don't know if it might be an error or if I don't understand something. Is it meant to say solve for S tau plus A where it says problem in bold?

PETER KEMPTHORNE: Let's see. Can you ask the question again?

AUDIENCE: So where it says problem in bold [INAUDIBLE] and then problem solved, is it meant to say probability that S tau--

PETER KEMPTHORNE: Oh, yes. It's probably the S tau equals plus A. Yes, definitely. No, no, that's fine. That's fine. There's always typos. Fewer typos on slides than on the Blackboard, I would say. That's the benefit of having the slides, and easier to fix. All right. Thank you for that.

All right. So we get this really nice formula for the likelihood of hitting A before hitting minus B. And so by symmetry, if A equals B, then this is simply 1/2. But if the player B has a larger bankroll than the chance of taking all the money from A, basically goes to 1. So it makes sense that if you're going to gamble with fair odds, you have an advantage if you have a bigger bankroll than your opponent.

Now, this second example of martingales, which is the square of the random walk Sn minus n, this is a martingale as well. And if we look at this Martingale, it's bounded by the max of A squared B squared plus tau. So we have a bounded transformation of the original Sn. And so this martingale has expectation 0 for all n. And so if we look at the expectation of M tau, this transformation of the original martingale Sn, it has 0 expectation, which is equal to the expected value of S tau squared minus the expected value of tau.

And if we look at what is the expected value of S tau squared, so this term here, we have that that's equal to being stopped at plus A times A squared plus the probability of being stopped at minus B times B squared. And this expectation is equal to 0. So we can actually solve for the expected value of tau. It's simply equal to the expected value of S tau squared.

So plugging everything in, it gets a little tricky until we simplify. And it's simply the product of A times B. So the amount of time it takes to defeat the opponent has an expected time equal to A times B. So if you're gambling in such a situation, the size of your bankroll times the size of the other person's bankroll estimates the approximate time it will take to finish the game, with one dominating the other.

Anyway, what's characteristic of these examples is we have a martingale transform of the original series. It has a 0 expectation as it's defined. And that 0 expectation applies for all time points as well as the stopping time.

All right. Well, a really challenging problem is to consider the random walk with bias. So we consider steps plus 1 or minus 1 with probability p and 1 minus p respectively. And then we're interested in what is the probability of hitting plus A, say, in this case. Certainly, as p increases, the probability of hitting plus A first increases. But how does it do so?

Now, some of you have taken stochastic processes already. Or how many people have taken the course in stochastic processes? Maybe just a couple. Do you recall this problem being covered in your course?

And when I taught the stochastic processes course about five or six years ago, the way this was solved was by brute force understanding the dynamics of the process and doing computations with first step analyses. And the solution was really quite complex, but it got the right answer.

What's really neat about martingales is that we can actually solve this problem pretty elegantly. And so we're going to use this moment generating function transformation. So our xn's have a moment generating function, which is p, the probability of plus 1, times e to the lambda times plus 1 plus q, the probability of negative 1 times e to the lambda times negative 1. So this definition of the moment generating function is very straightforward.

If we take e to the lambda xn divided by phi, these are independent random variables that have expectation equal to 1. And we can define n then to be the product of the first n yi's. And this will be a martingale by our example 4.

And if we solve for the lambda value, that makes phi of lambda equal to 1. Then for that fixed value of lambda, this Mn is going to be a martingale. It's actually a martingale for any lambda. But in addition, that special case where phi of lambda equals 1. And so if we graph this moment generating function as a function of lambda, we can see that it's equal to 1 when e to the lambda is equal to q over p.

So Mn for this particular lambda for which e to the lambda equals q over p is simply q over p to the power Sn. And with that, we basically have the initial value of M at time 1, basically, or at time 0 is 1. And that should equal the stopped value of q over p to the power SN2 at the stopped value. So it's q over p to the plus A times the probability of stopping at A plus q over p to the minus B times the probability of stopping at minus B.

