[SQUEAKING]

[RUSTLING]

[CLICKING]

PETER KEMPTHORNE: So today we're going to cover an introduction to time series analysis. And time series analysis is really a very powerful methodology in statistics, and especially powerful when it comes to modeling financial markets. So I hope you'll find this introduction engaging and motivating.

So let's talk about time series. We basically have notation indicating a stochastic process where X corresponds to the value of the process at time T, and the time periods T can be discrete. This is most commonly what we'll work with. But it can also correspond to a continuous time interval. And so we can imagine basically time periods, time indexed by positive numbers from zero.

And in terms of modeling such a time series, we want to be able to specify the probability model for any collection of time point values. So when we say, let's specify a probability model for a time series, then that corresponds to being able to specify the joint density of any set of points-- here we have m different time points. This generalizes to all cases of m.

And such a subset of the values of the process are a finite dimensional set of values, and so we'll have finite dimensional distributions for xt. Now importantly, we'll want to consider time series that are strictly stationary. And a strictly stationary time series is one for which the distribution of any set of time points is equal to the distribution at the same time points shifted by a constant.

So if we're looking at the index of time and we have xt, if we look at, say, four points here, T1, T2, T3, T4, then the distribution of the values at these time points will be the same as the distribution if we shift these time points all by the same amount tau. So T2 plus tau and so forth.

So we basically have that in this window and in this window, the distribution of outcomes is the same. So what does that mean? Well, that means that if there's a mean value in this period, then it's the same mean value in this other period, as well as all possible periods. So we basically have a probability model that is constrained in terms of the level being horizontal and the amount of spread about that level is constant.

Now with statistics, we like to estimate parameters. Parameters, like a sample mean, are very easy to estimate. We get a sample of values and estimate the population mean with the sample mean. And so we're maybe familiar with estimating constants and estimating means of populations is a trivial example or a simple example of that. With stationary time series, we want to consider perhaps a transform of our time series to a stationary scale.

And then there will be parameters of that distribution that we can estimate. And we can estimate them because the distribution is the same, basically across the whole range. So there's some hope that we'll be able to specify parameters consistently and with high precision. Now a less strict stationarity condition is called covariance stationarity.

And with covariance stationarity, we're looking just at the first and second order expectations of the time series values. So the mean value of Xt will be a constant mu for all t and the variance of Xt will also not depend on the time t. It will be stationary or constant over the time range. And the covariance of values that are tau units apart will have a covariance that simply is a function of how far apart the time points are.

So here, we're looking basically at a mean level mu and maybe plus or minus the square root of the variance of Xt being the lower and upper levels here. And then if we're interested in looking at how correlated different time values are, then if we consider, say, t and t plus tau, I guess, and we also consider another t prime and t prime plus tau over here, then when we have realizations of the process, they're going to be staying within a few standard deviations of the mean.

And the correlation between values at different time points will be the same if the time period between those time points are the same. So that's covariance stationarity. Now with such time series that are covariance stationary, then we can look at the autocorrelation function. And the correlation between Xt and Xt plus tau is the covariance between the two divided by the square root of the product of the variances.

So these are familiar formulas, generally, for correlations. And we call them autocorrelations because we're looking at the correlation of a time series with itself. And so we have rho of t is the autocorrelation. Now let's take a look at some financial time series. Let's begin with just looking at the S&P 500 index. So this is the value of the 500 major stocks in the US equity market.

This is from, I think, 2011 to 2020 plus. And you can see how this, over that period of time, trended up with some dramatic drops due to COVID. And with such a time series, it's clearly not stationary because it's trending over time. If we looked at different bands of years, they have different means, and so we really cannot consider this scaling of the index to be stationary or covariance stationary.

And what ends up happening with many, many financial time series is that we consider the time series Yt, say, to correspond to a random walk model on the log scale. So if we were to look at taking the log of the series and look at the increments of that, this is the log return of Yt.

Then the value of Yt is going to be an initial value Y0 times the exponential of the sum of these log returns. And we can relate these log returns to the percent returns on a fractional scale. This is log of 1 plus Rt, where Rt is equal to Yt minus Yt minus 1 over Yt minus 1. So what will commonly be the case is that this sequence of log returns is covariance stationary.

And so what I'd like to do next is to look at some time series plots of different financial time series, and see whether this transformation to log returns seems to make the series appear stationary. So here's the S&P 500 index. These graphs here actually update the graph from before to the end of September. And if we change initially the absolute price to the log of the price, you'll notice that an exponential growth tends to transform to a linear growth on the log scale.

