INFO 7374 **Yizhen Zhao**
Summer I, 2018

Lecture 6 Machine Learning in Risk Management

Risk Management Framework for Financial Institution

- **Market Risk**
- **Credit Risk**
- **Operation Risk**

Market Risk Management

Portfolio Profits and Losses (P & L) Analysis: An Example

Risk Metrics	Year/Month/Day, say, 2017Q1
Gross Profit	$3,442,322
Gross Loss	($1,365,035)
Net Profit	$2077,287
Profit Factor	2.52
Num. of Days	91
Num. of Winning Rate	1399
Daily Win Rate	65%
Best Month	8.77%
Worst Month	(1.84%)
Total Num. of Trades	92,372
Percent Profitable	56%
Winning Trades	51,439
Losing Trades	40,933
Even Trades	3,904
Average Num. of Trades/Day	43
Ave. Profit &Loss per Trade	$22

Total Return	1004%
Average Daily Return	0.11%
Average Annual Return	26.75%
Annual Standard Deviation	4.97%
Skewness	0.60
Kurtosis	2.46
Sharpe Ratio	5.38
Sortino Ratio	10.08
99% 1-Day Value-at-Risk	-0.62%
Max 1-Day Loss	-1.01%
Max Drawdown	-2.52%
Drawdown Period	19
Recovery Period	33
Commission	$659,874
Slippage	$241,386

- **Sharpe Ratio**

The Sharpe ratio is the average return earned in *excess* of the risk-free rate per unit of volatility or total risk. Generally, the greater the value of the Sharpe ratio, the more attractive the risk-adjusted return.

$$Sharpe\ Ratio = \frac{R_p - R_f}{\sigma_p}$$

R_p: the expected portfolio returns.
R_f: the risk-free interest rate, usually the 3-month Treasury Bill yield.
σ_p: the standard deviation of portfolio excess returns.

- **Sortino Ratio**

The Sortino ratio is a variation of the Sharpe ratio that differentiates harmful volatility from total overall volatility by using the asset's standard deviation of negative asset returns, called downside deviation.

$$Sortino\ Ratio = \frac{R_p - R_f}{\sigma_d}$$

σ_d: the standard deviation of negative portfolio returns.

- **Treynor Ratio**

The **Treynor ratio** (sometimes called the **reward-to-volatility ratio** or **Treynor measure**), is a measurement of the returns earned in*excess* of whatcould have been earned onan investment that has no diversifiable risk (e.g., Treasury bills or a completely diversifiedportfolio), per each unit of market risk assumed.

$$Treynor\ Ratio = \frac{R_p - R_f}{\beta_p}$$

β_p: portfolio beta from CAPM regression: $R_m - R_f = \alpha + \beta\left(R_p - R_m\right) + \epsilon_t$.

The Treynor ratio relates excess return over the risk-free rate to the additional risk taken; however, systematic risk is used instead of total risk. The higher the Treynor ratio, the better the performance of the portfolio under analysis.

Value-at-Risk

Value-at-Risk (VaR) can be interpreted as the cutoff point such that a loss will not happen with probability greater than $p = 95\%$ percent, say. If $f(u)$ is the distribution of profit and losses on the portfolio, VAR is defined from

