[SQUEAKING]

[RUSTLING]

[CLICKING]

PROFESSOR: All right, we'll get started. We're going to continue our discussion of principal components analysis. And what I want to emphasize with principal components is that we're dealing with a random vector X that's in m dimensional space. And it has an expected value of alpha, say. And it has a covariance matrix given by sigma, which is m by m.

This kind of setup can be used for modeling returns on stocks or assets that may have a mean return over some holding period and a variance-covariance matrix. And so in working with financial data, principal components analysis can be quite useful. Essentially, what principal components does is it transforms our X vector to a p vector of principal component variables.

And the definitions of these is we subtract from X the mean vector. And then we rotate the coordinate axes of this m vector by this gamma transpose matrix. And so this gamma transpose can equal gamma 1 transpose, gamma 2 transpose, down to gamma M transpose.

And so the first component-- p1 down to pm. The first component corresponds to using this gamma vector, which happens to be an eigenvector of the covariance matrix corresponding to where with associated eigenvalue lambda i. And in this representation, we basically have sigma gamma i is equal to lambda i gamma i.

And we can order our eigenvalues from largest to smallest. And the first principal component variable corresponds to lining up-- or looking at a coordinate system where we're looking along the dimension of greatest variability in the data. So this covariance structure it's actually useful to think of a special example, which is the multivariate normal distribution in m dimensions with mean vector alpha and covariance matrix sigma.

And so if we think of X1 and x2-- we'll just consider a two-dimensional case as an example. There could be an alpha 1, alpha 2 here for the mean of X1, first component mean of X2.

And with a bivariate normal distribution, we basically will have contours of the density, the joint density, of X1 and X2 that are given by ellipses centered at the mean vector. And the degree to which the components are highly correlated-- if it's a positive correlation, then we have ellipses that are tilted up. If it's a negative correlation, it would be tilted down.

Now, what principal components analysis does is it basically-- this step here corresponds to shifting the coordinate axes to, say, x1 star and x2 star. So we recenter our space at the point with the origin being at the mean point of the outcomes.

And then we consider rotating these axes in such a way that the first principal component axis corresponding to the largest eigenvalue, it will orient along the principal axis of the ellipse. And the second principal component axis will be orthogonal to the first. And it will-- well, in two dimensions it will have variability along that dimension that's smaller than P1.

When we extend from a two-dimensional case to an m-dimensional case-- so suppose we have 500 stocks or 1,000, then if we apply a multinomial model to that case, then our principal components will be identifying coordinates-- basically coordinate system variables that line up with different orthogonal dimensions of variability.

And so what ends up being very useful is that, at least when we look at a market of stocks or asset returns, some of those may be highly correlated with each other. Others may be, in a sense, orthogonal or uncorrelated with each other. And with principal components analysis, we can actually identify different directions in the space, which basically have decreasing levels of variability, but that are also orthogonal to each other or uncorrelated with each other.

So when we want to specify factors, underlying factors, say, in joint distributions of returns and markets, principal components can be very helpful in trying to identify those. And later on, we'll talk about factor modeling and see how close or how effective principal components can be. But there are also general factor models that we'll introduce later.

So with these transformations, as noted here, these principal component variables have basically a mean that's zero and a covariance matrix that's diagonal. So these principal component variables are uncorrelated with each other. And if we order the eigenvalues from largest to smallest, the first principal component variable will be the axis with the greatest variability.

Now, what's interesting to highlight is that in the transformation from the original X's to the principal component variables, the only thing we've done is shift the origin and then rotated the coordinate axes in an orthogonal way. So we're maintaining right angles between coordinate axes. And that's all we've done.

So if we have outcomes of X-- so if we have X is a random variable, but then we may have realizations-- realizations, say, X-- I don't know-- a, Xb. I don't want to have integers referring to these because I don't want to confuse the components. But we could have different realizations for different days or different analysis periods, and we would get different points of realizations of the outcomes of this vector.

And so principal components allows us just to view the data in a way that hasn't transformed the data in an unusual way. Basically, the configuration of points is the same before and after the construction of these PC variables.

So let's see. OK, there's this version of principal components where we consider splitting the first k principal components into a gamma 1 matrix and the remaining ones into a gamma 2 portion. And so one can then think of there being a linear model for our outcome x variable, which might be simplified to just a mean vector plus a matrix times the first k principal component variables plus error.