That comment is representing the fact that the lower panel follows more of a straight line path than the upper panel, which has perhaps some exponential curvature to it. Now, if we look at the monthly log returns, we can see that those monthly increments on the log scale have basically a flat time series process. And so it's rather consistent with this movement across a level that, in this case, is close to zero with constant variance.

Now, with this data, we might consider looking at the distribution of these log returns and see whether they're consistent with a normal model, just as a benchmark example that we might use. And here's the fitted normal curve to the data. Would you say that this is a good fit to the data? Do I see any nods or shakes? I think you had a shake.

AUDIENCE: It looks like relatively, but it follows like the up and down, but [INAUDIBLE].

PETER KEMPTHORNE: Right. Yeah. And so the true height is basically the realized values that are close to the mean, and the normal fit is underestimating those values that are close. And complimenting that, the tails of the distribution maybe are underestimating the magnitude of the tails as well. And so this property is a property that is called lepto kurtosis, which means slender. Lepto is Greek for slender.

So let's take a look at what happens if we look at weekly returns OK, with weekly returns, we get, again, a series that appears to be fairly flat. The volatility, though, looks like it maybe isn't stable. So the variation of the time series at different time points may not be constant. It certainly was extreme during the COVID crisis, as evident here. And here's, again, the normal curve, which has the same properties noted before.

And then here's the daily log returns. So looking at higher frequency measurements, we, again, get this property of stability in the mean but maybe not a stability or stationarity in the variance. And again, the normal model doesn't look like it's appropriate. It captures a center and a spread, but seems to be far off near the mean and far away. OK, let's take a look at Amazon stock.

So we're just going to look at the same plots here. You'll notice that the absolute price has a rather notable exponential growth in it. When we take the log scale, we have this plot. And what's really important to see is whether the variability in the time series is constant or not. And before taking the logs, we have a really narrow variation of the price series at the beginning compared with the end.

But on the log scale, the variation is more comparable. So that's actually a good thing to have. We still have a trend in the series, and taking the differences in the log values makes this series stationary, or at least appear stationary. And here's the log returns on a monthly frequency. And here's the histogram of those log returns and the normal fit. It's perhaps doing a slightly better job than for the S&P 500. It's not clear.

Here is the weekly frequency time series. Again, the top panel does look like it's consistent with a stationary series in terms of the properties we've highlighted so far. And the Gaussian distribution seems to be fitting a bit better. If we look at the daily log returns, then we have this lepto kurtosis appearing in this histogram of daily returns.

And in looking at the time series of the daily returns, it looks like there are some significant spikes that are not really consistent with the rest of the series. So maybe that feature is something that we would want to incorporate into our models. Well here's, getting away from the equity markets, this is looking at the crude oil futures contract. And with the crude oil futures contract, this future, it trades, I think, at the Chicago Mercantile Exchange.

It actually has a dramatic feature that the value of the contract went negative in 2020, which was a shock to financial markets, especially brokers and traders in those contracts. You can imagine if you had a hundred contracts of crude oil, on the day you woke up and the price was negative, you basically owed your broker 100 times that negative value. And brokers actually sometimes were unable to tell you what your position was because their systems didn't allow for negative prices.

If we tried to take the log of the series and look at the differences, with these negative values, that's actually a huge problem. But nonetheless, let's see. I did graph these, which essentially will exclude those days when there were negative values for the contract. And looking at this, here's the monthly log returns. So this is excluding the negative price periods on a monthly basis, on a weekly basis, and on a daily basis.

And so what's really important here is just how looking at different financial time series with the same framework of logging the series if it's positive and looking at increments on the log scale, which correspond to percentage changes, the data generally is transformed or often is transformed to a stationary scale. Yet another example is looking at the yields of the 10 year US Treasury.

And so the top panel shows-- it's called price, but it's really the yield in percentage terms of 10 year government bonds. And so you can see it started above 3%, dropped down to less than 1% following the COVID crisis, and then was rising ever since. And if we look at the price or the yield changes on the log scale, we get this plot, which isn't too different in its appearance.

But then if we look at the changes in log, the increments in log, which are log returns-- in mathematical definition, log returns are not really the right terminology for the percentage changes in yield. But we'll basically be using the same terminology as before. One can see that we basically get a distribution of increments, steps in the random walk on the log scale that are unimodal and symmetric.