$$F(x) = \int_{-\infty}^{x} f(u)du = 1 - p$$

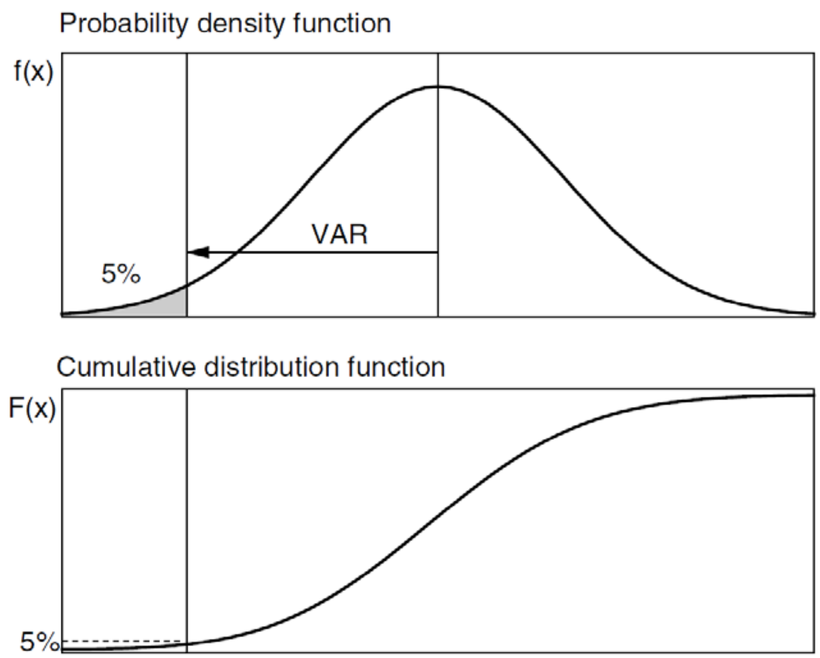
where is the right-tail probability, and c usual left-tail probability, VAR can then be defined as the deviation between the expected value and the quantile,

$$VAR(c) = E(X) - Q\ (X, c)$$

Important Classical Cases:

$$P(X \leq -1) = 1 - P(X \leq 1) = 1 - 0.8413 = 0.1587$$
$$P(X \geq 1.64) = P(x \leq -1.64) = 0.05$$
$$P(X \geq 1.28) = P(x \leq -1.28) = 0.1$$
$$P(X \geq 1.96) = P(x \leq -1.96) = 0.025$$

Examples



Risk-Neutral Density (State Price Density)

Options Basics

- A **call option** gives its holder the right to purchase an asset for a specified price, calledthe **exercise,** or **strike, price,** on or before some specified expiration date.
- A **put option** gives its holder the right to *sell* an asset for a specified exercise or strike price on or before some expiration date.
- An option is described as **in the money** when its exercise would produce profitsfor its holder. An option is **out of the money** when exercise would be unprofitable.Therefore, a call option is in the money when the asset price is greater than the exerciseprice. It is out of the money when the asset price is less than the exercise price; no onewould exercise the right to purchase for the strike price an asset worth less than thatprice. Conversely, put options are in the money when the exercise price exceeds theasset's value, because delivery of the lower-valued asset in exchange for the exerciseprice is profitable for the holder. Options are **at the money** when the exercise price andasset price are equal.
- The purchase price of the option is called the **premium.** It represents the compensationthe purchaser of the call must pay for the right to exercise the option if exercise becomesprofitable. If options are free,

$$Payoff\ to\ call\ holder = \begin{cases} S_T - X & if\ S_T > X \\ 0 & if\ S_T \leq X \end{cases}$$

$$Payoff\ to\ put\ holder = \begin{cases} 0 & if\ S_T \geq X \\ X - S_T & if\ S_T < X \end{cases}$$

Taking option premium into account,

$$Payoff\ to\ call\ holder = \begin{matrix} S_T - X - c & if\ S_T > X \\ -c & if\ S_T \le X \end{matrix}$$

$$Payoff\ to\ put\ holder = \begin{matrix} -p & if\ S_T \ge X \\ X - S_T - p & if\ S_T < X \end{matrix}$$

- Sellers of call options, who are said to *write* calls, receive premium income now as paymentagainst the possibility they will be required at some later date to deliver the asset inreturn for an exercise price less than the market value of the asset. If the option is left toexpire worthless, however, then the writer of the call clears a profit equal to the premiumincome derived from the initial sale of the option. But if the call is exercised, the profit tothe option writer is the premium income derived when the option was initially sold minus the difference between the value of the stock that must be delivered and the exercise pricethat is paid for those shares. If that difference is larger than the initial premium, the writerwill incur a loss. If options are free,

$$Payoff\ to\ call\ writer = \begin{matrix} -(S_T - X) & if\ S_T > X \\ 0 & if\ S_T \le X \end{matrix}$$

$$Payoff\ to\ put\ writer = \begin{matrix} 0 & if\ S_T \ge X \\ -(X - S_T) & if\ S_T < X \end{matrix}$$