If this, in fact, were an effective model, it might be the case that the last m minus k principal component variables essentially have no important information in them. And so this could be a very simple model for our x variable.

And importantly, if this were an appropriate model for an m vector x, then the covariance matrix of the-- the covariance matrix of X is just given by-- or let's see. Or the-- yeah, the covariance matrix of X, in terms of the diagonal entries which represent the correlations, are all due to common loadings on these different principal component variables in p1.

So the interesting question to think about is, how can we model random vectors with some covariance matrix? And are there simple structures to the covariance matrix that explain all of the intercorrelations?

Now, when we perform empirical principal components analysis, then what we do is we-- say, we have an X matrix, which has columns X1 through X sub capital T that are all m vectors. Then we compute the row means-- we subtract those row means from the matrix row by row.

And then with this X star matrix that's the demeaned matrix, that matrix times its transpose divided by the number of columns in X capital T, that is an estimate of the covariance matrix, essentially the sample covariance matrix. And we then can perform an eigenvalue eigenvector decomposition of this covariance matrix, and then use these results to define the empirical principal component coordinates of our data. And it corresponds to just taking our estimate of the gamma matrix transpose the demeaned matrix. So this is very simple to implement.

And let's see. The singular value decomposition is also a way of obtaining these principal component transformations. If we look at VDU transpose being the singular value decomposition of our X star demeaned matrix, then we have our lambda estimate just being the square of the D diagonal matrix divided by T. The estimate of the gamma matrix is simply the V. And our principal component variables-- or our data transformed to principal component coordinates is simply given by DU prime.

So principal components analysis is very easy to implement. And the lecture that was originally scheduled for today, Stefan Andreev, he's going to go through a number of examples of applying principal components in equity markets and in bond markets.

What's really useful to anticipate is that-- well, one of the examples that he's used in the past is where the X vector corresponds to index returns in different European markets on a daily basis, and attempting to understand the joint variability across all of the markets, but also to understand where there are different geographical clusters of countries like Northern European versus Southern European. Those tend to be separately highly correlated with each other.

Another example of principal components analysis is looking at interest rates in the fixed income market. It turns out that for different maturities of bonds ranging from, say, one month, three month, a year, to five years, 10 years, 20 years, 30 years, the interest rates can be measured in terms of their yields. And the yield curve will vary day by day. And it turns out that the principal components analysis of yields has a very strong structure to it. So I'll let him introduce those when he comes in a couple of weeks.

Well, with principal component variables, it's important to understand that there are alternative ways of deriving or defining principal component variables. One is to define the first principal component variable to be the linear combination, defined by weights w1 through wm, that maximize the variance of that weighted sum subject to the sum of squares of the weights equaling 1.

So we're looking at a unit vector w and trying to find the vector w that maximizes the variance of the linear combination in that direction. So when we get to the transformation of the data-- or the random variable by demeaning it, then if we look at w vectors that are unit vectors in this space, then we're looking at which w vector is pointing in the direction of maximum variability.

Then the second principal component variable is defined as the next w vector, now v, which maximizes the variance subject to having length one but being orthogonal to the first. So it's a rather nice way of characterizing principal component variables.

Now, with principal components analysis, one of the really neat things is that-- let's see. We often focus on the first few principal component variables because they explain most of the variability. But the last principal component variable defines the linear combination of the components that has the smallest variance. And sometimes that smallest variance can be zero, in which case, we're identifying a linear dependence in the returns.

So let's see. My first example of this case was before the euro existed. So I'm really dating myself. But the German deutschmark and the French franc were very, very highly correlated before the euro came out. And in looking at currency returns, the principal components analysis actually identified that empirically in a nice way.

Let's see. OK, mathematically with principal components analysis, we can talk about a decomposition of the total variance. So if we have this covariance matrix sigma, we could talk about the total variance is equal to the sum 1 to m of sigma jj. So the j-th diagonal entry of the covariance matrix is the variance of the j-th component. And the sum of that is the trace of sigma.

We can decompose that total variance as the sum of the eigenvalues. And so principal components analysis provides a decomposition of variability. And we can see what proportion of variability is explained by the first or the first few principal component variables.

These lecture notes finish with just reminding you of some important statistic-- important probability distributions in statistics. The chi-squared distribution arises again and again. And the chi-squared distribution with one degree of freedom is equal in distribution to the square of a standard normal distribution. So this is a useful thing to remember in different cases.