And as we consider higher frequency-- here's weekly and here's daily-- it's actually doing a better job, perhaps, using a normal model for this transformed series than for the equity market series. So with these cases, the same manipulations of the data, taking logs, looking at log returns, corresponds to transforming the series to a random walk process.

And with random walk processes, we can think of the sum from 0 to t of-- or the capital TC of X sub t and call this maybe S sub t. In modeling the random walk by this sum, there will be many different models we can consider for this. The simplest model would be one where these X's, the steps, are identically distributed and independent. So it's like a pure random walk.

And more complicated models will consider St or the Xt's to be dependent upon each other. So what we'd like to do, then, is-- let's see. Let's see here. If we think of this autocorrelation function, the correlation between X's that are tau units of time apart, we can actually calculate autocorrelations for each of the series we just looked at. And so here is the autocorrelation function of the monthly log returns on the S&P 500.

And one can see that there actually was a large negative autocorrelation for one month apart. And the other autocorrelations varied around this blue range. And what's important with autocorrelations, if we have a time series Xt, t equaling 1, 2, up to capital T, and then we consider the sample autocorrelation at lag K, then this is an estimate of the correlation between Xt and Xt plus K.

And we call this RK. And I'll maybe put a hat on it to indicate that this is estimated from the sample. It turns out that this is approximately distributed as a normal with a mean-- or I guess, OK, a normal with the rho K, the true autocorrelation, and a variance that's 1 over T minus K. So if we look at these bands, these bands correspond to the region of sample autocorrelations that are consistent with a zero.

True autocorrelation. If we want to test a null hypothesis that rho K is equal to zero, we can say reject if R hat K is greater than 1.96 times the square root of 1 over T minus K, applying a normal model for that return distribution. So we basically have, for the R hat K where we have a rho K here, we have a bell shaped curve with a standard deviation square root of 1 over T minus K. And if we consider rho K equal to zero, then we do the rejection if it's outside these bands.

Let's take a look at how these autocorrelations vary as we vary the frequency. Here's the weekly frequency for the computation of log returns. And we have this pattern where perhaps there are some significant negative autocorrelations at, say, five weeks apart. And what would you say is a property of a time series that has negative autocorrelations? Say, of order at lag 5? What does that really mean in terms of what you would expect? Yes?

AUDIENCE: So mean reversion. Mean reverting.

PETER KEMPTHORNE: Yes. Exactly. So the negative autocorrelations are consistent, perhaps, with mean reversion. So that means that if the price changes are too high, they'll revert back, or if they're too low, they'll revert back. Let's see. Also on these graphs are partial autocorrelations, which we'll discuss later.

But partial autocorrelations are correlations at a given lag after accommodating the dependence on lower order lags. We'll see how that works out in the case study later. But let's take a look at just these other series. So here's the daily series with the S&P 500. And so you can see, oh, there's actually some reasonable strong autocorrelations.

Here's Amazon stock. I'm not sure why it doesn't have the band on the lower side. But it basically is symmetrical with that blue band where the correlation is consistent with zero autocorrelation. Here's the weekly. Here's the daily. So interestingly, Amazon stock doesn't seem to have any significant autocorrelations, regardless of the frequency.

Here's the crude oil future. You can see that with the crude oil future, on a weekly basis, we have a negative autocorrelation at two weeks and then a positive autocorrelation at three weeks. So there's some interesting patterns that perhaps can be exploited. And here's daily. Again, one can see the rather significant autocorrelations at the low lag values.

And let's see with the Treasury yields. Looking at the autocorrelations and how the log yield changes, one can see that there is some strong time dependence at the weekly basis, as well as the daily basis. Now, let's see. This discussion of autocorrelations is a very important one in time series modeling.

And I just want to highlight how, if we have a time series and we fit a time series model. And so we compute, say, Y hat t, t equals 1 to T, then the residual series from this fit of the time series model is the set of errors which are Yt minus Y hat t. And we can just write Yt is equal to Y hat t plus epsilon or error t.

A good model. So the model is good if the et's basically have mean zero, constant variance, are uncorrelated with each other. And so these assumptions are basically assumptions of a white noise process. So if our model basically predicts up to this residual process, which is unpredictable, has mean zero, constant variance, is uncorrelated, there's no information left in the residual series.