Taking option premium into account

$$Payoff\ to\ call\ writer = \begin{matrix} c - (S_T - X) & if\ S_T > X \\ c & if\ S_T \le X \end{matrix}$$

$$Payoff\ to\ put\ writer = \begin{matrix} p & if\ S_T \ge X \\ p - (X - S_T) & if\ S_T < X \end{matrix}$$

Option Pricing - Black-Scholes Formula

Call Option

$$C = S_0 N(d_1) - Ke^{-rT} N(d_2)$$

where

$$d_1 = \frac{\log(S_0/K) + (r + \sigma^2/2)T}{\sigma\sqrt{T}}$$

$$d_2 = d_1 - \sigma\sqrt{T}$$

and

C_0: Current call option value.

S_0: Current stock price.

$N(d)$: The probability that a random draw from a standard normal distribution willbe less than d. This equals the area under the normal curve up to d.

X: Exercise price.

T: Time to expiration of option, in years.

σ: Standard deviation (volatility) of the stock return from now to expiration.

r: Risk-free interest rate (the annualized continuously compounded rate on a safeasset with the same maturity as the expiration date of the option, which may be different from r_f.).

e: The base of the natural log function, approximately 2.71828.
ln: Natural logarithm function.

Two Assumptions:

- Both the risk-free interest rate and the stock price volatility are constant over the life of the option.
- The distribution of the stock price at expiration progressively approaches the lognormal distribution.

Interpretation:

- $N(d)$: could be viewed as the risk-adjusted probabilities that the call option will expire in the money. If $N(d)$ terms are close to zero, meaning the option almost certainly will not be exercised.
- $\log(S_0/K)$: approximately the percentage amount by which the option is currently in or out of the money.
- $\sigma\sqrt{T}$: adjusts the amount by which the option is in the money or out of the money for the volatility of the stock price over the remaining life of the option.

Applications:

Delta Hedging and Greek Letters

- An option's hedge ratio (Option Delta) is the change in the price of an option for a $1 change in the stock price.

$$\Delta = \frac{\partial c}{\partial S}$$

- The hedge ratio for a call is $N(d_1)$. The hedge ratio for a put is $N(d_1) - 1$.
- Suppose the stock price is $100 per share and the option price is $10. Imagine that an investor has sold 20 call option contracts – that is, option on 2000 shares. If $\Delta = 0.6$, the investor's position could be hedged by

$$N + (-2000) \times 0.6 = 0 \quad \rightarrow N = 1200,$$

  which means to long 1200 shares of the stock. The gain (loss) position on the stock position would tend to offset the loss (gain) on the option position.
- A position with a delta of zero is referred to as *delta neutral*.
- Delta of a Portfolio $\frac{\partial \Pi}{\partial S} = \sum_{i=1}^{N} \mathrm{w_i}\Delta_i$

  Example: An option portfolio has the following three positions:

  1. A long position in 100,000 call options with strike price $55 and an expiration date in 3 months. The delta of each option is 0.533.
  2. A short position in 200,000 call options with strike price $56 and an expiration date in 5 months. The delta of each option is 0.48.

3. A short position in 50,000 put options with strike price \$56 and an expiration date in 2 months. The delta of each option is -0.508.

Delta of the whole portfolio is

$$100{,}000 \times 0.533 - 200{,}000 \times 0.48 - 50{,}000 \times (-0.508) = -14{,}900$$

The portfolio could be made delta neutral by buying 14900 shares.

- Theta: $\Theta = \frac{\partial \Pi}{\partial T}$, change in the value of the portfolio with respect to the passage of the time.
- Introduce Gamma: $\Gamma = \frac{\partial^2 c}{\partial S^2}$ Vega: $V = \frac{\partial c}{\partial \sigma}$ and Rho: $\rho = \frac{\partial c}{\partial r}$.

