Risk Simulator

ROV BizStats

ROV ANALYTICS: QUICK OVERVIEW

- **ANOVA: One-Way Single Factor with Multiple Treatments.** An extension of the two-variable t-test, looking at multiple variables simultaneously and when the sampling distribution is assumed to be approximately normal. A two-tailed hypothesis tests the null hypothesis such that the population means of each treatment is statistically identical to the rest of the group, indicating that there is no effect among the different treatment groups.
- **ANOVA: One-Way Randomized Block.** The sampling distribution is assumed to be approximately normal and when there exists a Block variable for which ANOVA will Control (i.e., Block the effects of this variable by controlling it in the experiment). This analysis can test for the effects of both the Treatments as well as the effectiveness of the Control or Block variable. If the calculated p-value for the Treatment or Block is less than or equal to the significance level used in the test, then reject the null hypothesis and conclude that there is a significant difference among the different Treatments or Blocks.
- **ANOVA: Two-Way.** An extension of the Single Factor and Randomized Block ANOVA by simultaneously examining the effects of two factors on the dependent variable, along with the effects of interactions between the different levels of these two factors. Unlike the randomized block design, this model examines the interactions between different levels of the factors or independent variables. In a two-factor experiment, interaction exists when the effect of a level for one factor depends on which level of the other factor is present. There are three sets of null and alternate hypotheses to be tested.
- **ARIMA.** Autoregressive Integrated Moving Average is used for forecasting time-series data using its own historical data by itself or with exogenous/other variables. The first segment is the autoregressive (AR) term corresponding to the number of lagged value of the residual in the unconditional forecast model. In essence, the model captures the historical variation of actual data to a forecasting model and uses this variation or residual to create a better predicting model. The second segment is the integration order (I) term corresponding to the number of differencing the time series to be forecasted goes through to make the data stationary. This element accounts for any nonlinear growth rates existing in the data. The third segment is the moving average (MA) term, which is essentially the moving average of lagged forecast errors. By incorporating this lagged forecast errors term, the model in essence learns from its forecast errors or mistakes and corrects for them through a moving average calculation. The ARIMA model follows the Box-Jenkins methodology with each term representing steps taken in the model construction until only random noise remains.
- **ARIMA (Auto).** Runs some common combinations of ARIMA models (low-order PDQ) and returns the best models.
- **Autocorrelation and Partial Autocorrelation.** One very simple approach to test for autocorrelation is to graph the time series of a regression equation's residuals. If these residuals exhibit some cyclicality, then autocorrelation exists. Another more robust approach to detect autocorrelation is the use of the Durbin-Watson statistic, which estimates the potential for a first-order autocorrelation. The Durbin-Watson test employed also identifies model misspecification, that is, if a particular time-series variable is correlated to itself one period prior. Many time-series data tend to be autocorrelated to their historical occurrences. This relationship can exist for multiple reasons, including the variables' spatial relationships (similar time and space), prolonged economic shocks and events, psychological inertia, smoothing, seasonal adjustments of the data, and so forth.
- **Autoeconometrics.** Runs some common combinations of Basic Econometrics and returns the best models.
- **Basic Econometrics/Custom Econometrics.** Applicable for forecasting time-series and cross-sectional data and for modeling relationships among variables, and allows you to create custom multiple regression models. Econometrics refers to a branch of business analytics, modeling, and forecasting techniques for modeling the behavior or forecasting certain business, financial, economic, physical science, and other variables. Running the Basic Econometrics models is similar to regular regression analysis except that the dependent and independent variables are allowed to be modified before a regression is run.
- **Charts: Pareto.** A Pareto chart contains both a bar chart and a line graph. Individual values are represented in descending order by the bars and the cumulative total is represented by the ascending line. Also known as the "80-20" chart, whereby you see that by focusing on the top few variables, we are already accounting for more than 80% of the cumulative effects of the total.