And so we can use the autocorrelation of the residual series. We can use the autocorrelation function of the error series to test if the model is good, basically. And a good model will be a model that has captured all of the time dependent information in it, and the predictions from these or forecasts from this model are ones we have confidence in.

And so there's actually another assumption here, which is that-- let's see. If the epsilon t are normal or Gaussian, then this gives good predictions. And by predictions, I'm thinking of not point forecasts of a series, but a prediction interval and how wide should that prediction interval be depending upon the variability.

When we work with time series, we're actually are going to work often with autocorrelations of residuals to see whether there's any time dependence in the series. What's rather interesting to note is that one can fix a single correlation, lag, and use that to test whether the data are consistent with a zero autocorrelation of a fixed lag. We can also use sets of lags and test whether the whole set is equal to zero or not.

Let's see. Let me just write it over here. So if, say, rho 1 equals rho 2 equals rho k is equal to zero, then R hat J, which is distributed as normal zero, roughly 1 over t for J equaling 1 to K. R hat J squared times t is approximately the square of a normal 0, 1, which is a chi squared 1 random variable.

And if we look at the sum of TR hat J squared from 1 to K, this will be approximately a chi squared distribution with K degrees of freedom. We can reject the null hypothesis if this test statistic, I'll call it-- well, actually, I'll call it BP because it's actually called the Box-Pierce test. If BP is large.

We'll say larger than a critical value C star, where this is determined by the chi squared distribution. And so this is actually called the Box-Pierce test. And so we'll actually be very interested in autocorrelations of the original series but then autocorrelations of the residual series. And whenever we have a model where the residual series still has significant autocorrelations in it, then we're really not done yet. We really need to continue.

OK. So we have-- now, the next topic is a really extraordinary theorem. The Wold representation theorem. And with stationary processes that are covariance stationary, then it turns out that any covariance stationary process can be decomposed into a Vt process which is linearly deterministic. So that means that basically the current value Vt as a linear combination of past values.

And St is a moving average process where we have these eta t's which are white noise. And our St process is just a weighted sum of those white noise terms. And the white noise terms are also uncorrelated with the linearly deterministic process.

So let's see. When I first learned of this theorem, I had an issue with, well, what is a linearly deterministic process? So can anyone suggest what a linearly deterministic process is? Or could be, or an example of one? So we have Vt is linearly deterministic.

So if we have Vt-- so if we observe Vt over a sequence of time points, well, for what patterns of Vt that determine the Vt process into the future? Can you think of any? So if Vt is equal to mu plus, say, cosine of theta t, then this thing will just move up and down, and we can actually determine that cosine function from enough values of Vt. Yes?

AUDIENCE: So does that mean like any periodic process would be linearly determined?

PETER KEMPTHORNE: Yes. Yeah. And it's interesting to know that. I mean, those of you who may have studied harmonic analysis could come across the result that if we have, say, different cosine or sine functions and take sums of those, then that will also be a linearly predictable process. And so if you look at summing signs of different frequencies, which basically corresponds to Fourier series that are flat, then those are linearly predictable.

So that's the Vt process that we should be aware of. But what's really interesting is that the variation about that deterministic process, the St process, can be expressed as a moving average process. And this representation holds for any covariance stationary time series. Now for any covariance stationary time series-- let's see.

With the covariance stationary process St, we have basically that mu is equal to the expected value of St. Sigma squared, which is the variance of St. This is constant over time. This is constant over time. And the covariance of Xt, Xt plus tau, is equal to gamma of tau for all t. So any covariance stationary process can be represented as this moving average of white noise process.

And so, well, how would we use this Wold representation theorem? Well, if we have a time series Xt that we want to specify a covariance stationary time series model to it, well, the first thing we would do perhaps is de-trend the series Xt so that it has a constant mean level. So if we had a linear trend or a quadratic trend, we would subtract that out. So Xt is a series that looks like it has a flat path.

And then we could initialize a parameter p indicating the past number of observations that explain the linearly deterministic term. And we can then project our time series Xt onto the space spanned by the lags, the p lags, and then look at the errors about that projection. So here we can think of having the time series values Y1 to Yn, and we can project those onto the lags of the Y values.

So for the first p rows, we actually need to go back in time. So that means we have to start not from the beginning of our time series, but maybe p units ahead. And then we can apply ordinary least squares to define the projection matrix onto those lagged values use. And then we have this projected residual.

