

The following content is provided under a Creative Commons license. Your support will help MIT OpenCourseWare continue to offer high quality educational resources for free. To make a donation or view additional materials from hundreds of MIT courses, visit MIT OpenCourseWare at [ocw.mit.edu](https://ocw.mit.edu).

**PROFESSOR:** Today's topic is factor modeling, and the subject here basically exploits multivariate analysis in statistics to financial markets where our concern is using factors to model returns and variances, covariances, correlations. And with these models, there are two basic cases. There's one where the factors are observable.

Those can be macroeconomic factors. They can be fundamental factors on assets or securities that might explain returns and covariances. A second class of models is where these factors are hidden or latent. And statistical factor models are then used to specify these models.

In particular, there are two methodologies. There's factor analysis and principal components analysis, which we'll get into some detail during the lecture. So let's proceed to talk about the setup for a linear factor model.

We have  $m$  assets, or instruments, or indexes whose values correspond to a multivariate stochastic process we're modeling. And we have  $n$  time periods  $t$ . And with the factor model we model the  $t$ -th value for the  $i$ -th object-- whether it's a stock price, futures price, currency-- as a linear function of factors  $F_1$  through  $F_K$ . So there's basically like a state space model for the value of the stochastic process, as it depends on these underlying factors.

And the dependence is given by coefficients  $\beta_1$  through  $\beta_k$ , which are depending upon  $i$ , the asset. So we allow each asset, say if we're thinking of stocks, to depend on factors in different ways. If a certain underlying factor changes in value, the  $\beta$  corresponds to the impact of that underlying factor.

So we have common factors. Now these common factors,  $f$ , this is really going to be a nice model if the number of factors that we're using is relatively small. So the

number  $k$  of common factors is generally very, very small relative to  $m$ . And if you think about modeling, say asset equity, asset returns in a market, there can be hundreds and thousands of securities.

And so in terms of modeling those returns and covariances, what we're trying to do is characterize those in terms of a modest number of underlying factors which simplifies the problem greatly. The vectors  $\beta_i$  are termed the factor loadings of an asset. And the  $\epsilon_{it}$ 's are a specific factor of asset  $i$ , period  $t$ .

So in factor modeling, we talk about there being common factors affecting the dynamics of the system. And the factor associated with particular cases are the specific factors. So this setup is really very familiar. It just looks like a standard sort of regression type model that we've seen before. And so let's see how this can be set up as a set of cross-sectional regressions.

So now we're going to fix the period  $t$ , the time  $t$ , and consider the  $m$  variate  $x$  variable as satisfying a regression model with intercept given by  $\alpha$ 's. And then the independent variables matrix is  $b$ , given by the coefficients of the factor loadings. And then we have the residuals  $\epsilon_t$  for the  $m$  assets.

So the cross-sectional terminology means we're fixing time and looking across all the assets for one fixed time. And we're trying to explain how, say, the returns of assets are varying depending upon the underlying factors. And so the-- well OK, what's random in this process? Well certainly the residual term is considered to be random. That's basically going to be assumed to be white noise with mean 0.

There's going to be possibly a covariance matrix  $\psi$ . And it's going to be uncorrelated across different time cross sections. Let's see if I can move the mouse, if this is what's causing the problem down here.

So in this model we have the realizations on the underlying factors being random variables. The returns on the assets depend on the underlying factors. Those are going to be assumed to have some mean,  $\mu_f$ , and some covariance matrix. And basically the dimension of that covariance matrix  $\omega_f$  is going to be  $k$  by  $k$ .

So in terms of modeling this problem, we go from an  $m$  by  $m$  system of covariances, correlations, to focusing initially on an  $a$   $k$  by  $k$  system of covariances and correlations between the underlying factors.  $\Psi$  is a diagonal matrix with the specific variances of the underlying assets. So we have basically  $\epsilon$ -- the covariance matrix of the  $\epsilon$ s is a diagonal matrix, and the covariance matrix of  $f$  is given by this  $\Omega$ .

Well, with those specifications we can get the covariance for the overall vector of the  $m$  variance stochastic process. And we have this model here for the conditional moments. Basically, the conditional expectation of the process given the underlying factors is this linear model in terms of the underlying factors  $f$ . And the covariance matrix is the  $\Psi$  matrix, which is diagonal.

And the unconditional moments are obtained by just taking the expectations of these. Well actually, the unconditional expectation of  $x$  is this. The unconditional covariance of  $x$  is actually equal to the expectation of this plus the variance of the conditional expectation, or the covariance of the conditional expectation.

So one of the formulas that's important to realize here is that if we're considering the covariance of  $x_t$ , which is equal to covariance of  $Bf_t$  plus  $\epsilon_t$ , that's equal to the covariance of  $Bf_t$  plus the covariance of  $\epsilon_t$  plus twice the covariance between this term and this, but those are uncorrelated. And so this is equal to  $B$  covariance of  $f_t$   $B$  transposed plus  $\Psi$ .

With  $m$  assets, how many parameters are in the covariance matrix if there's no constraints on the covariance matrix?