- **Charts: Box-Whisker.** Box plots or box-and-whisker plots graphically depict numerical data using their descriptive statistics: the smallest observation (Minimum), First Quartile or 25th Percentile (Q1), Median or Second Quartile or 50th Percentile (Q2), Third Quartile (Q3), and largest observation (Maximum). A box plot may also indicate which observations, if any, might be considered outliers.
- **Charts: Q-Q Normal.** This Quantile-Quantile chart is a normal probability plot, which is a graphical method for comparing a probability distribution with the normal distribution by plotting their quantiles against each other.
- **Combinatorial Fuzzy Logic.** Applies fuzzy logic algorithms for forecasting time-series data by combining forecast methods to create an optimized model. Fuzzy logic is a probabilistic logic dealing with reasoning that is approximate rather than fixed and exact where fuzzy logic variables may have a truth value that ranges in degree between 0 and 1.
- **Control Charts: C, NP, P, R, U, XMR.** Sometimes specification limits of a process are not set; instead, statistical control limits are computed based on the actual data collected (e.g., the number of defects in a manufacturing line). The upper control limit (UCL) and lower control limit (LCL) are computed, as are the central line (CL) and other sigma levels. The resulting chart is called a control chart, and if the process is out of control, the actual defect line will be outside of the UCL and LCL lines for a certain number of times.
    - **C Chart:** variable is an *attribute* (e.g., defective or nondefective), the data collected are in total *number of defects* (actual count in units), and there are *multiple measurements* in a sample experiment; when multiple experiments are run and the average number of defects of the collected data is of interest; and *constant number of samples* collected in each experiment.
    - **NP Chart:** variable is an *attribute* (e.g., defective or nondefective), the data collected are in *proportions of defects* (or number of defects in a specific sample), and there are *multiple measurements* in a sample experiment; when multiple experiments are run and the average proportion of defects of the collected data is of interest; and *constant number of samples* collected in each experiment.
    - **P Chart:** variable is an *attribute* (e.g., defective or nondefective), the data collected are in *proportions of defects* (or number of defects in a specific sample), and there are *multiple measurements* in a sample experiment; when multiple experiments are run and the average proportion of defects of the collected data is of interest; and with *different number of samples* collected in each experiment.
    - **R Chart:** variable has *raw data* values, there are *multiple measurements* in a sample experiment, *multiple experiments* are run, and the *range* of the collected data is of interest.
    - **U Chart:** variable is an *attribute* (e.g., defective or nondefective), the data collected are in total *number of defects* (actual count in units), and there are *multiple measurements* in a sample experiment; when multiple experiments are run and the average number of defects of the collected data is of interest; and with *different number of samples* collected in each experiment.
    - **XmR Chart:** *raw data* values, *single measurement* taken in each sample experiment, *multiple experiments* are run, and the *actual value* of the collected data is of interest.
- **Correlation (Linear and Nonlinear).** Computes the Pearson's linear product moment correlations (commonly referred to as the Pearson's R) as well as the nonlinear Spearman rank-based correlation between variable pairs and returns them as a correlation matrix. The correlation coefficient ranges between –1.0 and +1.0, inclusive. The sign indicates the direction of association between the variables while the coefficient indicates the magnitude or strength of association.
- **Cubic Spline.** Interpolates missing values of a time-series dataset and extrapolates values of future forecast periods using nonlinear curves. Spline curves can also be used to forecast or extrapolate values of future time periods beyond the time period of available data and the data can be linear or nonlinear.
- **Descriptive Statistics.** Almost all distributions can be described within four moments (some distributions require one moment, while others require two moments, and so forth). This tool computes the four moments and associated descriptive statistics.
- **Deseasonalizing.** This model deseasonalizes and detrends your original data to take out any seasonal and trending components. In forecasting models, the process eliminates the effects of accumulating datasets from seasonality and trend to show only the absolute changes in values and to allow potential cyclical