And what we would want is that this projected residual, if it's a stationary time series, we could use time series methods we'll learn about how to specify a moving average model to these series and get estimates of the size and also estimates of the white noise terms. And with this step, we then can evaluate whether our residuals are orthogonal to lagged values for lags greater than p back.

So these residuals should not have any dependence on lags of y that are further back than p. If those are not orthogonal, then we should increase p. And then we want to look at whether the eta t's in our moving average model, whether those are consistent with white noise or not.

And so if they are, we're good. But if not, then we need to consider changing the specification of the moving average model and perhaps adding other variables such as deterministic variables that could be useful in the projections. But what this demonstrates is how this theory can, in the abstract, be applied to real world data.

So as we increase p, the number of lags we use for the projection, we end up getting a limiting value as p gets large, that this is capturing the linearly predictable part of the series, the Vt's. But this assumption basically requires if we let p grow arbitrarily large, we probably want to make sure that p over n goes to zero, so that we're estimating p parameters with a number of points that is large relative to p.

So this is just a way of thinking about how we might approach our modeling effort, exploiting the Wold decomposition. But what's really important here is that any covariant stationary model has a representation as this moving average process of white noise. So that's a fairly powerful result.

It says that once we have transformed our data to a scale where covariant stationarity is appropriate, then we can restrict ourselves to the class of moving average processes. And what we'll see is that there are alternatives to moving average processes called autoregressions that are very closely related to moving average processes. Yes?

AUDIENCE: So I guess my question is, why do you need to specify covariance stationarity for the representation theorem?

PETER KEMPTHORNE: OK. Well, with the representation-- OK. OK. Let's see. Where is it here? No, it's fine. Here it is. Basically, if we consider this process St, which has some set of weights psi i times white noise eta t's, then if these conditions are satisfied for the psis-- so the sum, they're square summable, finite.

Then this process St is going to be covariant stationary. It'll have finite mean and variance that's constant, and the covariances will also be constant over increments of any lag. What's powerful is that this representation can be used for any covariant stationary model. So this is covariant stationary. But any covariant stationary model has this kind of representation.

Yeah. Sure. And of course, you know, covariant stationarity is a nice assumption. It may not be satisfied. So you need to have more complex models. But it's good to know when you have these assumptions satisfied that this is the representation. All right. So we next want to introduce you to this operator called a lag operator or backshift operator.

And because I call it a lag operator, I label it L. And so L operates on a time series Xt and just shifts the time index back one period. So we're thinking of discrete time series with integer values of t and the lag operating on Xt shifts the time units back one period. And we can think of L to the 0 power doing the identity. L1, 1 power of L, shifts back once. If we apply L twice or L squared, this shifts back two periods.

And of course, if we apply it n times then we shift back n periods. And so with these lag operators, we can think of inverses of those operators as well and consider shifting the time series forward if that were useful. And so when we think of the Wold representation, we can express it as the sum of the moving average process and the linearly deterministic process Vt.

And we can represent the moving average process as a polynomial of the lag operators, possibly of infinite order. That is, multiplying the eta t's. Now, a useful application of this moving average representation is that if we think about how was the value of the process at time t affected by the white noise process eta j lags ago, then the partial derivative of Xt with respect to that j lag back white noise term is given by psi j.

So when psi j is large in magnitude, that means that the current value of X is dependent on the j-th lag of the white noise. And with a representation of the Xt series by this psi function, if one graphs the psi function over lags. So let me do it over here. So if we look at psi j as a function of j and have-- we might have weights psi j that vary positive and negative.

And so if this were the case, it would indicate in this graph that the impact of the white noise three periods ago has a significant impact on X now. So it can capture delayed responses of the time series Xt to the white noise terms. And if we look at the long range impact of eta j-- or sorry, the long range impact of the-- or the cumulative response of Xt, basically, corresponds to the sum of the size.

And so there's basically a long run cumulative response of Xt to a given eta t. So this is the long run cumulative response to the white noise terms. Now, what's interesting about the psi of L process. So if we have psi of L is equal to the sum from 0 to infinity of psi j, L to the j, it's possible that this operator is invertible.

And if it's invertible, there exists an inverse operator with coefficients psi i star, and the product of these two polynomials is simply equal to the identity. And if that's the case that an inverse of the psi of L operator exists, then we can write our model for the stationary process as being equivalent to applying the inverse operator to both sides.