Example

	Delta	Gamma	Vega
Portfolio	0	-5000	-8000
Option 1	0.6	0.5	2.0
Option 2	0.5	0.8	1.2

To make the portfolio to be Gamma and Vega neutral, both options have to be used in way that

$$w_1 \cdot 0.5 + w_2 \cdot 0.8 - 5000 = 0$$
$$w_1 \cdot 2.0 + w_2 \cdot 1.2 - 8000 = 0$$

The solution is $w_1 = 400$ and $w_2 = 6000$, which means to make the portfolio Gamma and Vega neutral, we need 400 of Option 1 and 6000 of Option 2. The delta of the portfolio now is

$$w_1 \cdot 0.6 + w_2 \cdot 0.5 = 3240.$$

Therefore, 3240 units of the portfolio assets have to be sold to maintain the delta neutrality.

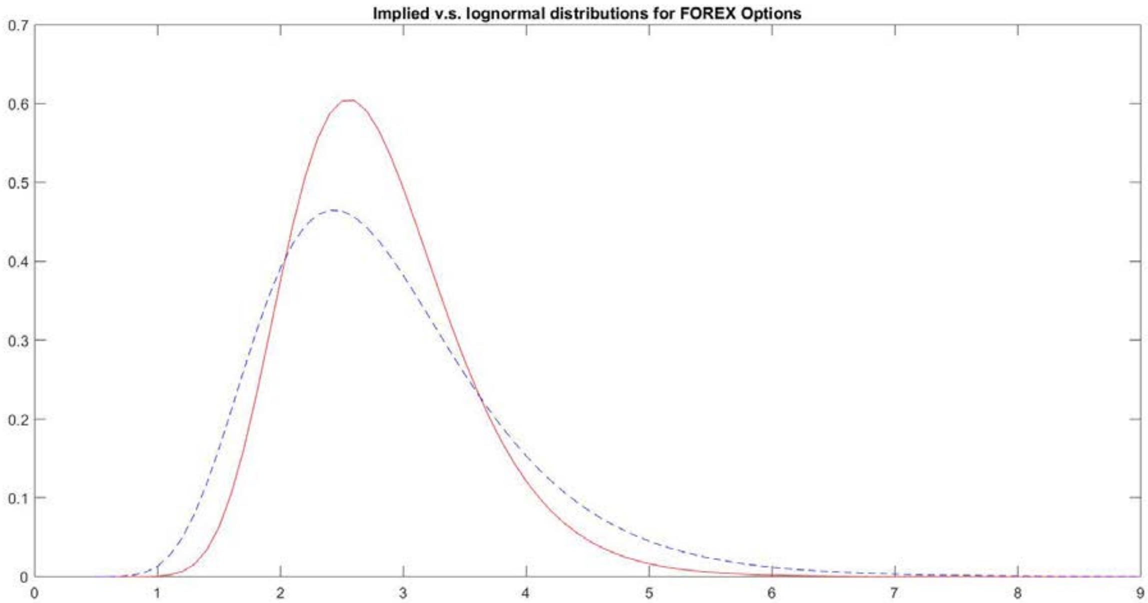
Implied Volatilities

- Given the option prices, one could back-engineer the volatility σ of the stock price from now to the maturity. This volatility is different from the standard deviation (physical) we calculated directly using the price sample. It contains market participants' perception about the future path of the asset price. It is risk-neutral.

- VIX is the implied volatility of S&P 500. It is called "Fear Gauge" indicator.

- A plot of the implied volatility of an option with a certain life as a function of its strike price is known as volatility smile. For example, for foreign currency options, the implied volatility is low for at-the-money options but moves progressively higher as the option moves inthemoney or out of the money.

- Using volatility smile, we could estimate the risk-neutral distribution of the underlying asset. It is also called implied distribution or state-price density.



In the Black-Scholes model, SPDs are log-normal distributions with constantvolatility. In practice, the volatility is time-varying and the return distribution deviatesfrom log-normality. The state price density is the second derivative (normalized to integrateto unity) of a call option pricing formula with respect to the strike price.

At the simplest level, one could form amoney spread with two call options, and assuming the probability that the stock pricefalls into the interval between the strike prices of these call options to be zero.