And again, this equation basically gives us two terms that are related to each other that we can solve for. So we get the probability that this biased random walk stops or hits plus A first is simply the ratio of q to p to the power of B minus 1 over the same ratio to the power A plus B. So this result is very direct and elegant. If you look at stochastic processes books that try to solve this biased random walk problem without martingale theory, it's a bit more involved and tricky.

All right, so here's just some R code that will graph these probabilities as functions of A and B. I think it's useful only just to know that you can do something like this. You can graph the probability surface as it varies with the hit levels A and B. And here are contours of the different probabilities, which can also be graphed easily. And that finishes this initial discussion of martingales.

Now, a second topic in stochastic processes that is extremely important is the concept of Markov processes. And so if we think of a general stochastic process xt, then we will say that the process is a Markov process if for any times of the process S that's between u and t, that xt will be independent of all xu's where u is less than s.

So when we think of a time scale, so we have u less than s less than t. If we imagine realizations of a stochastic process, and we are interested in looking at the value of xt, the value of xt given information on times s and before will be independent of all of the times before. And so the conditional distribution, so the distribution of xt, given all of the values of x prior to time u, is equal in distribution. It has the same random distribution as just the conditional distribution given xs.

So with a Markov process, the future depends on the past only through the last value of the past. It doesn't matter how one arrived at a given value at time s. For different realizations that have the same end point, it's only that point that indexes the distributions into the future. So this is a Markov process. It's a simple model where the past information can be encapsulated just in the last observed value, not in the realizations of how that value was achieved. So this is just some notation for trying to say what I just said.

Now, with Markov chains, these are interesting stochastic processes where we consider there being a state space, capital S, where we can index possible states that the process beginning, say, with 0 and the integers. And it could be a finite number or a countably infinite number. And then we have a time index set with n varying from 0 through the positive integers.

And so the Markov property with this stochastic process would have that the probability of the next state being j from time n, given the values of the process x0 through xn, this conditional probability is equal to the conditional probability of j state at time n plus 1, just given that xn is equal to i. So this is a nice property for processes that makes their analysis a bit less complex.

And what's important in this definition of the Markov process is that there's a one step transition probability from state i to state j, from time n to n plus 1, and these transition probabilities, together with the probabilities of the initial states, completely specifies the stochastic process.

Now, this issue of completely specifying a stochastic process is an important one, in that what we want with a stochastic process is to have a complete probability model of that process. And so if we have an initial state x0 and the probabilities over all states i in S, that characterizes the probability model for the state 0 value.

Then if we look at the conditional probability from time n equals 0 to n-- or from 0 to 1, we then have the conditional probabilities of moving from any state 0 to any state at time 1. And so those conditional probabilities, together with the marginal probability, specify the distribution of the process over state 0 and 1. And we can continue this computation for successive time points and completely specify our stochastic process.

Now, the general Markov process, as introduced there, is very complex, because we may have different transition probabilities depending upon how long the process has been operating. A simple case is called a stationary Markov process, where these transition probabilities, one step transition probabilities, do not depend on the time. So we have the same probabilities of moving from one state to another regardless of when that is occurring.

And so we can define a matrix P, which is the collection of all pairwise probabilities Pij moving from state i to state j. And we basically require that these transition probabilities are non-negative, and the sum over target states j from state i that those probabilities sum to 1 for a stationary transition probability matrix.

And to define the probability distribution of the entire process xn using this stationary transition probability matrix and initial probabilities, we can compute any probabilities by computing the collection of all such probabilities where we have n steps of the process, or n time points of the process, and states i0 through i little n there. So if we can calculate these probabilities, we will have specified the probability model.

Now, in calculating the joint probability of the first little n values states being i0 through in, we can decompose this probability into the marginal probability of the first n minus 1 times the conditional probability of the last given the others.

So if we think of A equaling x0 up to xn minus 1 equals i, n minus 1, and B is equal to xn equal to in, then we can look at the probability of A and B is equal to the probability of A times the probability of B given A.

And this is basically what's written up there. And this formula on the Blackboard is simply Bayes' formula relating the conditional probability of B given A to the probabilities of A and B. So probability of B given A.

