Machine Learning in Finance

Overview

Overview

FORDHAM | Gabelli School of Business

THE JESUIT UNIVERSITY OF NEW YORK

Dimensionality Reduction and Unsupervised Learning

Dimentionality Reduction and Unsupervised Learning

- We now examine **unsupervised learning**
- We are no longer using a model to see how x-variables (covariates) help predict target y-variables. That is called **supervised learning**. We how how well or how badly the covariates explain the target variables
- Now we are interested in how we can reduce the dimensions of large numbers of covariates..
- When we look at big data, data sets can be big in two ways: large number of data observations and large number of classifiers.
- We wish to see if we can reduce the number of classifiers to a smaller more manageable numbers

Dimensionality Reduction and Unsupervised Learning

- Another reason why we use these methods is that often we refer to "the risk-free rate" or "the market" when we refer to the stock market performance.
- In reality there is no one "interest rates", there are a variety of short-term rates on all sorts of highly-liquid assets.
- There is no one market indicator. Dow Jones and S&P are **averages** over a select number of share offerings.
- To see how the market is doing, we can see how the market share prices cluster around different indicators.

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK
Gabelli School of Business

Principal Components, Part I

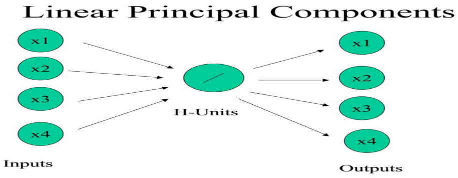
Principal Components, Part I

- The linear approach to reducing a larger set of variables into a smaller subset of "signals" from a large set of variables is called principal components analysis (PCA).
- PCA identifies linear "projections" or combinations of data which explain most of the variation of the original data, or extract most of the information from the larger set of variables, in decreasing order of importance.
- Obviously, and trivially, for a data set of K vectors, K linear combinations will explain the total variation of the data.
- But it may be the case that only two or three linear combinations or "principal components" may explain a very large proportion of the variation of the total data.

Principal Components, Part I

- If the underlying true structure interrelating the data is linear, then a few principal components or linear combinations of the data can capture the data "in the most succinct way"
- The resulting components are both uncorrelated and independent [Fotheringhame and Baddeley, Nonlinear Principal Components of Neuronal Spike Train Data, Springer-Verlag (1997): p. 1].
- The next figure illustrates the structure of the principal components "mapping".
- The H units in the hidden layer are linear combinations of the input variables.
- We call the "mapping" from the inputs to the H-units a "dimensionality reduction mapping".
- The mapping from the H-units to the output variables is a "reconstruction mapping"

Principal Components, Part I



FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Principal Components, Part II

Principal Components, Part II

- The method by which the coefficients linking the input variables to the H units are estimated is know as "orthogonal regression.
- Letting $X = [x_1 ... x_k]$ be a dimension T by k matrix of variables. We obtain the following eigenvalues λ_x and eigenvectors $\nu_x$through the process of "orthogonal regression" by the calculation of eigenvalues and eigenvectors:

$$[X\prime X - \lambda_x I]\nu_x = 0$$

- For a set of k regressors, there are, of course, at most k eigenvalues and k eigenvectors. The eigenvalues are ranked from the largest to the smallest.

Principal Components, Part II

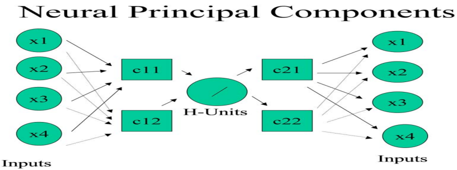
- We use the eigenvector ν_x associated with the largest eigenvalue to obtain the the first principal component of the matrix X. This first principle component is simply a vector of length T, computed as a weighted average of the k-columns of X, with the weighting coefficients being the elements of ν_x.
- In a similar manner, we may find second and third principal components of the input matrix by finding the eigenvector associated with the second and third largest eigenvalues of the matrix X, and multiplying the matrix by the coefficients from the associated eigenvectors.
- The normalized ratio of the first eigenvalue to the sum of the eigenvalues gives us the total explanatory power of the first principal component.