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** How many parameters? Right. So this is  $\sigma$ . So the number of parameters in  $\sigma$ .

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:**  $m$  plus what?

**AUDIENCE:** [INAUDIBLE].

**PROFESSOR:** OK, this is a square matrix,  $m$  by  $m$ . So there's possibly  $m$  squared, but it's symmetric. So we're double-counting off the diagonal. So it's  $m$  times  $m$  plus 1 over 2.

How many parameters do we have with the factor model? So if we think of  $\Sigma$ -- let's call this sigma star. The number of parameters in sigma star is what?

Well,  $B$  is an  $m$  by  $k$  matrix. This is  $m$  by  $k$ , so we have possibly  $m$  times  $k$  values. The  $\Sigma$  is-- or the covariance of  $f$  is the number of elements in the covariance matrix of  $f$ , which is  $k$  by  $k$ .

And then we have  $\psi$ , which is a diagonal of dimension  $m$ . So depending on how we structure things, we can have many, many fewer parameters in this factor model than in the unconstrained case. And we're going to see that we can actually reduce this number in the covariance matrix of  $f$  dramatically because of flexibility in choosing those factors.

Well let's also look at the interpretation of the factor model as a series of time series regressions. You remember when we talked about multivariate regression a few lectures ago, we talked about cross-sectional regressions and time series regressions, and looking at the collection of all the regressions in a multivariate regression setting. Here we can do the same thing. In contrast to the cross-sectional regression where we're fixing time and looking at all the assets, here we're looking at fixing the asset  $i$  and the regression over time for that single asset.

So the values of  $x_i$ , ranging from time one up to time capital  $T$  basically follows a regression model that's equal to the intercept  $\alpha_i$  plus this matrix  $f$  times  $\beta_i$ , where  $\beta_i$  corresponds to the regression parameters in this regression, but they are the factor corresponding to an asset  $i$  on the different  $k$  factors. In this setting, the covariance matrix of the  $\epsilon_i$  vector is the diagonal matrix  $\sigma^2$  times the identity. And so this is the classic Gauss Markov assumptions for a single linear regression model.

Well as we did previously, we can group together all of these time series regressions for all the  $m$  assets together by simply putting them all together. So we start off with  $x_i$  equal to basically  $f + \beta_i + \epsilon_i$ . And we can basically consider  $x_1, x_2, \dots, x_n$ .

So we have a  $T$  by  $m$  matrix for the  $m$  assets. And that's equal to a regression model given by basically what's on the slides here. So basically, we're able to group everything together and deal with everything all at once, which computationally is applied in fitting these.

Let's look at the simplest example of a factor model. This is the single factor model of Sharpe. We went through the capital asset pricing model, how returns on assets and stocks are basically-- the excess return on stock can be modeled in terms as a linear regression on the excess return of the market. And the regression parameter  $\beta_i$  corresponds to the level of risk associated with the asset.

And all we're doing in this model is, by choosing different assets we're choosing assets with different levels of risk scaled by the  $\beta_i$ . And they may have returns that vary across assets given by  $\alpha_i$ . The covariance matrix of the assets has-- the unconditional covariance matrix has this structure. It's basically equal to the variance of the market times  $\beta \beta'$  plus  $\psi$ .

And so that equation is really very simple. It's really self-evident from what we've discussed, but let me just highlight what it is.  $\sigma^2 \beta \beta' + \psi$ . And that's equal to  $\sigma^2$  times basically a column vector of all the betas,  $\beta_1$  down to  $\beta_m$  times its transpose plus a diagonal matrix with its  $\psi$ .

So this is really a very, very simple structure for the covariance. And if you had wanted to apply this model to thousands of securities, it's basically no problem. You can construct a covariance matrix.

And if this were appropriate for a large collection of securities, then the amount-- the reduction in terms of what you're estimating is enormous. Rather than estimating each cross correlation and covariance of all the assets, the factor model tells us

what those cross covariances are. So that's really where the power of the model comes in.

And in terms of why is this so useful, well in portfolio management one of the key drivers the asset allocation is the covariance matrix of the assets. So if you have an effective model for modeling the covariance, then that simplifies the portfolio allocation problem because you can specify a covariance matrix that you are confident with.

And also in risk management, effective models of risk management deal with, how do we anticipate what would happen if different scenarios occur in the market? Well, the different scenarios that can occur can be associated with what's happening to underlying factors that affect the system. And so we can consider risk management approaches that vary these underlying factors, and look at how that has an impact on the covariance matrix very directly.

Estimation of Sharpe's single index model. We went through before how we estimate the alphas and the betas. In terms of estimating the sigmas-- the specific variances-- basically, that comes from the simple regression as well. Basically, the sum of the squared estimated residuals divided by  $t - 2$ . Here we're assuming unbiasedness because we have two parameters estimated per model.

Then for the market portfolio, that basically has a simple estimate as well. The psi hat matrix would just be the diagonal of that-- the diagonal of the specific variances. And then the unconditional covariance matrix is estimated by simply plugging in these parameter estimates. So, very simple and effective if that single factor model is appropriate.