patterns to be identified by removing the general drift, tendency, twists, bends, and effects of seasonal cycles of a set of time-series data.

- **Distributional Fitting.** Which distribution does an analyst or engineer use for a particular input variable in a model? What are the relevant distributional parameters? The null hypothesis tested is that the fitted distribution is the same distribution as the population from which the sample data to be fitted comes.
    - **Akaike Information Criterion (AIC).** Rewards goodness of fit but also includes a penalty that is an increasing function of the number of estimated parameters (although AIC penalizes the number of parameters less strongly than other methods).
    - **Anderson-Darling (AD).** When applied to testing if a normal distribution adequately describes a set of data, it is one of the most powerful statistical tools for detecting departures from normality, and is powerful for testing normal tails. However, in non-normal distributions, this test lacks power compared to others.
    - **Kolmogorov-Smirnov (KS).** A nonparametric test for the equality of continuous probability distributions that can be used to compare a sample with a reference probability distribution, making it useful for testing abnormally shaped distributions and non-normal distributions.
    - **Kuiper's Statistic (K).** Related to the KS test making it as sensitive in the tails as at the median and also makes it invariant under cyclic transformations of the independent variable, making it invaluable when testing for cyclic variations over time. The AD provides equal sensitivity at the tails as the median, but it does not provide the cyclic invariance.
    - **Schwarz/Bayes Information Criterion (SC/BIC).** The SC/BIC introduces a penalty term for the number of parameters in the model with a larger penalty than AIC.
- **Exponential J Curve.** This function models exponential growth where value of the next period depends on the current period's level and the increase is exponential. Over time, the values will increase significantly from one period to another. This model is typically used in forecasting biological growth and chemical reactions over time.
- **Heteroskedasticity.** Several tests exist to test for the presence of heteroskedasticity, i.e., where the volatilities or uncertainties (standard deviation or variance of a variable is non-constant over time). These tests also are applicable for testing misspecifications and nonlinearities. The test is based on the null hypothesis of no heteroskedasticity.
- **Generalized Linear Models/Limited Dependent Variables: Logit.** Limited dependent variables techniques are used to forecast the probability of something occurring given some independent variables (e.g., predicting if a credit line will default given the obligor's characteristics such as age, salary, credit card debt levels, or the probability a patient will have lung cancer based on age and number of cigarettes smoked monthly, and so forth). The dependent variable is limited (i.e., binary 1 and 0 for default/cancer, or limited to integer values 1, 2, 3, etc.). Traditional regression analysis will not work as the predicted probability is usually less than zero or greater than one, and many of the required regression assumptions are violated (e.g., independence and normality of the errors). We also have a vector of independent variable regressors, X, which are assumed to influence the outcome, Y. A typical ordinary least squares regression approach is invalid because the regression errors are heteroskedastic and non-normal, and the resulting estimated probability estimates will return nonsensical values of above 1 or below 0. This analysis handles these problems using an iterative optimization routine to maximize a log likelihood function when the dependent variables are limited.
- **Generalized Linear Models/Limited Dependent Variables: Probit.