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK
Gabelli School of Business

Nonlinear Principal Components, Part I

Nonlinear Principal Components, Part I

- In this approach, input variables in this network are "encoded" by intermediate logsigmoid units, in a "dimensionality reduction" mapping.
- These encoding units are combined linearly to form H neural nonlinear principal components.
- The H-units in turn are "decoded" by two decoding logsigmoid units, in a "reconstruction mapping", which are combined linearly to "regenerate" the inputs as the output layers.
- It is not strictly required that such networks have equal numbers in the encoding and decoding layers. The next slide shows one type of network.


Neural Principal Components
x1
x2
x3
x4
c11
c12
H-Units
c21
c22
x1
x2
x3
x4
Inputs
Inputs

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Nonlinear Principal Components, Part II

- Such a system has the following representation, with EN as an "encoding neuron", and DN as a "decoding neuron".

$$EN_j = \sum_{k=1}^{K} \alpha_{j,k} X_k$$

$$EN_j = (1/(1 + exp(-EN_j)))$$

$$H_p = \sum_{j=1}^{J} \beta_{p,j} EN_j$$

$$DN_j = \sum_{p=1}^{P} \gamma_{j,p} H_p$$

$$DN_j = (1/(1 + exp(-DN_j)))$$

$$X_k = \sum_{j=1}^{J} \delta_{k,j} DN_j$$

Nonlinear Principal Components, Part II

- The H principal component units from linear orthogonal regression or neural network estimation are particularly useful for evaluating "expected" or "required" returns for new investment opportunities, based on the capital asset pricing model, better known as the CAPM.
- In its simplest form, this theory requires that the minimum required return for any asset or portfolio k, r_k, net of the risk free rate r_f,is proportional,by a factor β_k,to the difference between the observed market return, $r_m$less the risk free rate,:

$$r_k = r_f + \beta_k[r_m - r_f]$$

$$\beta_k = Cov(r_k, r_m)/Var(r_m)$$

- The coefficient β_k is widely known as the CAPM "beta"

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK
Gabelli School of Business

Nonlinear Principal Components, Part III

Nonlinear Principal Components, Part III

- In the CAPM literature, the actual return on any asset $r_{k,t}$ is a compensation for risk.
- The required return $r_{k,t}$ represents "diversifiable" risk in financial markets
- The appeal of the CAPM is its simplicity in deriving the minimum expected or required return for an asset or investment opportunity. In theory, all we need is information about the return of a particular asset k, the market return, the risk free rate, and the variance and covariance of the two return series.
- As a decision rule, it is simple and straightforward: if the current observed return on asset k at at time t, $r_{k,t}$, is greater than the required return, r_k, then we should invest in this asset.
- However, the limitation of the CAPM is that it identifies the market return with only one particular "market" return.

Nonlinear Principal Components, Part III

- Usually the "market" return is an index, such as the Standard and Poor or the Dow Jones, but for many potential investment opportunities, these indices do not reflect the relevant or benchmark "market return".
- The market average is not a useful signal representing the news and risks coming from the market. Not surprisingly, the CAPM model does not do very well in explaining for predicting the movement of most asset returns.
- The Arbitrage Pricing Theory (APT) was introduced as an alternative to the CAPM.
- The APT provides an approximate relation for expected or required asset returns by replacing the single benchmark "market return" with a number of unidentified factors, or principal components, distilled from a wide set of asset returns observed in the market.
- The Intertemporal Capital Asset Pricing Model (ICAPM) differs from the APT in that it specifies the benchmark market return index as one argument determining the required return,

Nonlinear Principal Components, Part III

- We may use one observed market return as one variable for determining the required return. But one may include other arguments as well, such as macroeconomic indicators which capture the systematic risk of the economy.
- The final remaining arguments can be the principal components, either from the linear or neural estimation, distilled from a wide set of observed asset returns.
- Thus, the "required return" on asset k, r_k, can come from a regression of these returns, on one overall market "index" rate of return, a set of macroeconomic variables (such as the yield spread between long and short-term rates for government bonds, the expected and unexpected inflation rates, industrial production growth, the yield between corporate high and low-grade bonds), and a reasonably small set of principal components obtained from a wide set of returns observed in the market.