And so we get an inverse operator, which is a polynomial lags operating on Xt is equal to the eta t. So this corresponds to actually an autoregression model that has white noise terms. And so when this exists, this is the format of that. And what we'd like to do is show you how this works with some different examples. Right now it's perhaps a bit abstract, but when we look at the special cases of an autoregressive order one model, many of these terms will become clear.

Now, we can introduce with this notation of lags and polynomial lag operators the ARMA p,q, models, where ARMA stands for autoregressive and M-A stands for moving average models, and the p corresponds to the order of the autoregressive, and the q corresponds to the order of the moving average. We can think of our time series Xt being a sum of terms, autoregressive terms on the top row, and then the second row corresponds to moving average terms of the white noise.

So notice in the top row, we have Xt. And actually, if you look at that equation, if we take Xt minus mu, this turns out to be equal to phi 1 Xt minus 1 minus mu plus phi 2, and so forth. So these are lag terms of the mean adjusted X's. And then we have the next row is the sum of the white noise terms at time t, t minus 1 up to t minus q.

So ARMA p,q models are a parsimonious case of time series models where we just have p autoregressive parameters, phi 1 to phi p, and q moving average parameters, theta 1 to theta q. If we group all the terms with X's on the left-hand side and all the terms with the etas on the right-hand side, we then get this phi of L operator times Xt mu view is equal to theta of L eta t.

And the Wold decomposition of this ARMA model is to multiply this ARMA equation on both sides by the inverse of phi of L. And so mu plus psi of L eta t is the Wold decomposition representation of the process. So let's look at autoregressive models. So we'll assume that there's no moving average term. So we have the phi of L operator operating on Xt minus mu is equal to the white noise.

And so that model basically says that Xt is a linear combination of the p lags of Xt plus white noise. So this is a regression model for our time series Xt using predictor variables the lags of this time series. And if we take expectations of both sides, we basically get mu. The constant expectation of the covariance stationary process is equal to C plus the sum of phi j times mu, the expectation of each of the lags, plus 0.

And so C turns out to equal mu times phi of 1. So we have that representation of the constant term. Now, what's important with time series models is whether the time series model is stationary or not. And it turns out that for an autoregressive model with parameters phi 1 to phi p, that if we look at what's called this characteristic equation, 1 minus phi 1z minus phi 2z squared up to phi pz to the p power.

If we consider the roots of that polynomial, then, well, if we have roots lambda 1 to lambda p, then phi of L has to be the product of these terms. Each of these terms, each of these factors are equal to zero when z equals lambda 1. So if z equals lambda 1, then the first term is zero. If z equals the p-th root, lambda p, then this last factor is zero. So this is a representation of this p-th order polynomial in terms of its roots.

And the condition for stationarity is that all of these roots have to lie outside the unit circle in the complex plane. And so if we look at one of these roots, fix just lambda as a root, then if we take the inverse of this factor, it has this geometric metric sum expression of 1 plus 1 over lambda L plus 1 over lambda squared, L squared, and so forth.

And so it's basically an infinite order polynomial in lags. And the coefficients are 1 over lambda to the I th power. So if these lambdas are bigger than 1 inch magnitude, then the inverse of those, the reciprocal of those, is smaller than 1 in magnitude. So these terms will die down with a limiting value of zero. And that's necessary for this representation to be appropriate.

So we get a covariance stationary sum of lagged errors, lagged white noise terms with this representation. So the phi inverse of L is actually the product of p of these reciprocals. And so the condition that each of those reciprocals or each of those terms is defined by lambda j's that are bigger than 1 in magnitude makes their geometric series expression converge.

If the lambdas were inside the unit circle, the magnitude, then when we multiply them together, square them, they get bigger and bigger. OK. So let's take a look at what this works out to be, just for an autoregressive process with p equal to 1. With p equal to 1, our characteristic equation is simply 1 minus phi z equals 0. And so the root of that is clearly 1 over phi.

And so the p equals 1 autoregressive model is going to be covariance stationary if and only if the magnitude of phi is smaller than one, or equivalently, its inverse magnitude is greater than 1. And so we end up having these first and second moments and these covariances which are very easy to compute. And this will probably be a homework exercise problem. But with the first order autoregression model, we have a constant mean, a constant variance, which is sigma squared over 1 minus phi times sigma squared.

And the covariances basically are the variance times phi to the j-th power. So with an autocorrelation, the phi is the first order autocorrelation, and phi to the j-th is the j-th order autocorrelation. OK. Now, if phi is between minus 1 and 1-- for phi between 0 and 1, the process exhibits exponential mean reversion. And when phi is negative but no larger than negative 1 in magnitude, then it exhibits some oscillating exponential mean reversion.