Consider theportfolio obtained by buying one call options struck at X_j and selling another call optionwith identical expiration date, but at higher exercise priceX_j.

$$\text{This portfolio pays } \begin{cases} 0 & if S_T \leq X_j \\ X_{j+1} - X_j & if S_T > X_{j+1} \end{cases}$$

Now denote by $H(S_t, X, T)$ the market price of acall option at time t with strike price X, time-to-maturity T, and the underlying asset priceS_t. The price of the money spread must be

$$H(S_t, X, T) - H(S_{t+1}, X, T) = 0 \times P\left(S_T < X_j\right) + (X_{j+1} - X_j) \times P(S_T > X_{j+1})$$

Using a fine grid of strike prices,$j = 1, \cdots, n$,

$$P\left(S_T \leq X_j\right) = 1 - P(S_T > X_{j+1})$$

yields thecumulative distribution function of the future stock price S_T. The state price density f^* isobtained by taking the derivative of the cumulative distribution function with respect tothe strike price.

Credit Risk Management

Probability of Default

Merton Model

One could estimate the default risk of a firm based on stock prices, because equity prices are available for a larger number of companies and are more actively traded than corporate bonds. The Merton (1974) model views equity as akin to a call option on the assets of the firm, with an exercise price given by the face value of debt.

To simplify to the extreme, consider a firm with total value V that has one bond due in one period with face value K. If the value of the firm exceeds the promised payment, the bond is repaid in full and stockholders receive the remainder. However, if V is less than K, the firm is in default and the bondholders receive V only. The value of equity goes to zero. Throughout, we assume that there are no transaction costs. Hence, the value of the stock at expiration is

$$S_T = Max(V_T - K, 0)$$

Because the bond and equity add up to the firm value, the value of the bond must be

$$B_T = V_T - S_T = V_T - Max(V_T - K, 0) = Min\ (V_T, K)$$

The current stock price, therefore, embodies a forecast of default probability, in thesame way that an option embodies a forecast of being exercised.

The value of the stock is given by the BS formula

$$S = call = VN(d_1) - Ke^{-rT}N(d_2)$$

In the Black-Scholes model, $N(d_2)$ is also the probability of exercising the call, or thatthe bond will not default. Conversely, $1 - N(d_2) = N(-d_2)$ is the risk-neutral probabilityof default.

Risk Modeling of Loss Given Default and Exposure at Default

Probability of Default (PD) modeling is supported by widely known methodologies used in Marketing, Account Management and Risk. LGD and EAD modeling are much less supported by business practices. As a result, modeling methodologies for LGD and EAD are still in the developmental stages.

Overview of LGD and EAD Modeling Methodologies

LGD Modeling Approaches

Loss given default (LGD) is defined as:

$$LGD = 1 - RecoveryRate$$

$$RecoveryRate = \frac{Amount\ \ Recovered}{Amount\ \ Outstanding\ \ at\ \ Default}$$

where amount recovered sums up all discounted cash flows received during the recovery process after default, less the total cost incurred.

There are major differences between PD and LGD modeling. While LGD is a continuous variable and usually follows a beta distribution, default events (PD) are binomial. LGD depends on the recovered amount, which may take several years after default to resolve, whereas PD describes the likelihood of a default event occurring within a specified period (usually 1 year). Information about events occurring after default has no effect on PD.

The non-normality of LGD (and EAD Factor) distribution calls for an explicit transformation so that the target variable follows a standard normal distribution. This will allow one to use a linear regression with a normally distributed target variable to get an LGD prediction.

Many lenders and business people believe that the relationship between collateral values and LGD is linear. However, very often data suggest that LGD distribution over collateral values is non-linear,

Parameter Estimation Procedure

- **Example: Define EAD-factor (Credit Conversion Factor)**
  Denote
  Bal_0- the facility outstanding dollar amount at current time
  Bal_1- the facility outstanding dollar amount at default time
  $Auth_0$- the facility authorized dollar amount at current time
  $Undrawn_0 = (Auth_0 - Bal_0)$- the facility undrawn dollar amount at current time

Define EAD factor as follows:

$$\text{EAD factor } = \begin{cases} \dfrac{\mathrm{Max}(\mathrm{Bal}_1 - \mathrm{Bal}_0, 0)}{\mathrm{Undrawn}} & \text{if undrawn}_0 > 0 \\ 0 & \text{if undrawn}_0 \leq 0 \end{cases}$$

Thus EAD Factor is the proportion of $undrawn_0$ to drawn down at default time. The predicted exposure amount at default is then calculated as

$$EAD = Bal_0 + EAD\ Factor\ \times Undrawn_0$$

in which we model the EAD Factor between -1 and 1 (floored at -1 and capped at 1), then floor the prediction at 0.