And let's see. If we-- well, here's its density function. It turns out to be a special case of the gamma distribution. And here's this moment-generating function. It turns out that if we take independent normal 0, 1 random variables, square them and sum them, then the sum of n independent ones is a chi-squared with N degrees of freedom. And we can prove these results using moment-generating functions and get densities for those.

Importantly, t distributions often are-- let's see-- arise, especially in the study of regression models and the significance of regression parameter estimates. But basically, with a t distribution, if we take a standard normal and then divide it by the square root of a chi-squared divided by its degrees of freedom, where the chi-squared variable U is independent, then this will give us a t distribution with r degrees of freedom.

And so, let's see. With t distributions, t distributions are examples of distributions that are heavier tailed than a normal distribution. And so if you think of taking a normal z variable and dividing by this chi-squared distribution divided by this degree of freedom and square root, we're basically dividing a random normal by something that's random, but around 1. And the randomness of that factor leads to greater variability than the normal. So it's obvious that that could be the case. And it's indeed true.

And let's see. The F distribution turns out to be the ratio of two chi-squared distributions, each divided by their respective degrees of freedom. And in classical statistics, we often are interested in measuring variability of data and variability due to different driving factors. And so F distributions arise commonly when we have estimates of variance given by a chi-squared divided by a degree of freedom. And if we have two different estimates of variability, then if those distributions have common variances, then the F distribution applies.

Some of you may have heard of analysis of variance in statistics. Anyway, the analysis of variance methods work with studying F statistic distributions. Yeah?

AUDIENCE: Are noncentral distributions used a lot where there's a noncentrality parameter? Or are they mainly just like the centered ones?

PROFESSOR: Well, the noncentrality-- or when you have null hypotheses in analysis of variance that are true, then the F statistics will be F's with zero noncentrality. And if the null hypothesis is false, then you'll have a F distribution with a noncentrality parameter.

So the noncentrality parameter comes into specifying the distribution of the test statistic when the null hypothesis is false.

AUDIENCE: So it's not used sometimes then like when [INAUDIBLE] is false.

PROFESSOR: Yes. Actually, the next two lectures after today will be on regression analysis and testing and regression. And so we'll see those come into play there. Yep.

All right. What else? Oh, OK. Just the weak law of large numbers and the central limit theorem are the remaining topics here. If we have sample data realizations of, say, N-- or N realizations of a random variable-- and here, we say IID for Independent and Identically Distributed. If those random variables have mean mu and variance sigma squared, then the sample mean, X bar n, will converge in probability to that mean.

So the larger the sample size in the mean, the closer the sample mean will be to the true mu parameter. But what we-- what's important from probability theory is how do we characterize convergence of random variables? And so the convergence in probability to a constant means that the probability that the deviation, the absolute deviation from mu exceeding epsilon, that that goes to zero.

And this theorem-- or this result can be proven with Chebyshev's inequality. And it's sort of a neat example where if we want to look at the variance of X bar n, which is equal to, what, the expected value of X bar minus mu squared. Then if we were to draw X bar n, and here's mu, and then have the graph of X bar n minus mu squared, if we were to transform our scale to X bar n minus mu, then this mu value goes to zero.

And if we were to look at the epsilon value for X bar n minus mu, we can consider a function other than this parabola, which is zero, up to this epsilon away from mu point, and then is simply equal to epsilon squared beyond, and do the same at minus epsilon. So zero up to there.

And the variance, which is the probability weighted average of the parabola over the distribution, has to be greater than epsilon squared times the probability of exceeding-- of being further away from you than epsilon.

So the fact that this function is greater than or equal to epsilon squared times the probability that X bar n minus mu is greater than epsilon. This gives us the result. And we basically have that-- we're basically giving a positive bound for this. Or we have an upper bound that decreases with n for this probability. So that gives us our nice result.

Let's see. If you really like probability theory, one can actually extend this result to the case where we don't have a variance existing. We just have a mean existing. And that's a rather strong result that's neat. At least it's neat in terms of how the mathematics can handle more complex cases.

Now, finally, we have the central limit theorem, which tells us what the asymptotic distribution is of a sum of independent and identically distributed random variables is. And so if we have variables X1 to Xn that are mean 0 variance sigma squared, it turns out that the sum of the excise divided by the square root of the sample size converges to a normal distribution with mean zero and variance sigma squared.