** A probit model (sometimes also known as a *normit* model) is a popular alternative specification for a binary response model. It employs a probit function estimated using maximum likelihood estimation and is called probit regression. The probit and logistic regression models tend to produce very similar predictions where the parameter estimates in a logistic regression tend to be 1.6 to 1.8 times higher than they are in a corresponding probit model. The choice of using a probit or logit is entirely up to convenience, and the main distinction is that the logistic distribution has a higher kurtosis (fatter tails) to account for extreme values. For example, suppose that house ownership is the decision to be modeled, and this response variable is binary (home purchase or no home purchase) and depends on a series of independent variables X_i such as income, age, and so forth, such that $I_i = \beta_0 + \beta_1 X_1 + ... + \beta_n X_n$, where the larger the value of I_i, the higher the probability of home ownership. For each family, a critical I^* threshold exists, where if exceeded, the house is purchased, otherwise, no home is purchased, and the outcome probability (P) is assumed to be normally distributed, such that $P_i = CDF(I)$ using a standard normal cumulative distribution function (CDF). Therefore, use the estimated coefficients exactly like that of a regression model and, using the estimated Y, apply a standard normal distribution to compute the probability.
- **Generalized Linear Models/Limited Dependent Variables: Tobit.** The tobit model (censored tobit) is an econometric and biometric modeling method used to describe the relationship between a non-negative dependent variable Y_i and one or more independent variables X_i. A tobit model is an econometric model in which the dependent variable is censored; that is, the dependent variable is censored because values below zero are not observed. The tobit model assumes that there is a latent unobservable variable Y^*. This variable is linearly dependent on the X_i variables via a vector of β_i coefficients that determine their interrelationships. In addition, there is a normally distributed error term U_i to capture random influences on this relationship. The observable variable Y_i is defined to be equal to the latent variables whenever the latent variables are above zero, and Y_i is assumed to be zero otherwise. That is, $Y_i = Y^*$ if $Y^* > 0$ and $Y_i = 0$ if $Y^* = 0$. If the relationship parameter β_i is estimated by using ordinary least squares regression of the observed Y_i on X_i, the resulting regression estimators are inconsistent and yield downward-biased slope coefficients and an upward-biased intercept.
- **Linear Interpolation.** Sometimes interest rates or any type of time-dependent rates may have missing values. For instance, the Treasury rates for Years 1, 2, and 3 exist, and then jump to Year 5, skipping Year 4. We can, using linear interpolation (i.e., we assume the rates during the missing periods are linearly related), determine and "fill in" or interpolate their values.
- **Logistic S Curve.** The S curve, or logistic growth curve, starts off like a J curve, with exponential growth rates. Over time, the environment becomes saturated (e.g., market saturation, competition, overcrowding), the growth slows, and the forecast value eventually ends up at a saturation or maximum level. The S curve model is typically used in forecasting market share or sales growth of a new product from market introduction until maturity and decline, population dynamics, growth of bacterial cultures, and other naturally occurring variables.
- **Markov Chain.** Models the probability of a future state that depends on a previous state (a mathematical system that undergoes transitions from one state to another), forming a chain when linked together (a random process characterized as memoryless: the next state depends only on the current state and not on the sequence of events that preceded it) that reverts to a long-run steady state level. Used to forecast the market share of two competitors.
- **Multiple Regression (Linear and Nonlinear).** Multivariate regression is used to model the relationship structure and characteristics of a certain dependent variable as it depends on other independent exogenous variables. Using the modeled relationship, we can forecast the future values of the dependent variable. The accuracy and goodness of fit for this model can also be determined. Linear and nonlinear models can be fitted in the multiple regression analysis.
- **Neural Network.** Commonly used to refer to a network or circuit of biological neurons, modern usage of the term often refers to artificial neural networks comprising artificial neurons, or nodes, recreated in a software environment. Such networks attempt to mimic the neurons in the human brain in ways of thinking and identifying patterns and, in our situation, identifying patterns for the purposes of forecasting time-series data.
    - **Linear.** Applies a linear function.
    - **Nonlinear Logistic.** Applies a nonlinear logistic function.
    - **Nonlinear Cosine-Hyper Tangent.** Applies a nonlinear cosine with hyperbolic tangent function.
    - **Nonlinear Hyper Tangent.** Applies a nonlinear hyperbolic tangent function.