Now needless to say, a single factor model doesn't characterize the structure of the covariances and, or the returns typically. And so we want to consider more general models, multi-factor models. And the first set of models we're going to talk about are common factor variables that can actually be observed. Basically any factor that you can observe is a potential candidate for being a relevant factor in a linear factor model. The effectiveness of a potential factor is determined by fitting the model and

seeing how much contribution that factor makes to the explanation of the returns and the covariance structure.

Chen, Ross, and Roll wrote a classic paper in 1986. Now Ross is actually here at MIT. And with their paper, they looked at modeling. Rather than looking at these factors directly, including those in the model, they looked at transforming these factors into surprise factors. So rather than having interest rates just as a simple factor directly plugged into the model, it would be the change in interest rates. And additionally, not only just the change in interest rate, but the unanticipated change in interest rates given market information.

So their paper goes through modeling different macroeconomic variables with vector autoregression models, and then using those specify unanticipated changes in these underlying factors. And so that's where the power comes in. And that highlights how when you're applying these models, it does involve some creativity to get the most bang for the buck with your models. And the idea they had of incorporating unanticipated changes was really a very good one and is applied quite widely now.

Now with this setup, one basically-- if one has empirical data over times 1 through capital T, then if one wants to specify these models, one has their observations on the  $x_i$  process. You basically have observed all the returns historically. We also, because the factors are observable, have the  $f$  matrix as a set of observed variables. So we can basically use those to estimate the beta  $i$ 's and also estimate the variances of the residual terms with simple regression methods.

So implementing these is quite feasible, and basically applies methods that we have from before. So what this slide now discusses is how we basically estimate the underlying parameters. We need to be a little bit careful about the Gauss Markov assumptions.

You'll remember that if we have a regression model where the residual terms are uncorrelated and constant variance, then the simple linear regression estimates are

the best ones. If there is unequal variances of the residuals, and maybe even covariances, then we need to use generalized least squares. So the notes go through those computations in the formulas, which are just simple extensions of our regression model theory that we had in previous lectures.

Let me go through an example. With common factor variables that are using either fundamental or asset specific attributes, there's the approach of-- well, it's called the Barra Approach. This is from Barr Rosenberg. Actually, I have to say, he was one of the inspirations to me for going into statistical modeling and finance. He was a professor at UC-Berkeley who left academics very early to basically apply models in trade money.

As an anecdote, his current situation is a little different. I'll let you look that up. But anyway, this approach basically provided the Barra Approach for factor modeling and risk analysis, which is still used extensively today.

So with common factor variables using asset specific attributess-- in fact, sort the factor realizations are unobserved but are estimated in the application of these models. So let's see how that goes. Oh, OK, this slide talks about the Fama-French approach, which concerns-- OK, Fama and French, Fama of course we talked about him in the last lecture.

He got the Nobel Prize for his work in modeling asset price returns and market efficiency. Fama and French found that there were some very important factors affecting asset returns in equities. Basically, returns tended to vary depending upon the size of firms.

So if you consider small firms versus large firms, small firms tended to have returns that were more similar to each other. Large firms tended to have returns that were more similar to each other. So there's basically a big versus small factor that is operating in the market. Sometimes the market prefers small stocks, sometimes it prefers large stocks.

And similarly, there's another factor which is value versus growth. Basically, stocks

that are considered good values are stocks which are cheap, basically, for what they have. So you're basically getting a stock at a discount if you're getting a good value.

And value stocks can be measured by looking at the price to book equity. If that's low, then the price you're paying for that equity in the firm is low, and it's cheap. And that compares with stocks for which the price relative to the book value is very, very high.

Why are people willing to pay a lot for stocks? In that case, well it's because the growth prospects of those stocks is high, and there's an anticipation basically that the current price is just reflecting a projection of how much growth potential there is. Now the Fama French approach is for each of these factors to basically rank order all the stocks by this factor and divide them up into quintiles.

So say this is market cap. We can divide up all the stocks in-- basically consider a histogram, or whatever, of all the market caps of all the stocks in our universe. And then divide it up into basically the bottom fifth, the next fifth, and then-- it probably needs to go up-- the top fifth. And the Fama French approach says, well, let's look at an equal-weighted average of the top fifth. And basically, buy that and sell the bottom fifth.

And so that would be the big versus small market factor of Fama and French. Now, if you look at their work, they actually do the bottom minus the top, because the value tends to outperform the other. So they have a factor who's more positive values and associated more generally with positive returns. But that factor can be applied and used to correlate with individual asset returns as well.

Now, with the Barra Industry Factor-- this is just getting back to the Barra Approach-- the simplest case of understanding the Barra industry factor models is to consider looking at dividing stocks up into different industry groups. So we might expect that, say oil stocks will tend to move together and have greater variability or common variability. And that could be very different from utility stocks, which tend to actually be quite low risk stocks.

Utility companies are companies which are very highly regulated. And the profitability of those firms is basically overlooked by the regulators. They don't want the utilities to gouge consumers and make too much profit from delivering power to customers. So utilities tend to have fairly low volatility but very consistent returns, which are based on reasonable, from a regulatory standpoint, levels of profitability for those companies.