So we have, what, Zn equaling 1 over root n sum 1 to n of the Xi's converges to a normal with mean zero and variance sigma squared. And if we were to think of X bar, which is equal to 1 over root n times Zn, this basically is converging, in some sense, to a normal with mean zero and variance sigma squared over n.

Now, I put quotes around this arrow here because it's not really legitimate to talk about a limiting distribution that depends on n. But this is the way we approximate the distributions of sample means. Basically, they converge to the mean of the distribution, in this case, zero. And the variance is sigma squared over n.

Now, what's neat in this proof is just to highlight the use of moment-generating functions. And so if we-- say, let's calculate the moment-generating function of Zn, we're going to show that the moment-generating function of Zn converges to this moment-generating function, which in fact is the normal with mean zero and variance sigma squared.

So the theorem that says moment-generating functions are of unique, and if the limit of moment-generating functions exists and equals some function, we can identify the distribution by that limiting moment-generating function. Well, if we calculate just the moment-generating function of Zn, if we just plug-in Zn within the expectation, then noting that the X's are independent of each other, we basically are taking the expectation of a product of independent terms. And so that expectation of a product is the product of the expectations.

So we have the product of the expected value of Xi times t over root n. And this expectation here is actually the moment-generating function of X evaluated at t over root n. So it's the n-th power of that when we take the product of all n of those.

Now, this is where it gets interesting. If we look at our moment-generating function of t over root n, well, we don't really know much about the distribution of the X's. We know it has a mean. We know it has a variance. The mean is zero. The variance is sigma squared.

If we expand the moment-generating function into a Taylor series, then we get 1 plus the expected value of X times t over root n plus the expected value of X squared divided by 2 times t over root n squared plus, an order term that is given by the third power of the argument.

And so as n gets large, these higher order terms are of smaller order. And so we finally have that this in the limit is 1 plus this multiple over n plus a small o of t squared over n. And if we take this expression to the n-th power, then it converges to the exponent of the argument sigma squared t squared over 2.

So let's see if we have 1 plus r-- let's -- 1 plus r over n to the n-th power, this converges to e to the r. And Vasiliy used this argument in the first lecture to talk about continuous compounding and what the limiting value is of continuous compounding at some interest rate. Anyway, that same exponential result is true here. So this actually proves the central limit theorem using the identity theorem for MGFs. Yeah?

AUDIENCE: Is CLT actually used in practice? Because obviously, like in finance nothing is going to be IID, or at least they're not going to be independent. So is it actually still useful in a practical setting?

PROFESSOR: It can be very useful in a practical setting. And right now, it's specified in terms of identically distributed X's. It actually generalizes to X's that are independent but with different distributions as well.

So also what's I guess really important is that if a variable we're working with corresponds to the sum of independent or uncorrelated terms, the sum can approach a normal distribution well. And if one looks at, I guess, linear projections of multidimensional vectors, X's, if you had a random direction in the n dimensional space, the projection onto that random direction will correspond to sums of close to identically distributed random variables, but will very likely have a Gaussian distribution as well.

This will come up when we talk about regression modeling. But when we look at the error vector from a regression, the feasibility of the error vector being close to Gaussian is plausible and reasonable given the way it's constructed sometimes.

OK. Well, let's move on then to this note on probability theory for asset pricing. What's really neat about this note is that we're going to introduce you just with basic probability theory to one of the more important models in quantitative finance, or at least in early quantitative finance, which is the capital asset pricing model.

And so let's consider a simple context where we have one risky asset, like a stock. And we have a one period model where we have-- the beginning of the period is t0, and the end of the period is t1, I guess. And there's this risky asset, a-- so there's a risky asset, a. And we'll consider a market agent having Qa shares of the asset, and just Q units of a risk-free asset.

Now, the risk-free asset is like putting money in the bank and being confident that you'll get it back with interest. And we'll assume that the risk-free asset costs $1-- so $1 per unit. And the risky asset will be of price Pa per share.

So this market agent has a wealth at time 0. So W0 is going to be equal to qa pa plus q times $1. And X-- OK, so maybe we'll put an f here just to indicate that it's the risk-free asset.