****

Nonparametric Hypothesis Tests

Nonparametric techniques make no assumptions about the specific shape or distribution from which the sample is drawn. This lack of assumptions is different from the other hypotheses tests such as ANOVA or t-tests (parametric tests) where the sample is assumed to be drawn from a population that is normally or approximately normally distributed. If normality is assumed, the power of the test is higher due this normality restriction. However, if flexibility on distributional requirements is needed, then nonparametric techniques are superior. In general, nonparametric methodologies provide the following advantages over other parametric tests:

- Normality or approximate normality does not have to be assumed.
- Fewer assumptions about the population are required; that is, nonparametric tests do not require that the population assume any specific distribution.
- Smaller sample sizes can be analyzed.

- Samples with nominal and ordinal scales of measurement can be tested.
- Sample variances do not have to be equal, which is required in parametric tests.

However, two caveats are worthy of mention:

- Compared to parametric tests, nonparametric tests use data less efficiently.
- The power of the test is lower than that of the parametric tests.

- **Chi-Square Goodness of Fit.** The chi-square test for goodness of fit is used to examine if a sample dataset could have been drawn from a population having a specified probability distribution. The probability distribution tested here is the normal distribution. The null hypothesis tested is such that the sample is randomly drawn from the normal distribution.
- **Chi-Square Independence.** The chi-square test for independence examines two variables to see if there is some statistical relationship between them. This test is not used to find the exact nature of the relationship between the two variables, but to simply test if the variables could be independent of each other. The null hypothesis tested is such that the variables are independent of each other.
- **Chi-Square Population Variance.** The chi-square test for population variance is used for hypothesis testing and confidence interval estimation for a population variance. The population variance of a sample is typically unknown, and hence the need for quantifying this confidence interval. The population is assumed to be normally distributed.
- **Friedman's Test.** The Friedman test is the extension of the Wilcoxon Signed-Rank test for paired samples. The corresponding parametric test is the Randomized Block Multiple Treatment ANOVA, but unlike the ANOVA, the Friedman test does not require that the dataset be randomly sampled from normally distributed populations with equal variances. The Friedman test uses a two-tailed hypothesis test where the null hypothesis is such that the population medians of each treatment are statistically identical to the rest of the group, that is, there is no effect among the different treatment groups.
- **Kruskal-Wallis Test.** The Kruskal-Wallis test is the extension of the Wilcoxon Signed-Rank test by comparing more than two independent samples. The corresponding parametric test is the One-Way ANOVA, but unlike the ANOVA, the Kruskal-Wallis does not require that the dataset be randomly sampled from normally distributed populations with equal variances. The Kruskal-Wallis test is a two-tailed hypothesis test where the null hypothesis is such that the population medians of each treatment are statistically identical to the rest of the group, that is, there is no effect among the different treatment groups.
- **Lilliefors Test.** The Lilliefors test evaluates the null hypothesis of whether the data sample was drawn from a normally distributed population, versus an alternate hypothesis that the data sample is not normally distributed. If the calculated p-value is less than or equal to the alpha significance value, then reject the null hypothesis and accept the alternate hypothesis. Otherwise, if the p-value is higher than the alpha significance value, do not reject the null hypothesis. This test relies on two cumulative frequencies: one derived from the sample dataset and one from a theoretical distribution based on the mean and standard deviation of the sample data. An alternative to this test is the chi-square test for normality. The chi-square test requires more data points to run compared to the Lilliefors test.
- **Runs Test.** The Runs test evaluates the randomness of a series of observations by analyzing the number of runs it contains. A run is a consecutive appearance of one or more observations that are similar. The null hypothesis tested is whether the data sequence is random, versus the alternate hypothesis that the data sequence is not random.
- **Wilcoxon Signed-Rank (One Var).** The single-variable Wilcoxon Signed-Rank test looks at whether a sample dataset could have been randomly drawn from a particular population whose median is being hypothesized. The corresponding parametric test is the one-sample t-test, which should be used if the underlying population is assumed to be normal, providing a higher power on the test.
- **Wilcoxon Signed-Rank (Two Var).** The Wilcoxon Signed-Rank test for paired looks at whether the median of the differences between the two paired variables are equal. This test is specifically formulated for testing the same or similar samples before and after an event (e.g., measurements taken before a medical treatment are compared against those measurements taken after the treatment to see if there is a difference). The corresponding parametric test is the two-sample t-test with dependent means, which should be used if the underlying population is assumed to be normal, providing a higher power on the test.