But when phi equals 1, then the Wold decomposition does not exist and the process is a simple random walk which is non-stationary. And when phi is greater than 1, then the process is explosive. So in financial markets, these first order autoregressive models actually are used quite frequently. With interest rates where there's maybe a constant long term interest rate level and there's variation about that, this Ornstein-Uhlenbeck process corresponds to the first order autoregressive process.

With these other series, these are financial or time series in financial markets where the level of the series is expected to be stable over time. So perhaps interest rate spreads would be modeled as being stable or stationary. Real exchange rates, perhaps, or valuation ratios. So these models arise in these application areas. Now to fit an autoregressive process, we could simply just fit a linear regression model with p lags of the series as explanatory variables.

But what's rather interesting to know is this approach called solving Yule-Walker equations. If we take our model equation, that first equation line, we have the mean deviation of Xt is equal to a linear combination of phi j times the j-th lag deviation from mu. There's a typo there, that second phi 2 should have an Xt minus 2.

Then if we multiply both sides of this model equation by Xt minus j minus mu, then we basically, for each term, we're adding Xt minus j minus mu as a factor, and then take expectations of that. We have our j-th lag autocovariance, and it's equal to the phi 1 times the j minus 1 lag autocovariance plus phi 2 times the j minus 2 lag, and so forth.

And if the j is equal to zero, then we also have a sigma squared term where we're looking at basically that. Well, that term comes in. So we basically have equations, a system of equations if we vary j. And so if we consider j equal to 1 to p, we get a system of p linear equations in the phi j's.

And this is our Yule-Walker equations. Gamma 1 is equal to the first row of the matrix times the vector of phis, and so forth. And with the properties of the autocovariances, they are symmetrical. So we can actually use these to make the computations of the matrices.

And when we use this Yule-Walker approach to estimate the parameters, we're actually using a method of moments principle for estimating the autoregressive model parameters. Method of moments is a principle in statistical estimation, where we basically equate sample moments and the population moments and solve for what population parameter values match the sample moments.

Now, we can do a similar analysis of moving average models. So here, we have a q-th order moving average model represented by a polynomial theta of L of order q in terms of powers of L. And the properties of this process can be studied. And one thing we might want to do is invert this moving average model into an autoregressive model by basically multiplying the model equation on both sides by the inverse of the theta of L operator.

And so if we do that, then we basically have an autoregressive representation of the moving average order q model. And if you calculate simple second order properties of this moving average process, well, the mean is simply mu. The right-hand side of the equation is just weighted sums of etas which have mean zero. And so the expectation of the left-hand side is simply equal to mu.

And if we look at the variance of Xt, it basically is the variance of the weighted sum of the etas. And it turns out to have this simple form with covariances of the series. With moving average processes, if we look at time increments j that are longer than the q order, then there's no overlap of white noise terms in Xt and Xt plus j, and so their covariance is going to be zero. When there is overlap, then we get this formula.

OK. So next, we want to deal with how we can accommodate non-stationarity in time series. So far we've just assumed examples and models, like autoregressive and moving average models, where this series is stationary. There will be cases when our time series is not stationary and differencing the series will generally render a series to a stationary transformation.

And so if we consider first order differencing, we have-- let's see here. If we do delta Yt equaling 1 minus L Yt equals Yt minus L Yt, which is Yt minus Yt minus 1. OK. This is first order differences. If we apply the first difference of this, we end up getting Yt minus Yt minus 1 minus Yt minus 1 minus Yt minus 2. And this turns out to be equal to Yt minus 2, Yt minus 1 plus Yt minus 2. And there could be k-th order differencing, but in practice, we'll only generally use first or second order differencing.

And what's really neat about differencing is that first order differencing will eliminate a linear trend in the series. Basically, if we have a linear trend, then the difference will just be a constant difference. And the second order difference is appropriate if there is a quadratic trend in the series and second order differencing will eliminate that quadratic trend.

So we'll finish there for today. But this issue of modeling differences of the series or second differences of the series as stationary, corresponds to having time series models that are looking at the dynamics of the slope of the time series, which is like the first derivative or the second derivative, and assuming that those are perhaps constant level and constant variability. So those mathematical features end up being very intuitively applied actually with different cases. OK. We'll finish there for today.