Now, at time T1-- so this is the t0 prices are given here. At T1, the risk-free asset is going to be worth 1 plus rf. And the price of the risky asset, we'll assume that it's a random variable given by f tilde. And for simplicity, we'll assume that it has some mean, mu, and variance sigma squared.

Now, right away this is maybe a bad assumption to make because Gaussian or normal distributions can be arbitrarily large negative. But we'll assume that the likelihood of impossible values is very close to 0, so that this is a useful model. So this is the time T1 state of the market, which has certainty for the risk-free investment but uncertain outcome for the risky asset.

Well, one could consider the agent being able to buy x units of the risky asset at the initial time point, t0, and then have end of period wealth that reflects that trade of buying X units at t0.

So at the end of the period t1, time, there will be Qa plus X shares with value f tilde. And the x shares will have been bought by paying Pa, the t0 price at time 0. And so there will be money invested in the risk-free asset earning the interest rate, or the absolute return of 1 plus rf.

Now, what we want to do is consider what is the optimal choice of X-- the number of shares to buy for this agent? And in order to make that choice, one can apply the decision theory under uncertainty and associate a utility of the end period wealth given by some utility function, u. And we can choose the X value to maximize the expected utility of wealth.

So importantly, If we have u of w as a function of w, and there's basically a w0 value here-- utility of w0-- there are some standard assumptions one makes as to what the utility function is like. And we might assume that the first derivative and second derivative exist. So we'll assume that these exist.

And so-- well, the first derivative is likely to be positive. More wealth is preferred to less. So we have a better direction being up and to the right. And the second derivative, this actually could be less than 0 if there's a decreasing marginal utility to more wealth. So this function might go and look like this with a U prime greater than 0 and U double prime less than 0.

So these kinds of assumptions are reasonable to assume. And the theory of rational decision-making under uncertainty, if this is indeed your utility function, is to find the action X that maximizes this. And if we are lucky, we might be able to solve for X by satisfying the first order condition, the derivative of the expected utility equaling 0 at the maximum value. And-- yeah?

AUDIENCE: Why would you use anything other than the identity function for U? Because surely you just want to maximize your end of period wealth?

PROFESSOR: Let's see. If you use the identity, then there's-- well, it will turn out there's no penalty for a risk that you're undertaking. And so with this decreasing marginal utility of wealth, you end up having a penalty for taking risks. And so the-- yeah. So I'll leave it there, but it's a very good question.

And let's see. There's actually some-- well, there's some really interesting discussions and about how to optimally invest, as this case. And there was a controversy between Claude Shannon and I think it was Milton Friedman about maximizing the exponential-- or the expected log return of investments just period and that being optimal. So anyway, we'll return to that discussion later as well.

But for now, let's assume that this is a reasonable thing to do. And in order to solve this problem, we're actually going to use a lemma called Stein's Lemma, which is a curious mathematical result dealing with normally distributed y variable and y star. They're jointly distributed with each other and Gaussian.

If we look at the expected value of g of Y times Y minus mu, and the expected value of g prime of Y, if both of those exist, then we get these simple formulas for essentially the covariance between g of Y and Y, and the covariance of g and Y and Y star. So this is a mathematical theorem that is abstract as it's stated, but it ends up being applicable in our problem.

So if we define our wealth at the end of the period at time t1 to be given by this w tilde vector, then if the outcome of the asset price, f tilde, if this is Gaussian, then this wealth at time t1 will also be Gaussian because we're just taking a multiple of that plus a constant. So that note is simple.

Then if we look at the expected utility and try to take the derivative of that with respect to X, then under certain conditions we can take the expectation of the derivative with respect to X of the integrand with the expectation.

Now, under what conditions is it feasible to differentiate underneath the integral sign when you're doing computations in calculus? I mean, basically what we want to do is take the expected value of U of w tilde plus delta and subtract the expectation of U of w tilde and divide by delta, and take the limit as this goes to 0. So that's the formula for computing the derivative.

It turns out we can do that if the first derivative of U is bounded or has bounded expectation. And so if that's possible to do, then we can take the derivative underneath the integral sign. And we have basically the derivative of U multiplied by the chain rule, the derivative of w tilde with respect to X. So if we take the derivative of w tilde with respect to X, we get f tilde minus Pa times 1 plus rf. So this is just the chain rule of differentiation there.