Well with an industry factor model, what we can do is associate factor loadings, which basically are loading each asset in terms of which industry group it's in. So we actually know the beta values for these stocks, but we don't know the underlying factor realizations for these stocks. But in terms of the betas, with these factors we can basically get a well defined beta vectors and b matrix for all the stocks. And the problem then is, how do we specify the realization of the underlying factors?

Well the realization of the underlying factors basically is just estimated with a regression model. And so if we have all of our assets  $x_i$  for different times  $T$ , those would have a model given by factor realizations corresponding to the  $k$  industry groups with known beta  $\beta_{ij}$  values.

And the estimation of these, we basically have a simple regression model where the realizations of the factor returns  $f_t$  are given by essentially a regression coefficient in this regression, where we have the asset returns  $x_t$ , the known factor loadings  $b$ , the unknown factor realizations  $f_t$ . And just plugging into the regression, if we do it very simply we get this expression for  $\hat{f}_T$ , which is the simple linear regression model estimating those realizations.

Now this particular estimate of the factor realizations is assuming that the variability of the components of  $x$  have the same variance. This is like the linear regression estimates under normal Gauss Markov assumptions. But basically the epsilon  $\epsilon_i$ 's will vary across the different assets. The different assets will have different variabilities, different specific variances. So that's actually going to be heteroscedasticity in these models. So this particular vector of industry averages should actually be extended to accommodate for that.

So we have the estimation of the covariance matrix of the factors can then be estimated using these estimates of the realizations. And our estimation of the residual covariance matrix can then be estimated. So I guess an initial estimate of the covariance matrix  $\hat{\sigma}$  is given by this known matrix  $B$  times our sample estimate of the factor realizations plus the diagonal matrix  $\hat{C}$ .

And a second step in this process can incorporate information about there being heteroscedasticity along the diagonal of the size to adjust the regression estimates. So we basically get a refinement of the estimates that does account for the non-constant variability. Now this property of heteroscedasticity versus homoscedasticity in estimating the regression parameters, it may seem like this is a nicety of the statistical theory that we just have to try and check, but it's not too big a deal.

But let me highlight where this issue comes up again and again. With portfolio optimization, we went through last time for mean variance optimization, we want to consider a weighting of assets that basically weights the assets by the expected returns, premultiplied by the inverse of the covariance matrix. And so we basically in portfolio allocation want to allocate to assets with high return, but we're going to penalize those with high variance.

And so the impact of discounting values with high variability arises in asset allocation. And then of course arises in statistical estimation. Basically with signals with high noise, you want to normalize by the level of noise before you incorporate the impact of that variable on the particular model.

So here are just some notes about the inefficiency of estimates due to heteroscedasticity. We can apply generalized least squares. A second bullet here is that factor realizations can be scaled to represent factor mimicking portfolios. Now with the Fama French factors, where we have say big versus small stocks or value versus growth stocks, it would be nice to know, well what's the real value of trading that factor? If you were to have unit weight to trading that factor, would you make money or not? Or under what periods would you make money?

And the notion of factor mimicking portfolios is important. Let's go back to the specification of the factor realizations here.  $f_{tT}$ , the  $t$ -th realization of the factors, their  $k$  factors, is given by essentially the regression estimate of those factors from the realizations of the asset returns. And if we're doing this in the proper way, we'd be correcting for the heteroscedasticity.

Well this realization of the factor returns is a weighted average or a weighted sum of the  $x_t$ . So we have basically  $f_t$  is equal to a matrix times  $x_T$ , where this is  $B$  prime. So our  $k$  dimensional realizations-- let's see, this is basically  $k$  by  $1$ .

Each of these  $k$  factors is a weighted sum of these  $x$ 's. Now the  $x$ 's, if these are returns on the underlying assets, then we can consider normalizing these factors. Or basically normalizing this matrix here so that this sort of row weights sum to  $1$ , say, for a unit of capital.

So if we were to invest and next unit of one capital unit in these assets, then this factor realization would give us the return on a portfolio of the assets that is perfectly correlated with the factor realization. So factor mimicking portfolios can be defined that way. And they have a good interpretation in terms of the realization of potential investments. So let's go back.

The next subject is statistical factor models. This is the case where we begin the analysis with just our collection of outcomes for the process  $x_T$ . So basically our time series of asset returns for  $m$  assets over  $T$  time units. And we have no clue initially what the underlying factors are, but we hypothesize that there are factors that do characteristic the returns. And factor analysis and principal components analysis provide ways of uncovering those underlying factors and defining them in terms of the data themselves.

So we'll first talk about factor analysis. Then we'll turn to principal components analysis. Both of these methods are efforts to model the covariance matrix. And the underlying covariance matrix for the assets  $x$  can be estimated with sample data in terms of the sample covariance matrix.

So here I've just written out in matrix form how that would be computed. And so with this  $m$  by  $T$  matrix  $x$ , we basically take that matrix, take out the means by computing the means with multiplying by this matrix, and then take the sum of deviations about the means for all the  $m$  assets individually and across each other, and divide that by capital  $T$ .