Another option in selecting a target variable in Exposure at Default modeling is to model the facility utilization change, which is defined as:

$$Util_{Ch} = \frac{Bal_1 - Bal_0}{Auth_0}$$

floored at 0 and capped at 1.

- **Weight of Evidence (WOE)Transformation**.

The first step for LGD/EAD modeling is the technique of transformation for independent variables. It applies to both types of variables; numeric and character. As will be shown later, such variable transformation allows one to tackle problems with optimum selection of variables, issues with outliers, as well as problems with imputation of missing values.

Consider the linear regression model

$$y \sim a + bx$$

where x is a vector of independent variables, a is the intercept, and b is the vector of parameters.

The WOE transformation for $z = (x, y)$ consists of the following steps:
(a) Partition the variable z into intervals.
(b) Calculate the average of z each interval
(c) Derive a new variable $w(z)$ by assigning the average from (b) when the value of variable z falls in that interval.
We call the derived variable $w(z)$ the WOE transformation for the variable z.

For instance, one could use $k - Means$ Clusteringfor WOE Transformation:For a fixed $k > 1$, the $k - Means$ Clustering estimates $E(y|x)$ by taking average of $y_1, \cdots y_T$ around $k$class level (means). Then grouping based on business and statistical considerations.

- **Model Factor Selection**.

Naïve Bayesian

$$\ln \mathrm{BIC} = \left(\frac{\mathrm{k}}{\mathrm{T}}\right) \ln \mathrm{T} + \ln(\frac{\mathrm{RSS}}{\mathrm{T}})$$

$$RSS = \sum_{t=1}^{T} \left(Y_t - \beta_0 - \sum_{j=1}^{k} x_{tj}\, \beta_j \right)^2$$

Ridge Regression

$$\hat{\beta}^{ridge} = \underbrace{argmin}_{\beta} \{ \sum_{t=1}^{T} \left(Y_t - \beta_0 - \sum_{j=1}^{k} x_{tj}\, \beta_j \right)^2 + \lambda \sum_{j=1}^{k} \beta_j^2 \}$$

$$\hat{\beta}^{ridge} = (X'X + \lambda I)^{-1} X'Y$$

LASSO

$$\hat{\beta}^{LASSO} = \underbrace{argmin}_{\beta} \{ \frac{1}{2} \sum_{t=1}^{T} \left(Y_t - \beta_0 - \sum_{j=1}^{k} X_{tj}\, \beta_j \right)^2 + \lambda \sum_{j=1}^{k} |\boldsymbol{\beta_j}| \}$$

Elastic Net

$$\hat{\beta}^{Elastic\ Net} = \underbrace{argmin}_{\beta}\{\sum_{t=1}^{T}\left(Y_t - \beta_0 - \sum_{j=1}^{k} X_{tj}\,\beta_j\right)^2 + \lambda\sum_{j=1}^{k}(\alpha\beta_j^2 + (1-\alpha)|\beta_j|)\}$$

- **Choose a Model Structure**.

We consider the following models, which are either trained by maximizing likelihood or minimizing the least square error:

Logit Model:Let $P_i = \frac{e^{z_i}}{1+e^{z_i}}$ (where $Z_i = \beta_1 + \beta_2 X_i$) be the probability of default, then $1 - P_i$ the probability of not to default

$$\mathrm{L_i} = \ln\left(\frac{P_i}{1-P_i}\right) = \beta_0 + \beta_1 X + \epsilon_i$$

L ,the log of odds ratio, is not only linear in X, but also linear in the parameters.L is called the logit.