Now, with this second equation, we can simplify that to the red probability, which is only depending upon the n minus 1 value. And so we have that this probability is equal to the marginal probability of the first n minus 1 times the transition probability from state in minus 1 to in. And that won't depend on n.

And so we can basically, by induction, basically replace each of the successive state transitions by the probability of moving from one state to another and the successive transitions of all those states. So this joint probability can be represented in this form, exploiting the Markov property to simplify the one step transitions, and using the stationery property to define these Pij's.

OK, well, with stationary transition probabilities, the multi-step transition probabilities turn out to have a really nice form. And so we can represent the transition probability matrix over n steps as the n-th power of the one step transition probabilities matrices. So this property of Markov chains with stationary transition probabilities is very nice and gives us this result.

Now, let's see. In terms of proving this, There's this analysis called first step analysis in stochastic processes. That's a really useful tool for proving different results. And this first step analysis basically says, OK, if we want to calculate the probability of going from state 0, sorry, state i to j in n steps, that's like starting at time 0 in state i and transitioning to state j. Then we can decompose this probability into the sum over all possible first steps going from i to k in state 1 and then from state k to j by time n.

And so this first equation here is a first step equation of the probabilities. We can basically decompose the process of going from state i to j in n steps as going from state i to each k state in step one, and then going from state k to n in n minus 1 steps. So this is that formula.

And so if we write this out, we end up getting the xn equaling-- or this in blue is equal to the term above in black because of the Markov nature of the Markov property of the process. And so we basically have this sum of Pik times probability of xn minus 1 equaling jx0 equals k. So we end up basically getting this matrix equation of the n step transition probabilities is equal to the one step matrix P times the n minus 1 step transition probability matrix in green.

OK. Well, these transition probabilities, together with the marginal distribution of the initial state, gives us the marginal distribution of xn at any time. So these formulas just indicate how we do these computations.

Now, where do Markov chains arise in finance? Well, one example to think about is how companies have credit ratings. And those credit ratings range from high credit, which is AAA, and low credit being I guess C. D being default is maybe a terminal state, at least in some sense.

But companies are graded or rated by Standard and Poor's and Fitch and other rating agencies, and depending upon how good their credit rating is, they can borrow money at lower interest rates. And the way corporations borrow money is to issue a bond, paying a certain interest rate and having that interest rate be attractive to lenders. And the way that such a loan contract is attractive to lenders is with a high interest rate to the lender, but a low risk of default by the borrower.

And so one can think about how these credit ratings vary from one period to the next. And here's a table of migration probabilities from one year to the next. And one can see that AAA initial ratings had only a 43% chance of staying AAA at the end of that one year. And there was a very low probability of downgrades to C, CCC, or B.

And with, say, very low credit rating companies, the chance of basically increasing their credit rating to B was 41%, staying at CCC is 32%, but almost a 20% chance of defaulting perhaps. So these kinds of metrics basically can be analyzed with the tools of Markov chains.

And let's see. Then another example that I put together is looking at the stock price dynamics of any stock, but I chose Apple, and using daily prices to consider states of the Apple stock price, which is it an up day or a down day? Is the price at day t higher than t minus 1, or is it a down day? I guess one of those should probably be a less than or equal to if we wanted to incorporate corporate everything.

But we can consider defining a two day state, which, let's see, the formatting here is a little off, but we could have two up days in a row, an up day followed by a down day, a down day followed by an up day, and a down day followed by a down day. And look at trying to model the state changes for a two day state with a Markov chain.

And in R, there's a nice r package that will do this for us. And this was distributed in the Canvas site. And so you should be able to replicate this example with Apple or choose any other stock you might be interested in.

And so importantly, this uses the library quantmod here. Sort of hard to see. Let's see. So there's a library quantmod, which allows us to collect price data from the internet. And then there's this package Markov chain, which allows us to specify Markov chain models.

And so with this, our code, it basically collects the data on Apple stock. And I think what I really want to emphasize for everyone is that these examples-- and so it's a bit tedious if you want to go through these examples yourself. However, the examples should work just with clicking a button, saying run, and you can replicate these.