Now, the setup for statistical factor models is exactly the same as before, except the only thing that we observe is  $x_t$ . So we're hypothesizing a model where  $\alpha$  is basically a vector that is basically the vector of mean return of the individual assets.

$B$  is a matrix of factor loadings on  $k$  factors  $f_t$ . And  $\epsilon_t$  is white noise with mean 0 covariance matrix given by the diagonal. So the setup here is the same basic set up as before, but we don't have the matrix  $B$  or the vector  $f_t$ . Or, of course,  $\alpha$ .

Now in this setup, it's important that there is an indeterminacy of this model, because for any given specification of the matrix  $B$  or the factors  $f$ , we can actually transform those by a  $k$  by  $k$  invertible matrix  $H$ . So for a given specification of this model, if we transform the underlying factor realizations  $f$  by the matrix  $H$ , which is  $k$  by  $k$ , then if we transform the factor loadings  $B$  by  $H$  inverse, we get the same model.

So there is an indeterminacy here, or a-- OK, there's an indeterminacy of these particular variables, but there's basically flexibility in how we define the factor model. So in trying to uncover a factor model with statistical factor analysis, there is some flexibility in defining our factors. We can arbitrarily transform the factors by an invertible transformation in the  $k$  space.

And I guess it's important to note that what changes when we do that transformation? Well the linear function stays the same in terms of the covariance matrix of the underlying factors. Well, if we have a covariance matrix for those underlying factors, we need to accommodate the matrix transformation  $h$  in that. So that has an impact there.

But one of the things we can do is consider trying to define a matrix  $H$ , that diagonalizes the factors. So in some settings, it's useful to consider factor models where you have uncorrelated factor components. And it's possible to define linear factor models with uncorrelated factor components by a choice of  $H$ . So with any linear factor model in fact, we can have uncorrelated factor components if that's useful.

So this first bullet highlights that point that we can get orthonormal factors. And we can also have 0 mean factors by adjusting the data to incorporate the mean of these factors. And if we make these particular assumptions, then the model does simplify to just being the covariance matrix  $\Sigma$  is the factor loadings  $B$  times its transpose plus a diagonal matrix.

And just to reiterate, the power of this is basically no matter how large  $m$  is, as  $m$  increases the  $B$  matrix just increases by  $k$  for every increment in  $m$ . And We also have an additional [INAUDIBLE] on the side. So as we add more and more assets to our modeling, the complexity basically doesn't increase very much.

With all of our statistical models, one of the questions is how do we specify the particular parameters? Maximum likelihood estimation is the first thing to go through, and with normal linear factor models we have normal distributions for all the underlying random variables.

So the residuals  $\epsilon_T$  are independent and identically distributed, multivariate normal dimension  $m$  with diagonal matrix  $\Psi$  given by the individual elements variances.  $f_T$ , the realization of the factors the  $k$  dimensional factors can have mean 0. And just to have the identity covariance we can scale them and make them uncorrelated. And then  $x_T$  will be normally distributed with mean  $\alpha$  and covariance matrix  $\Sigma$  given by the formulas in the previous slide.

With these assumptions, we can write down the model likelihood. The model likelihood is the joint density of our data given the unknown parameters. And the standard setup actually for statistical factor modeling is to assume independence over time. Now we know that there can be time series dependence. We won't deal

with that at this point. Let's just assume that they are independent across time.

Then we can consider this as simply the product of the conditional density of  $x^T$ , given  $\alpha$  and  $\sigma$ , which has this form. This form for the density function of a multivariate normal should be very familiar to you at this point. It's basically the extension of the univariate normal distribution to  $m$  variate.

So we have  $1$  over the square root of  $\pi$  to the  $m$  power. There are  $m$  components. And then we divide by the square root of the individual variants or the determinant of the covariance matrix. And then the exponential of this term here, which for the  $t$ -th case is a quadratic form in the  $x$ 's. So this multivariate normal  $x$ , we take off its mean and look at the quadratic form of that with the inverse of its covariance matrix.

So there's the log likelihood function. It reduces to this form here. And maximum likelihood estimation methods can be applied to specify all the parameters of  $B$  and  $\psi$ . And there's an em algorithm, which is applied in this case. I think I may have highlighted it before, but the em algorithm is a very powerful estimation methodology for maximum likelihood in statistics.

When one has very complicated models which can be simplified-- well, models that are complicated by the fact that we have hidden variables-- basically the hidden variables lead to very complex likelihood functions. A simplification of the em algorithm is that if we could observe some of the hidden variables, then our likelihood functions are very simple and can be computed directly.

And the em algorithm alternates estimating the hidden variables, assuming the hidden variables are known doing the simple estimates with the observed hidden variables, and then estimating the hidden variables again, and just iterating that process again and again. And it converges.

And their paper demonstrates that the supplies in many, many different application settings. And it's just a very, very powerful estimation methodology that is applied here with statistical factor analysis. I indicated that for now we could just assume independence over time of the data points and computing its likelihood. You recall

our discussion a couple of lectures back about the state space models, linear state space models. Essentially, that linear state space model framework can be applied here as well to incorporate time dependence in the data as well.