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

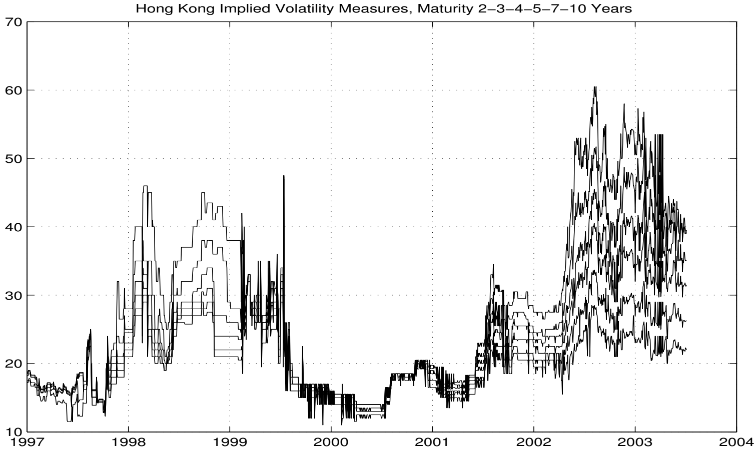
Implied Volatility in Hong Kong, Part I

Implied Volatility in Hong Kong, Part I

- We compare and contrast the implied volatility measures for Hong Kong and the United States, since we expect both of these to have similar features, due to the currency peg of the Hong Kong dollar to the U.S. dollar.
- But there may also be some differences, since Hong Kong was more vulnerable to the Asian financial crisis which began in 1997, and also had the SARS crisis in 2003. W
- We discuss both of these experiences in turn, and apply the linear and nonlinear dimensionality reduction methods for in-sample as well as for out-of-sample performance.

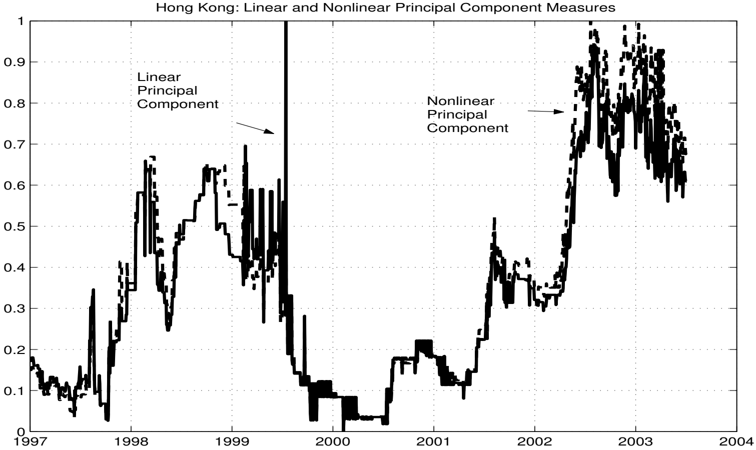
Implied Volatility in Hong Kong, Part I

- Hong Kong Implied Volatility



Implied Volatility in Hong Kong, Part I

- Linear and Nonlinear Principal Components for Hong Kong Implied Volatility



Implied Volatility in Hong Kong, Part I

- The first solid curve comes from the linear method. The second, given by the broken curve, comes from an auto-associative map or neural network.
- We estimate the network with five encoding neurons and five decoding neurons.
- For ease of comparison, we scaled each series between zero and one. What is most interesting is how similar both curves are. The linear principal component shows a big "spike" in mid-1999 but the overall volatility of the nonlinear principal component is slightly greater.
- The standard deviations of the linear and nonlinear components are, respectively, .233 and .272.

Implied Volatility in Hong Kong, Part I

- Hong Kong: In-sample Performance
- We see that both methods do rather well.