Now with this expression in black, we can use the-- well, let's see. We can express this as the expected product of two random variables is equal to the covariance plus the product of the expectations. So here, we have U prime and f tilde minus Pa1 plus rf. The covariance between those two plus the product of their expectations is equal to that expected product.

Then, by Stein's Lemma, one actually can simplify or re-express this covariance as the expected second derivative of U times the covariance of w and f. And so we can then substitute this expression into this last expression here and get 0 equals that covariance plus the product of the first expectations there. And so we're substituting in this expected second derivative of U times the covariance in that formula.

Now, with this expression here, we have this covariance of w and f. Well, the covariance between w and f is simply the covariance of f with a multiple of f. So it equals that multiple times the variance of f. And so we plug that in to this covariance term.

And what's curious, and is the punch line, is that this final equation gives us a formula for how we solve for X. And so the value of X that we should choose will basically satisfy this equation. And so our X value will depend upon what the price of the asset is. And it actually is equal to a discounted value of the expected price plus this other term here.

And this other term here-- the second derivative is negative. First derivative is positive. So this is actually a negative factor here. And so we end up getting that-- we have a formula that depends upon that ratio of second derivative expectation to first derivative expectation of the utility function.

Now, in economics classes, maybe some of you have come across the constant absolute risk aversion utility function. Has anybody come across that in an economics class? Maybe. Maybe not. But that function, which is a constant minus e to the minus Aw, this is the constant absolute risk aversion utility function.

And so it basically has the right shape in terms of marginal utilities decreasing and utilities always increasing. And so with this particular utility function, we have the U double prime divided by U prime, those actually giving us just the factor A here.

So we have our price of the asset should equal the discounted expected cash flow minus a risk adjustment. And the risk adjustment in the price, that's like the fair price for the asset, varies depending upon what the risk aversion coefficient is of the agent. And also the magnitude of the shares of the risky asset that are obtained and the variance of the price outcome. So the fair price for the market agent basically is balancing all of these factors.

And let's see if-- OK. Well, with this result for a single asset, what's really neat is that we can generalize from one asset to many and set up the whole problem with almost the same notation. So we consider defining a Q vector of shares of assets 1 through n-- a1 through an-- and Qf units of the risk-free asset.

And we can go through the same logic of the initial wealth being the Q shares per asset times the price at time 0, t0, of the assets plus the value of the risk-free investment. And we can assume that the random prices at time t1 are jointly normally distributed. And so we then can think of the market agent basically adjusting the initial shares Q by the X vector and choosing the X vector to maximize expected utility.

And we can just plug everything in now with vectors instead of scalars, and the same arguments apply. And we end up getting, at the end of the day, different equilibrium prices for the different assets that are discounted expected values of the price plus a penalty depending upon the risk aversion of the agent. And so we end up getting this final result here.

Now, what's particularly interesting from this setup is to then say, well, what prices should be holding if this market model is in equilibrium? And so if the market is in equilibrium, then we should have expected returns of risky assets that satisfy this formula here. And if we look at an initial endowment of no shares at all, and define the X allocation to be an equilibrium investment in the market portfolio, then our market portfolio defined by X will have a market value of f tilde m, the sum of the Xi's shares times f tilde i.

The price initially of the market portfolio will be just the sum of the Xi shares times the initial price as Pai. And the return of the market portfolio will just be the change in value divided by the initial value, which corresponds to a weighted sum of the asset changes or percentage returns.

So the percentage return r tilde m is simply the weighted returns of the individual assets. And this is our expected return of the market portfolio when the underlying assets satisfy these pricing conditions.

Now, if we expand this expected market return expression, we can successively derive that it equals the risk-free rate, plus this blue factor times the covariance of the initial w vector with rm. And this blue expression can be solved for as the excess return of the market portfolio divided by the covariance term here. And so we end up getting that this covariance of w and rm end up being simply pm times the covariance of rm, or the variance of rm.

And plugging this in, we finally get that the expected returns of different assets are simply a beta factor times the excess return-- expected return of the market portfolio over the risk-free rate. And so this expression is actually the expression for what's called the security market line, where we have the expected return of asset i is the risk-free rate, plus a beta factor times the excess return of the market over the risk-free rate.

And this beta factor is the ratio of the covariance of asset i and the market portfolio divided by the variance of the market portfolio. And so this beta i is equivalent to the regression coefficient of the excess return of stocks regressed on the excess return of the market.