So that simplification is it's not binding in terms of holding this up and estimating these models. Let me go back here, OK. So the maximum likelihood estimation process will give us estimates of the B matrix and the psi matrix. So applying this em algorithm, a good computer can actually get estimates of B and psi and the underlying alpha, of course.

Now from these we can estimate the factor realizations  $f_T$ . And these can be estimated by simply this regression formula, using our estimates for the factor loadings B had our estimates of alpha, we can actually estimate the factor realizations. So with statistical factor analysis, we use the em algorithm to estimate the covariance matrix parameters. Then the next step, we can estimate the factor realizations.

So as the output from factor analysis, we can work with these factor realizations. And those realizations or those estimates of the realizations of the factors can then be used basically for risk modeling as well. So we could do a statistical factor analysis of returns in, say, the commodities markets. And identify what factors are driving returns and covariances in commodity markets.

We can then get estimates of those underlying factors from the methodology. We could then use those as inputs to other models. Certain stocks may depend on significant factors in the commodity markets. And what they depend on, well we can use statistical modeling to identify where the dependencies are. So getting these realizations of the statistical factors is very useful, not only to understand what happened in the past with the process and how these underlying factors vary, but you can also use those as inputs to other models.

Finally, let's see, there was a lot of interest with statistical factor analysis on the interpretation of the underlying factors. Of course, in terms of using any model it's once confidence rises when you have highly interpretable results. One of the initial

applications of statistical factor analysis was in measuring IQ. And how many people here have taken an IQ test?

Probably everybody or almost everybody? Well actually if you want to work for some hedge funds, you'll have to take some IQ tests. But basically in an IQ test there are 20, 30, 40 questions. And they're trying to measure different aspects of your ability. And statistical factor analysis has been used to try and understand what are the underlying dimensions of intelligence.

And one has the flexibility of considering different transformations of any given set of estimated factors by this H matrix for transformation. And so there has been work in statistical factor analysis to find rotations of the factor loadings that make the factors more interpretable. So I just raise that as there's potential to get insight into these underlying factors if that's appropriate.

In the IQ setting, the effort was actually to try and find what are interpretations of different dimensions of intelligence? We previously talked about factor mimicking portfolios. The same thing applies.

One final thing is with likelihood ratio tests, one can test for whether the linear factor model is a good description of the data. And so with likelihood ratio tests, we compare the likelihood of the data where we fit our unknown parameters, the mean, vector,  $\alpha$ , and covariance matrix  $\sigma$  without any constraints. And then we compare that to the likelihood function under the factor model with, say,  $k$  factors.

And the likelihood ratio tests are computed by looking at twice the difference in log likelihoods. If you take an advanced course in statistics, you'll see that basically this difference in the likelihood functions under many conditions is approximately a chi squared random variable with degrees of freedom equal to the difference in parameters under the two models. So that's why it's specified this way. But anyway, one can test for the dimensionality of the factor model.

Before going into an example of factor modeling, I want to cover principal components analysis. Actually, principal components analysis comes up in factor

modeling, but it's also a methodology that's appropriate for modeling multivariate data and considering dimensionality reduction. You're dealing with data in very many dimensions. You're wondering is there a simple characterization of the multivariate structure that lies in a smaller dimensional space? And principal components analysis gives us that.

The theoretical framework for principal components analysis is to consider an  $m$  variate random variable. So this is like a single realization of asset returns in a given time, which has some mean and covariance matrix  $\sigma$ . The principal components analysis is going to exploit the eigenvalues and eigenvectors of the covariance matrix.

[INAUDIBLE] went through eigenvalues and singular value decompositions. So here we basically have the eigenvalue eigenvector decomposition of our covariance matrix  $\sigma$ , which is the scalar eigenvalues times the eigenvector  $\gamma_i$  times its transpose. So we actually are able to decompose our covariance matrix with eigenvalues eigenvectors.

The principal component variables are to consider taking away the mean from the random vector  $x$ . And then consider the weighted average of those de-meaned  $x$ 's given by the eigenvector. So these are going to be called the principal component variables, where  $\gamma_1$  is the first one corresponding to the largest eigenvalue.  $\gamma_m$  is going to be the  $m$ -th, or last, corresponding to the smallest.

The properties of these principal component variables is that they have mean 0, and their covariance matrix is given by the diagonal matrix of eigenvalues. So the principal component variables are a very simple sort of affine transformation of the original variable  $x$ . We translate  $x$  to a new origin, basically to the 0 origin, by subtracting the means off it.

And then we multiply that de-meaned  $x$  value by an orthogonal matrix  $\gamma'$ . And what does that do? That simply rotates the coordinate axes.

So what we're doing is creating a new coordinate system for our data, which hasn't changed the relative position of the data or the random variable at all in the space. Basically, it just is using a new coordinate system with no change in the overall variability of what we're working with. In matrix form, we can express this principal component variables  $p$ . So that's considered partitioning  $p$  into the first  $k$  elements and the last  $m$  minus  $k$  elements  $p_2$ .