****

Parametric Hypothesis Tests

- **One Variable (T).** The one-variable t-test of means is appropriate when the population standard deviation is not known but the sampling distribution is assumed to be approximately normal (the t-test is used when the sample size is less than 30). This t-test can be applied to three types of hypothesis tests—a two-tailed test, a right-tailed test, and a left-tailed test—to examine if the population mean is equal to, less than, or greater than the hypothesized mean based on the sample dataset.
- **One Variable (Z).** The one-variable Z-test is appropriate when the population standard deviation is known and the sampling distribution is assumed to be approximately normal (this applies when the number of data points exceeds 30).
- **One-Variable (Z) Proportion.** The one-variable Z-test for proportions is appropriate when the sampling distribution is assumed to be approximately normal (this applies when the number of data points exceeds 30, and when the number of data points, N, multiplied by the hypothesized population proportion mean, P, is greater than or equal to 5, $NP \geq 5$). The data used in the analysis have to be proportions and be between 0 and 1.
- **Two-Variable (T) Dependent.** The two-variable dependent t-test is appropriate when the population standard deviation is not known but the sampling distribution is assumed to be approximately normal (the t-test is used when the sample size is less than 30). In addition, this test is specifically formulated for testing the same or similar samples before and after an event (e.g., measurements taken before a medical treatment are compared against those measurements taken after the treatment to see if there is a difference).
- **Two-Variable (T) Independent Equal Variance.** The two-variable t-test with equal variances is appropriate when the population standard deviation is not known but the sampling distribution is assumed to be approximately normal (the t-test is used when the sample size is less than 30). In addition, the two independent samples are assumed to have similar variances.
- **Two-Variable (T) Independent Unequal Variance.** The two-variable t-test with unequal variances (the population variance of sample 1 is expected to be different from the population variance of sample 2) is appropriate when the population standard deviation is not known but the sampling distribution is assumed to be approximately normal (the t-test is used when the sample size is less than 30). In addition, the two independent samples are assumed to have similar variances.
- **Two-Variable (Z) Independent Means**. The two-variable Z-test is appropriate when the population standard deviations are known for the two samples, and the sampling distribution of each variable is assumed to be approximately normal (this applies when the number of data points of each variable exceeds 30).
- **Two-Variable (Z) Independent Proportions.** The two-variable Z-test on proportions is appropriate when the sampling distribution is assumed to be approximately normal (this applies when the number of data points of both samples exceeds 30). Further, the data should all be proportions and be between 0 and 1.
- **Two-Variable (F) Variances.** The two-variable F-test analyzes the variances from two samples (the population variance of Sample 1 is tested with the population variance of Sample 2 to see if they are equal) and is appropriate when the population standard deviation is not known but the sampling distribution is assumed to be approximately normal.

****

- **Principal Component Analysis.** Principal component analysis, or PCA, makes multivariate data easier to model and summarize. To understand PCA, suppose we start with N variables that are unlikely to be independent of one another, such that changing the value of one variable will change another variable. PCA modeling will replace the original N variables with a new set of M variables that are less than N but are uncorrelated to one another, while at the same time, each of these M variables is a linear combination of the original N variables, so that most of the variation can be accounted for just using fewer explanatory variables.
- **Seasonality.** Many time-series data exhibit seasonality where certain events repeat themselves after some time period or seasonality period (e.g., ski resorts' revenues are higher in winter than in summer, and this predictable cycle will repeat itself every winter).
- **Segmentation Clustering.** Taking the original dataset, we run some internal algorithms (a combination or k-means hierarchical clustering and other method of moments in order to find the best-fitting groups or natural statistical clusters) to statistically divide or segment the original dataset into multiple groups.
- **Stepwise Regression (Backward).** In the backward method, we run a regression with Y on all X variables and reviewing each variable's p-value, systematically eliminate the variable with the largest p-value. Then run a regression again, repeating each time until all p-values are statistically significant.