Mixture Model:

$$Y_t = \beta_0 + \beta_1 h_t + \epsilon_t$$
$$h_t = \mu + \phi\, h_{t-1} + \eta_t$$
$$\eta_t = \rho\epsilon_t + \sqrt{1-\rho^2}u_t \quad and \quad u_t \sim N(0,1)$$

ϵ_t follows a time-varying distribution $f_t(\cdot)$. This could be approximated by a multivariate Gaussian finite mixture

$$f_t(x,p,\mu,\Sigma) = \sum_{k=1}^{c} p_k\phi(x,\mu_k,\Sigma_k)$$

where ϕ represents a multivariate normal probability density function given by

$$\phi(x_i;\mu_k,\Sigma_k) = \frac{\exp\{-\frac{1}{2}(x-x_i)^T\Sigma_k^{-1}(x_i-\mu_k)\}}{(2\pi)^{d/2}\sqrt{|\Sigma_k|}}$$

$\mu_k$is a d-dimensional vector of means, and$\Sigma_k$is a $d \times d$covariance matrix.

Single Layer Neural Network:

Single Layer Feed-forward Network Model

$$r_t = \alpha + \sum_{i=1}^{p}\beta_i r_{t-i} + \sum_{j=1}^{d}\eta_j\, G(\alpha_j + \sum_{i=1}^{p}\gamma_i s_{t-i}^{n_1,n_2}) + \epsilon_t\ \epsilon_t \sim ID(0,\sigma_t^2)$$

where G is the **activation function** which is chosen to be a sigmoidal function:

$$G(x) = \frac{1}{1 + e^{-\alpha x}}$$

Single Layer Feed-forward Network Model with lagged returns alone:

$$r_t = \alpha + \sum_{i=1}^{p} \beta_i r_{t-i} + \sum_{j=1}^{d} \beta_j G(\alpha_j + \sum_{i=1}^{p} \gamma_i r_{t-i}) + \epsilon_t \; \epsilon_t \sim ID(0, \sigma_t^2)$$

d is the total types of signals you want to enclose in the prediction.
p is the total numbers of lags you choose to enclose in the information set of prediction.

As both LGD and EAD Factor distributions usually exhibit high concentration around 0 and 1, the mixture model or single layer neural network demonstrate significant improvement over the logit model that uses only raw variables (with no WOE transformation).

- **Boost Modeling**

When we are not satisfied with the model built, probably because of its bias in prediction for some score bands. In addition to the segmentation methodology, we can use the following boost strategies to improve the accuracy:
(a) boosting the model by a decision tree
(b) boosting the model by a scalar model
(c) boosting the model by a linear regression
With methodology (a), we simply choose the model prediction error as the new dependent variable and train a decision tree using all available variables (no WOE transformation is required), including the built model score. In general, we get a decent improvement in accuracy. Specifically,

Boosting by a Scalar Model

Let $ptrend$ denote the value of y predicted by the model built (base model). A scalar model can be trained by scaling an exponential factor to the base model:

$$y \sim \exp(a_0 + a_1 z_1 + a_2 z_2 + \cdots + a_{k+1} z_k) \times ptrend$$

where $z_1, z_2, \cdots, z_k$ are either indicator variables denoting the score bands that require adjustment for the prediction error, or other available that give lift to this scalar model. The resulting boosty model needs to be capped at 1 (because 0 < y <1).

Scalar boost model can be implemented and trained similarly by using the conjugate gradient algorithm and the Fletcher-Reeves method.

Boosting by a Linear Regression

We can divide the sample into high/low segments using indicators H and L. where:

$$H = \begin{cases} 1 & ptrend > c \\ 0 & else \end{cases}$$

$$L = 1 - H$$

Where number c can be chosen to be the mean or mode of variable y. Let p_1 and p_2 be the sub-model scores for H and L segments respectively, both capped at 1. Then the final model can be given by:

$$p = H \times p_1 + L \times p_2$$

To generate new predictive variables, we consider variables of the form:

$$ptrend \times w(z)$$

where $w(z)$ is the WOE transformed variable for variable z.