And so for the Apple stock, we just needed to substitute in the Apple symbol and collect the data. And so here's the Apple stock price from 2014 or 2013 through a week ago.

And I considered just fitting a Markov chain model for these up and down states to the last 4 and 1/2 years. And so the code here goes through. And basically, with Apple prices, Apple closing prices here, we define differences in closing prices for two lags. And then talk about defining up, up days in this section here, up down days, down up, and so forth.

And then if we graph these states, say, from May of this year to the end of August, this shows how this 4 state Markov chain evolved over these dates. And we can just use the Markov chain function. Or let's see. We can plot a Markov chain function or a Markov chain transition states as this graph here.

So what this code does is it looks at transitions from one day to the next and calculates the probabilities of going from one state to another. So the sums across rows equals 1. And this graph, this plot shows basically what is happening with the Markov chain. We have different states given by the nodes. And then we can transition from between nodes by which arrows are feasible. And this Markov chain package will give us the probabilities, empirical probabilities, of moving from one state to another.

And so one can see that two up days in Apple are likely to be followed by two up days, again, so that we'd get a third up day with that one step transition. Two down days are not likely to be followed by a third down day, because it won't go to that down, down state with probability or more than 0.5. So there tends to perhaps be reversals from double down days and a continuation of up days, as suggested by these graphs.

Now, with this example, I thought, well, maybe there would be some interesting structure if we defined three day states. And so the same logic gives us 8 states of up or down on 3 successive days. And so 2 times 2 times 2, or 8, is the total number of possible states. And one can calculate the empirical probabilities of transitioning and then get this display of the Markov chain with these eight states.

And so what's rather curious just to focus on our three down days are not likely to be followed by a fourth. And three up days maybe are likely to be followed by a fourth up day. So you've got some insight perhaps into the overall dynamics. And what's great to me is that we can visualize the process and the graphical representation of nodes and edges between the nodes that have probabilities is quite useful.

So let's see. If you study Markov chains and then one can classify different states or different nodes as communicating states. So can you move from one state to another or not? And in this case, one can move from one state to another, possibly not in one time period, but in multiple ones. And so this happens to be a strongly connected set of states.

And let's see. This little note ends with printing out the columns of the transition probability matrix for up moves and down moves. And so if we wanted to move from one state to an up move in the next period, then looking at this table, one can see that the highest probability of the three day state being up in the third day, given the prior three day state, then it corresponds to three down days lead to an up day the next day with almost a 60% chance.

And one can also do the same for down moves and see under what state are we most likely to have a down move. And so I think 0.510 is the highest, which isn't much higher than 0.5. But basically, the past few days being a bit mixed, but with two up days preceded by a down day, give us a higher likelihood of a down day.

So let's see. With this note, let's see, there's a-- OK, so here's the reference for this Markov chain package. And what's quite convenient with different R packages on different application topics is that these articles often go through the mathematical background of the models. And so if you check out this discussion, this article on the Markov chain package, it actually will include the mathematical issues of how you might classify different states and do various computations. So that's quite useful to have as a resource with these.

Now, let's see. With the RStudio Cloud files that were distributed, if you use those programs, you can change the symbols from Apple to any other and use different time periods to see whether these properties may be strong or not with other stocks. And I think that might be an interesting thing to check out.

Our next topic, which is on regression analysis. And with regression, we want to just review the notation and setup for multiple linear regression. If you've taken any statistics class prior to this, which most of you have, you're very familiar with simple linear regression where you have one dependent variable and one predictor or explanatory variable. We want to consider the case of multiple linear regression, where we have a single dependent variable and multiple predictor or explanatory variables given by x's, and we want to describe the relationship between the dependent and independent variables.

Now, different regression problems can focus on prediction. This is actually one that I find most engaging, at least trying to, say, predict stock prices over fixed horizons is of real interest. But regression models can also be used to make inferences of causality. Regression models arise in medical experiments where one of the explanatory variables corresponds to a treatment effect. One can also be interested in trying to approximate the dependent variable by functions of the independent variables and understand their functional relationships. So all of these objectives can be achieved with regression.