- **Stepwise Regression (Correlation).** In the correlation method, the dependent variable Y is correlated to all the independent variables X, and starting with the X variable with the highest absolute correlation value, a regression is run. Then subsequent X variables are added until the p-values indicate that the new X variable is no longer statistically significant. This approach is quick and simple but does not account for interactions among variables, and an X variable, when added, will statistically overshadow other variables.
- **Stepwise Regression (Forward).** In the forward method, we first correlate Y with all X variables, run a regression for Y on the highest absolute value correlation of X, and obtain the fitting errors. Then, correlate these errors with the remaining X variables and choose the highest absolute value correlation among this remaining set and run another regression. Repeat the process until the p-value for the latest X variable coefficient is no longer statistically significant and then stop the process.
- **Stepwise Regression (Forward-Backward).** In the forward and backward method, apply the forward method to obtain three X variables, and then apply the backward approach to see if one of them needs to be eliminated because it is statistically insignificant. Repeat the forward method and then the backward method until all remaining X variables are considered.

****

Stochastic Processes

Sometimes variables cannot be readily predicted using traditional means, and these variables are said to be stochastic. Nonetheless, most financial, economic, and naturally occurring phenomena (e.g., motion of molecules through the air) follow a known mathematical law or relationship. Although the resulting values are uncertain, the underlying mathematical structure is known and can be simulated using Monte Carlo risk simulation.

- **Brownian Motion Random Walk Process.** The Brownian motion random walk process takes the form of $\frac{\delta S}{S} = \mu(\delta t) + \sigma\varepsilon\sqrt{\delta t}$ for regular options simulation, or a more generic version takes the form of $\frac{\delta S}{S} = (\mu - \sigma^2/2)\delta t + \sigma\varepsilon\sqrt{\delta t}$ for a geometric process. For an exponential version, we simply take the exponentials, and as an example, we have $\frac{\delta S}{S} = \exp\left[\mu(\delta t) + \sigma\varepsilon\sqrt{\delta t}\right]$, where we define S as the variable's previous value, δS as the change in the variable's value from one step to the next, μ as the annualized growth or drift rate, and σ as the annualized volatility
- **Mean-Reversion Process.** The following describes the mathematical structure of a mean-reverting process with drift: $\frac{\delta S}{S} = \eta(\bar{S}e^{\mu(\delta t)} - S)\delta t + \mu(\delta t) + \sigma\varepsilon\sqrt{\delta t}$. Here we define η as the rate of reversion to the mean, $\bar{S}$ as the long-term value the process reverts to Y as the historical data series, β_0 as the intercept coefficient in a regression analysis, and β_1 as the slope coefficient in a regression analysis.
- **Jump-Diffusion Process.** A jump-diffusion process is similar to a random walk process but includes a probability of a jump at any point in time. The occurrences of such jumps are completely random but their probability and magnitude are governed by the process itself. We have the structure $\frac{\delta S}{S} = \eta(\bar{S}e^{\mu(\delta t)} - S)\delta t + \mu(\delta t) + \sigma\varepsilon\sqrt{\delta t} + \theta F(\lambda)(\delta t)$ for a jump diffusion process and we define θ as the jump size of S, $F(\lambda)$ as the inverse of the Poisson cumulative probability distribution, and λ as the jump rate of S.
- **Jump-Diffusion Process with Mean Reversion.** This model is essentially a combination of all three models discussed above (geometric Brownian motion with mean-reversion process and a jump-diffusion process).