So in this model, we end up having that different assets will have different returns. And those returns are bounded below by rf, but then are increased by having a higher beta factor multiplying the excess return of the market. So you may be familiar with issues of betas in equity markets and their being low betas or high beta stocks.

In this model, we have that the way you get higher expected return is to have a higher beta. And so this also just leads to understanding what the equilibrium prices are of individual assets. If we take those expressions for returns and then reverse them into expressions for the equilibrium prices, we have that the equilibrium prices are the expected terminal price divided by a discount factor that discounts by the discount rate of the risk-free investment, but adding to that a penalty based on the level of risk that's being taken.

And here are just some references on this topic that you may want to check out. The development here was actually taken from the first reference by Frank de Jong and Barbara Rindi. There are other important citations. One of them that's of interest is John Lintner, "The Valuation of Risky Assets."

The capital asset pricing model was actually formulated by John Lintner and William Sharpe back in the early '60s. And I distributed one of the articles of Lintner in the Canvas site. And it's actually nice to be able to read a very technical paper in an accessible way. And so I think that you'll enjoy checking that paper out.

OK. Well, let me turn then to the main topic in our syllabus for today, but it's the tail end of today, to talk about stochastic processes, concepts. And one of the most important concepts in stochastic processes is the concept of a Martingale process.

And so if we have a stochastic process given by a sequence of random variables Xn, we can define a related process Mn, which is defined as functions of the first n X's for each n. So if we have Mn depending only on the first nX's in some well-defined way given by a function fn, then we will say that this Mn sequence, which is random, because of its dependence on the X's, it will be a Martingale if the expected value of the n-th M, given the first n minus 1 X's, is simply equal to the n minus first value of M. And we also are going to require that the magnitude of the Mn's as a finite expectation for all n too.

Now, let's consider perhaps the simplest example of a Martingale, which is basically a random walk. Suppose the Xn sequence is just independent random variables that each has mean 0, and Sn is the sum of the first nX's. Then this Sn is a random walk with steps given by X. And it's going to be a Martingale with respect to the original process of Xn's.

Now, we can make this example, even simpler by assuming that the X's are independent and identically distributed, and maybe that they're discrete with just a plus 1 or minus 1 step increment with equal probability. So these X's have expectation equal to 0 and variance equal to 1.

And the process Sn has expectation 0, the sum of the expectations. And the variance of Sn is the variance of a sum of independent steps, increments. And so it's just the sum of the variances. So that's n.

Well, here's a random walk realization of this model with 100 steps. And so it basically starts at time 0, and then as steps occur, increments up or down plus or minus 1. If we increase from 100 steps to 1,000, this is what a realization looks like. And then here's an extension from 1,000 steps to 10,000 steps.

And what's rather significant is the general shape of this random walk process as we increase the number of steps. It becomes this very jagged path. And it actually corresponds to Brownian motion models that we'll come across.

Now, if we were to realize 10 different paths of length 10,000, here are examples of that. And so one can see that for low number of steps there's not very much variability that increases as we increase the number of steps. And here, we're just drawing bands at plus or minus multiples of the standard deviation of the random walk value.

And so these standard deviation levels basically apply just through the mathematics of calculating one or two standard deviations of the sum. And what we know from our central limit theorem is that the distribution of possible values at any given time step will be closely modeled by a Gaussian distribution because of the central limit theorem.

[BELL RINGING]

We have a stop. OK, I'm going to take two more minutes just to show you a more-- not a more interesting, but a related but very interesting Martingale. Suppose that once we've calculated this random walk Sn, we look at the square of Sn and subtract n sigma squared.

It turns out that this is a martingale as well. And here's one realization over 10,000 steps, and here are 10 realizations. And so these look very, very different from the simple random walk. But these, as well, are martingales. And so one of the-- what we'll see next time is how martingales have very special properties which can be used to solve challenging probability problems.

And just as a closing remark, does anyone here know where the term "martingale" comes from? Do we have any equestrians? There's a bridle and a horse called a martingale bridle that keeps the horse straight, looking straight. So it is a bridle that has pressure on both sides of the mouth to keep the horse looking straight.

And what's true about a martingale is that the expected value in the future is always a constant level equal to the last level of the series. So that property will end up being very, very convenient, and we'll see examples of that next time.