Hong Kong Implied Volatility Estimates
Goodness of Fit: Linear and Nonlinear Components
Multiple Correlation Coefficient

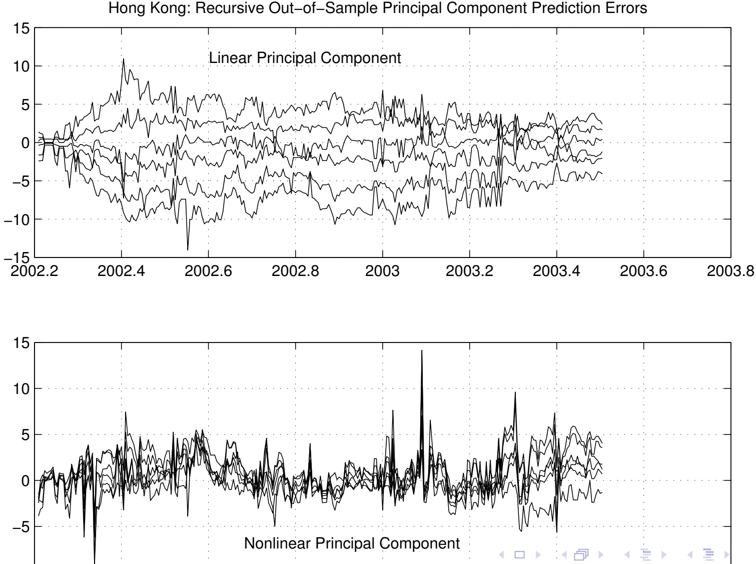
	Maturity in Years:					
	2	3	4	5	7	10
Linear	0.965	0.986	0.990	0.981	0.923	0.751
Nonlinear	0.988	0.978	0.947	0.913	0.829	0.698

FORDHAM
THE JESUIT UNIVERSITY OF NEW YORK
Gabelli School of Business

Implied Volatility in Hong Kong, Part II

Implied Volatility in Hong Kong, Part II

- HK Volatility Forecast errors with recursive estimation: Nonlinear PC wins!



Implied Volatility in Hong Kong, Part II

Hong Kong Implied Volatility Estimates
Out-of Sample Prediction Performance
Root Mean Squared Error

	Maturity in Years:					
	2	3	4	5	7	10
Linear	4.195	2.384	1.270	2.111	4.860	7.309
Nonlinear	1.873	1.986	2.985	2.479	1.718	1.636

Diebold-Mariano Tests*

	Maturity in Years:					
	2	3	4	5	7	10
DM-0	0.000	0.000	1.000	0.762	0.000	0.000
DM-1	0.000	0.000	1.000	0.717	0.000	0.000
DM-2	0.000	0.000	1.000	0.694	0.000	0.000
DM-3	0.000	0.000	1.000	0.678	0.000	0.000
DM-4	0.000	0.000	1.000	0.666	0.000	0.000

Note: * P-values
DM-0 to DM-4: tests at autocorrelations 0 to 4.

FORDHAM

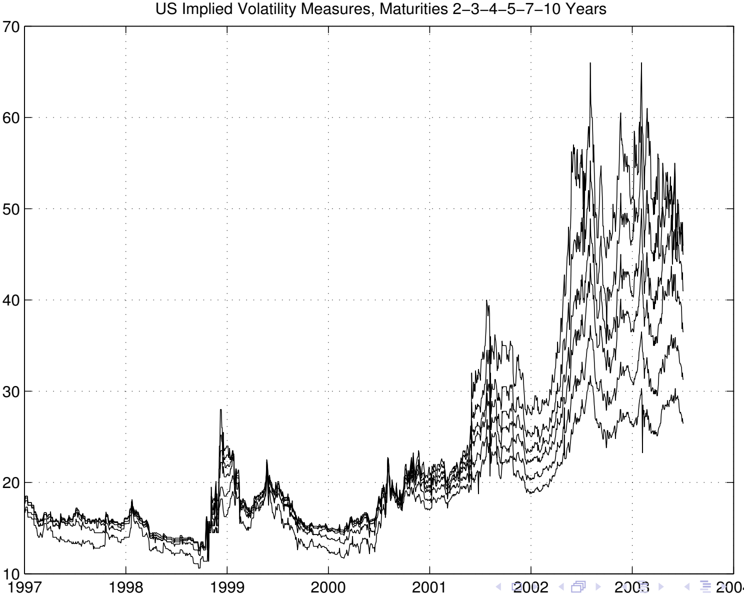
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