****

- **Structural Break.** Tests if the coefficients in different datasets are equal, and is most commonly used in time-series analysis to test for the presence of a structural break. A time-series dataset can be divided into two subsets and each subset is tested on the other and on the entire dataset to statistically determine if, indeed, there is a break starting at a particular time period. A one-tailed hypothesis test is performed on the null hypothesis such that the two data subsets are statistically similar to one another, that is, there is no statistically significant structural break.
- **Time-Series Analysis.** In well-behaved time-series data (e.g., sales revenues and cost structures of large corporations), the values tend to have up to three elements: a base value, trend, and seasonality. Time-series analysis uses these historical data and decomposes them into these three elements, and recomposes them into future forecasts. In other words, this forecasting method, like some of the others described, first performs a back-fitting (backcast) of historical data before it provides estimates of future values (forecasts).
- **Time-Series Analysis (Auto).** Selecting this automatic approach will allow the user to initiate an automated process in methodically selecting the best input parameters in each model and ranking the forecast models from best to worst by looking at their goodness-of-fit results and error measurements.
    - **Time-Series Analysis (DES).** The double exponential-smoothing (DES) approach is used when the data exhibit a trend but no seasonality.
    - **Time-Series Analysis (DMA).** The double moving average (DMA) method is used when the data exhibit a trend but no seasonality.
    - **Time-Series Analysis (HWA).** The Holt-Winters Additive (HWA) approach is used when the data exhibit both seasonality and trend.
    - **Time-Series Analysis (HWM).** The Holt-Winters Multiplicative (HWM) approach is used when the data exhibit both seasonality and trend.
    - **Time-Series Analysis (SA).** The Seasonal Additive (SA) approach is used when the data exhibit seasonality but no trend.
    - **Time-Series Analysis (SM).** The Seasonal Multiplicative (SM) approach is used when the data exhibit seasonality but no trend.
    - **Time-Series Analysis (SES).** The Single Exponential Smoothing (SES) approach is used when the data exhibit no trend and no seasonality.
    - **Time-Series Analysis (SMA).** The Single Moving Average (SMA) approach is used when the data exhibit no trend and no seasonality.
- **Trending and Detrending: Difference, Exponential, Linear, Logarithmic, Moving Average, Polynomial, Power, Rate, Static Mean, and Static Median.** Detrends your original data to take out any trending components. In forecasting models, the process removes the effects of accumulating datasets from seasonality and trend to show only the absolute changes in values and to allow potential cyclical patterns to be identified after removing the general drift, tendency, twists, bends, and effects of seasonal cycles of a set of time-series data. For example, a detrended dataset may be necessary to discover a company's true financial health—one may detrend increased sales around Christmas time to see a more accurate account of a company's sales in a given year more clearly by shifting the entire dataset from a slope to a flat surface to better see the underlying cycles and fluctuations. The resulting charts show the effects of the detrended data against the original dataset, and the statistics reports show the percentage of the trend that was removed based on each detrending method employed, as well as the actual detrended dataset.
- **Volatility: GARCH Models.** The Generalized Autoregressive Conditional Heteroskedasticity model is used to model historical and forecast future volatility levels of a time-series of raw price levels of a marketable security (e.g., stock prices, commodity prices, and oil prices). GARCH first converts the prices into relative returns, and then runs an internal optimization to fit the historical data to a mean-reverting volatility term structure, while assuming that the volatility is heteroskedastic in nature (changes over time according to some econometric characteristics). Several variations of this methodology are available in Risk Simulator, including **EGARCH, EGARCH-T, GARCH-M, GJR-GARCH, GJR-GARCH-T, IGARCH, and T-GARCH.** The dataset has to be a time series of raw price levels.
- **Volatility: Log Returns Approach.** Calculates the volatility using the individual future cash flow estimates, comparable cash flow estimates, or historical prices, computing the annualized standard deviation of the corresponding logarithmic relative returns.
- **Yield Curve (Bliss).** Used for generating the term structure of interest rates and yield curve estimation with five estimated parameters. Some econometric modeling techniques are required to calibrate the values of several input parameters in this model. Virtually any yield curve shape can be interpolated using these models, which are widely used at banks around the world.
- **Yield Curve (Nelson-Siegel).** An interpolation model with four estimated parameters for generating the term structure of interest rates and yield curve estimation. Some econometric modeling techniques are required to calibrate the values of several input parameters in this model.

Real Options Valuation, Inc.
4101F Dublin Blvd., Ste. 425, Dublin, California 94568 U.S.A.
admin@realoptionsvaluation.com | www.realoptionsvaluation.com | www.rovusa.com