And we define a general linear model with these equations here. We basically have our observed yi values being a sum of a fitted value plus error. And the fitted value y hat i is written here as a linear function of the independent variables xi with coefficients beta 1 through beta p. And there's actually a typo there. It shouldn't be beta sub ip, it should just be beta p there.

And with this setup, well, we could consider different explanatory variables that happen to be just different powers of one independent variable x. We could also consider Fourier series of x variables. We could also have time series regressions where the independent variables x may correspond to prior time values of the y series. And in all of these cases, what's really important is linearity of the fitted value y hat.

And what's linear is not that it's the sum of x's so much as it is the regression parameters beta are a linear function of those. So the linearity of y hat in the regression parameters is a key property of this model, and that linearity dependence on the parameters allows us to specify the parameters.

Now, let me just go to another example here. Let's see here. OK. Here's a time series of the high yield spread in the bond markets. And so this is looking at the yield on bonds with a credit rating of B AA minus the yield on a US Treasury. And so this difference in yield tends to be positive, well above 2. It means that these lower credit rated companies have to pay more interest to lenders when they borrow money. But that excess rate above the Treasury evolves over time and sometimes drops moderately or rises moderately.

Now, if we were to use just polynomials to try and fit these, well, as we extend to eighth order polynomials, everyone here probably knows that with a high enough order polynomial, you can match any smooth curve. No surprise. Do we expect this spread to actually follow some high order polynomial? Probably not.

Here's another example of using Fourier series with the same spread time series. And so as we add in more terms in the Fourier series, we're able to model higher frequency variations. And what this shows is just how, I guess, frequency analysis in financial series potentially can be useful.

What's really surprising, though, is if we do a autoregression, where here is the high yield spread, and then we basically have an autoregression saying that the spread tomorrow is the same as the spread today or very, very close. Then we get this blue curve fitting the black curve, which is over, underneath the blue curve. So we get almost a perfect fit with this autoregression.

So there's a multiple r squared, which we will introduce you to, of 0.995 for the autoregression. So the use of historical values of a time series can be very powerful in predicting future values.

All right. So in fitting a multiple regression model, whether it's with time series factors or not, what we do is go through these steps of proposing a model, step one, and make assumptions about the error term epsilon. And then second, we define a criterion for judging different estimators. And that criterion can be things like least squares or minimum mean squared prediction error or others. And then characterize the best estimator using that criterion.

Then we check whether the assumptions we made are, in fact, satisfied. And very often, we'll discover that certain assumptions we make are not satisfied and we need to revise those and create a new model. The model assumptions we apply include what are called Gauss-Markov assumptions, which is where we assume that the errors in the model have mean 0 and constant variance.

And so with this epsilon hat corresponding to the n cases in the series, we basically can assume these are mean 0 and have standard deviation sigma that's a constant and that these are uncorrelated.

We could additionally assume that these errors are normally distributed or not. We can generalize the Gauss-Markov to assume non-constant variances of the errors and possibly correlations of errors over time. We could also consider non-normal non-Gaussian distributions. And all of these cases actually have a role in different contexts. In particular, I guess generalized Gauss-Markov methods apply with time series regressions, and the non-normal non-Gaussian distributions arise as well, often with asset returns data.

Now, in terms of what kind of criterion we should use for specifying parameters, we have least squares, maximum likelihood, robust methods, Bayes' methods, a whole variety of methods. And what's really important to me is for people to have an understanding of what these different methods are, and when you would prefer one method over another.

And so what I think is a really important question to ask when doing empirical analyses with different models is, are your assumptions reasonable, if not satisfied by the data you're modeling? And with different criteria for judging different specifications of models, are you using an appropriate criterion or not? Are there variations in that criterion that would lead to higher performance levels? So that for me is a guiding question throughout these different checks.

So with that, we'll finish today. And next time, we'll finish these notes going through ordinary least squares estimation. And what's useful about this development of least squares estimation is that it does so with linear algebra. And what we'll see is that with linear algebra, we get very straightforward formulas for least squares estimates, as well as interesting analytical properties of those least squares estimates. So we'll cover that next time.