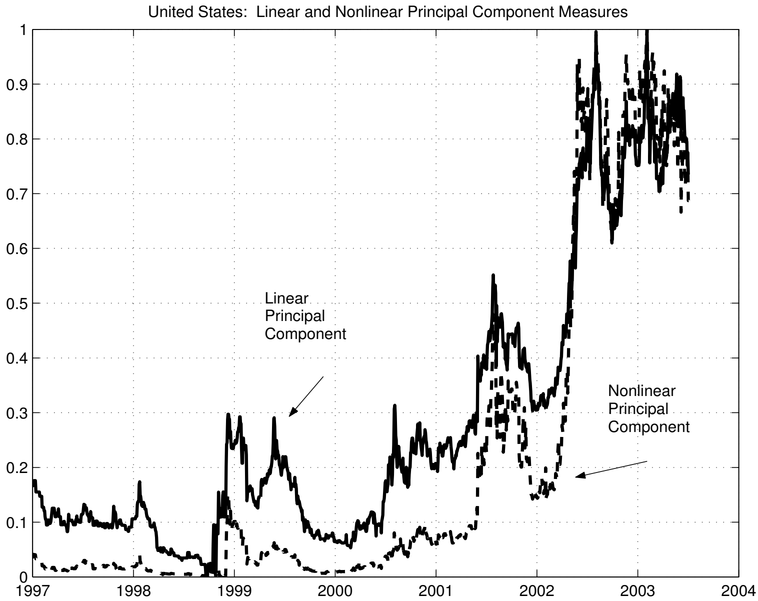
Implied Volatility in USA, Part I

Implied Volatility in USA, Part I

- USA: Implied Volatility



Implied Volatility in USA, Part I



Implied Volatility in USA, Part I

- In-sample performance for USA volatilities

United States Implied Volatility Estimates
Goodness of Fit: Linear and Nonlinear Components
Multiple Correlation Coefficient
Maturity in Years:

	2	3	4	5	7	10
Linear	0.983	0.995	0.997	0.998	0.994	0.978
Nonlinear	0.995	0.989	0.984	0.982	0.977	0.969

FORDHAM

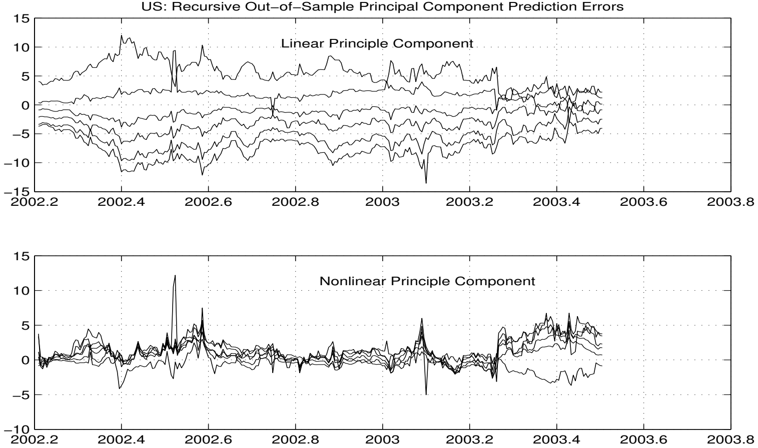
THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business

Implied Volatility in USA, Part II

Implied Volatility in USA, Part II

- USA Volatility Forecast Errors: Nonlinear PC wins.



Implied Volatility in USA, Part II

United States Implied Volatility Estimates
Out-of Sample Prediction Performance
Root Mean Squared Error

	Maturity in Years: 2	3	4	5	7	10
Linear	5.761	2.247	1.585	3.365	5.843	7.699
Nonlinear	1.575	2.249	2.423	2.103	1.504	1.207

Diebold-Mariano Tests*

	Maturity in Years: 2	3	4	5	7	10
DM-0	0.000	0.000	0.997	0.000	0.000	0.000
DM-1	0.000	0.002	0.986	0.000	0.000	0.000
DM-2	0.000	0.006	0.971	0.000	0.000	0.000
DM-3	0.000	0.011	0.956	0.000	0.000	0.000
DM-4	0.000	0.017	0.941	0.001	0.000	0.000

Note: * P-values
DM-0 to DM-4: tests at autocorrelations 0 to 4.

FORDHAM

THE JESUIT UNIVERSITY OF NEW YORK

Gabelli School of Business