Then our original random vector  $x$  has this decomposition. And we can think of this as being approximately a linear factor model. We can consider from principal components analysis essentially if  $p_1$  the principle component variables correspond to our factors, then our linear factor model would have  $B$  as given by  $\gamma_1$ ,  $f$  as given by  $p_1$ . And our epsilon vector would be given by  $\gamma_2$   $p_2$ .

So the principal components decomposition is almost a linear factor model. The only issue is this  $\gamma_2$   $p_2$  is an  $m$  vector, but it may not have a diagonal covariance matrix. Under the linear factor model with a given set of factors  $k$  less than  $m$ , we always are assuming that the residual vector has covariance matrix to a diagonal. With a principal components analysis, that may or may not be true.

So this is like an approximate factor model, but that's why this is called principal components analysis. It's not called principal factor analysis yet. The empirical principal components analysis now. We've gone through just a description of theoretical principal components, where if we have a mean vector  $\alpha$ , covariance matrix  $\sigma$ , how we would define these principle component variables.

If we just have sample data, then this slide goes through the computations of the empirical components results. So all we're doing is substituting in estimates of the means and covariance matrix, and computing the eigenvalue eigenvector decomposition of that. And we get sample principal component variables which are - we basically compute  $x$ , the de-meanded vector matrix of realizations and premultiply that by  $\hat{\gamma}'$ , which is the matrix of eigenvectors corresponding to the eigenvalue eigenvector decomposition of the sample covariance matrix.

This slide goes through the singular value decomposition. You don't have to go through and compute variances, covariances to derive estimates of the principle component variables. You can work simply with the singular value decomposition. I'll let you go through that on your own.

There's an alternate definition of their principle component variable thought that's very important. If we consider a linear combination of the components of  $x$ ,  $x_1$  through  $x_m$  given by  $w$ , if we consider a linear combination of that which maximizes the variability of that linear combination subject to the norm of the coefficients  $w$  equals 1, then this is the first principal component variable.

So if we have in two dimensions the  $x_1$  and  $x_2$ , if we have points that look like an ellipsoidal distribution, this would correspond to having  $\alpha_1$  there,  $\alpha_2$  there, a sort of degree of variability. The principal components analysis says, let's shift to the origin being at  $\alpha_1$   $\alpha_2$ . And then let's rotate the axes to align with the eigenvectors.

Well the first principal component variable finds the dimension at which the coordinate axis at which the variability is a maximum. And basically along this dimension here, this is where the variability would be the maximum. And that's the first principle component variable. So this principal components analysis is identifying essentially where's there the most variability in the data, where it's the most variability without doing any change in the scaling of the data? All we're doing is shifting and rotating.

Then the second principle component variable is basically the direction which is orthogonal to the first, which has the maximum variance. And continuing that process to define all  $m$  principle component variables. In principle components analysis, there's discussions of the total variability of the data and how well that's explained by principle components.

If we have a covariance matrix  $\sigma$ , the total variance can be defined and is defined as the sum of the diagonal entries. So it's the trace of a covariance matrix. We'll call that the total variance of this multivariate  $x$ . That is equal to the sum of the eigenvalues as well.

So we have a decomposition of the total variability into the variability of different principal component variables. And the principal component variables themselves are uncorrelated. You remember the covariance matrix of the principle component variables was the lambda, the diagonal matrix of eigenvalues.

So the off-diagonal entries are 0. So the principle component variables are uncorrelated, and have variability  $\lambda_k$ , and basically decompose the variability. So principal components analysis provides this very nice decomposition of the data into different dimensions, with highest to lowest information content as given by the eigenvalues.

I want to go through a case study here of doing factor modeling with the U.S. Treasury yields. I loaded in data into  $r$  which ranged from the beginning of 2000 to the end of May, 2013. And here are the yields on constant maturity U.S. Treasury securities ranging from 3 months, 6 months, up to 20 years.

So this is essentially the term structure of US Government [INAUDIBLE] of varying levels of risk. Here's a plot of over that period. So starting in the [INAUDIBLE], we can see this, the rather dramatic evolution of the term structure over this entire period. If we wanted to have set any [INAUDIBLE].

If we wanted to do a principle components analysis of this, well if we did the entire period we'd be measuring variability of all kinds of things. When things go down, up, down. What I've done in this note is just initially to look at the period from 2001 up through 2005. So we have five years of data on basically the early part of this period that I want to focus on and do a principal components analysis of the yields on this data.

So here's basically the series over that five year period. Beginning of this analysis, this analysis is on the actual yield changes. So just as we might be modeling say asset prices over time and then doing an analysis of the changes, the returns, here we're looking to yield changes.

So first, you can see there's basically the average daily value for the different yield

tenants is ranging from 3 months up to 20. Those are actually all negative. That corresponds to the time series over that five year period. Basically the time series were all at lower levels from beginning to end on average.

The daily volatility is the daily standard deviation. Those vary from 0.0384 up to .0698 for-- is that the three year? And this is the standard deviation of daily yield changes where 1 is like 1%. And so basically it's between three and six basis points a day are the variation on the yield changes.

So that's something that's reasonable. When you look at the news or a newspaper and see how interest rates change from one day to the next, it's generally a few basis points from one day to the next. This next matrix is the correlation matrix of the yield changes. If you look at this closely, which you can when you download these results, you'll see that near the diagonal the values are very high, like above 90% for correlation.

And as you move across away from the diagonal, the correlations get lower and lower. Mathematically that is what is happening. We can look at these things graphically, which I always like to do.

Here is just a graph bar chart of the yield changes and the standard deviations of the yield changes, daily volatilities ranging from very short yields to long tenor yields, up to 20 years. So there's variability there. Here is a [INAUDIBLE] plot of the data. So what I've done is just looked at basically for every single tenor, this is say the 5 year, 7 year, 10 year, 20 year.

I basically plotted the yield changes of each of those against each other. We could do this with basically all nine different tenors, and we'd have a very dense page of a [INAUDIBLE] plot. So I split it up into looking just at the top and bottom block diagonals. But you can see basically how the correlation between these yield changes is very tight and then gets less tight as you move further away.

With the long tenors-- let's see, the short tenors-- one, one more. Here the short tenors, ranging from 3 year, 2 year, 1 year, 6 month, and so forth. So here you can

see how it gets less and less correlated as you move away from a given tenor. Well the principal components analysis gives us-- if you conduct a principle components, basically the standard output is first a table of how the variability of the series is broken down across the different component variables.

And so there's basically the importance of components for each of the nine component variables where it's measured in terms of the relative squared standard deviations of these variables relative to the sum. And the proportion of variance explained by the first component variable is 0.849. So basically 85% of the total variability is explained by the first principle component variable.

Looking at the second row, second in, 0.0919, that's the percentage of total variability explained by the second principle component variable. So 9%. And then for third it's around 3%. And it just goes down closer to 0,

There's a scree plot for principle components analysis, which is just a plot of the variability of the different principle component variables. So you can see whether the principle components is explaining much variability in the first few components or not. Here there's a huge amount of variability explained by the first principle component variable.

I've plotted here the standard deviations of the original yield changes in green, versus the standard deviations of the principle component variables in blue. So we basically are modeling with principal component variables most of the variability in the first few principle components. Now let's look at the interpretation of the principal component variables.

There's the loadings matrix, which is the gamma matrix for the principle components variables. Looking at numbers is less informative for me than looking at graphs. Here's a plot of the loadings on the different yield changes for the first principle component variable. So the first principle component variable is a weighted average of all the yield changes, giving greatest weight to the five year.

What's that? Well that's just a measure of a level shift in the yield curve. It's like,

what's the average yield change across the whole range? So that's what the first principle component variable is measuring.

The second principle component variable gives positive weight to the long tenors, negative weight to the short tenors. So it's looking at the difference between the yield changes on the long tenors versus the yield change on the short tenors. So that's looking at how the spread in yields is changing.

Then the third principle component variable has this structure. And this structure for the weights is like a double difference. It's looking at the difference between the long tenor and medium tenor, minus the medium tenor, minus the short tenor. So that's giving us a measure of the curvature of the term structure and how that's changing over time. So these principle component variables are measuring the level shift for the first, the spread for the second, and the curvature for the third.

With principle components analysis, many times I think people focus just on the first few principle component variables and then say they're done. The last principle component variable, and the last few, can be very, very interesting as well, because these are the variables of the original scales, the linear combinations which have the least variability.

And if you look at the ninth principle component variable-- there were nine yield changes here-- it's basically looking at a weighted average of the 5 and 10 year minus the 7 year. So this is like the hedge of the 7 year yield with the 5 and 10 year. So that difference in yield change is-- that combination of yield change is going to have the least variability.

The principle component variables have zero correlation. Here's just a [INAUDIBLE] plot of the first three principle component variables and the ninth. And you can see that those have been transformed to have zero correlations with each other.

One can plot the cumulative principal component variables over time to see how the evolution of these underlying factors has changed over the time period. And you'll recall that we talked about the first being the level shift. Basically from 2001 to 2005,

the overall level of interest rates went down and then went up. And this is captured by this first principal component variable accumulating from 0 down to minus 8, back up to 0.

And the scale of this change from 0 to minus 8 is the amount of greatest variability. The second principal component variable accumulates from 0 up to less than 6, back down to 0. So this is a measure of the spread between long and short rates. So the spread increased, and then it decreased over the period.

And then the curvature, it varies from 0 down to minus 1.5 back up to 0. So how the curvature changed over this entire period was much, much less, which is perhaps as it should be. But these graphs indicate basically how these underlying factors evolved over the time period.

In the case note I go through and fit a statistical factor analysis model to these same data and look at identifying the number of factors. And also comparing the results over this five year period with the period from 2009 to 2013, and comparing those different results. They are different, and so it really matters over what period one specifies these models to. And fitting these models is really just a starting point where you want to ultimately model the dynamics in these factors and their structural relationships. So we'll finish there.