Research Methods

By

Dr. Paul Cottrell

June 12, 2016

Overview

In this Breadth component, an analysis of research methods in quantitative finance pertaining to experimental, cross-sectional, quasi-experimental, pre-experimental, sample design, econometric models, and stochastic simulations is put forward. A compare and contrast will be performed for each of the above research methods in quantitative finance. This component will include a discussion of how these methods contribute to better understanding on hedging and pricing of financial assets. A discussion of the strengths and limitations of each research design method investigated pertaining to quantitative finance will be put forth.

Experimental design is an important research method due to the power of inference. Frankfort-Nachmias and Nachmias (2008) proposed four major types of research design with differing degrees of inference and safeguards. The four major types of research design proposed by Frankfort-Nachmias and Nachmias are: (1) experimental, (2) cross-sectional, (3) quasi-experimental, and (4) pre-experimental (p. 103). Experimental designs involve assigning random samples to the experimental and control groups with independent variables applied to the experimental group. Cross-sectional designs usually are associated with survey research, whereby manipulation to determine inference is diminished compared to experimental design. Quasi-experimental research involves random sampling from a population but not random assignment within testing groups. Pre-experimental are the weakest form of research methods due to the lack of random selection and assignment.

In sample designs there are probability samples and non-probability samples (Frankfort-Nachmias and Nachmias, 2008, p. 167). Frankfort-Nachmias and Nachmias (2008) considered three types of non-probability sample designs: convenience, purposive, and quota samples (p. 168). Frankfort-Nachmias and Nachmias also considered four probability sample designs: simple random, systematic, and stratified (pp. 169-173).

Another important set of tools in research design, especially in the field of finance is econometrics. Econometrics is utilizing regression analysis to determine relationships between dependent and independent variables. This breadth component will also provide an investigation by Wooldridge (2009) in three broad categories in econometrics: introduction to econometrics, simultaneous equation models, and advanced time series topics. Simultaneous equation models are a set of system of equations that describe financial and economic systems. A review of time series analysis for the field of finance and economics is important to understand the temporal dynamics of the investigated variables. For example, a time series analysis on price of the S&P 500 might reveal certain volatility dynamics.

In the stochastic modeling section a review of some random sampling methods and how to improve on the random sampling homogeneity are investigated. In complex financial problems Monte Carlo simulations are frequently utilized with stochastic processes built in the simulation. This type of research method can help in making a more realistic and statistically sound research design. Stochastic models can also help in forecasting price curves when the simulation parameterizes a Levy process, whereby the Levy process defines the drift, volatility, and jumpdiffusion components of the historic time series. By taking the Levy process parameters the Monte Carlo simulation can generate probability paths of the price curve of interest.

In the conclusion section an important question was examined. How to best construction a research study that is related to quantitative finance research questions, especially relating to time series data from asset markets? It's within the intent of this Breadth component to answer this important question by using different research methods in quantitative finance pertaining to the categories of experimental, cross-sectional, quasi-experimental, pre-experimental, sample design, econometric models, and stochastic simulations.

EXPERIMENTAL

In this section a discussion on the components of a research design and types of experimental designs are investigated. There are four components of a typical research design: comparison, manipulation, control, and generalization (Frankfort-Nachmias and Nachmias, 2008, p. 94). Frankfort-Nachmias and Nachmias (2008) considered the comparison component as a way to establish correlation among independent and dependent variables (p. 94). In quantitative research a comparison could be between price and volatility of asset returns. The price could be representing the dependent variable, whereas volatility represents the independent variable. Frankfort-Nachmias and Nachmias contended that the manipulation component is important to establish causality, whereby changing the independent variable and measuring the change in the dependent variable might suggest a statistical relationship (p. 95). In our example of price and volatility: if volatility changes x percent then price might change y percent which might establish statistical significance. For the third component, control is important to consider in the research design. Internal validity can be established if the proper controls are in place during the experiment (p. 95). If a researcher can establish that the independent variable did cause a change in the dependent variable and not some other mediating set of variables then proper control was conducted in the experiment. In quantitative finance perhaps price is also affected by investor memory and not just volatility. The experiment would need to investigate how price changes with volatility and investor's memory of previous price conditions. The last component is generalization, which is defined as the ability to project the conclusions of the experiment to the larger population (p. 101). This is also considered external validity.

It is possible that the sample size is not truly representative of the population dynamics. For example, our time series on price of an asset might be during times of financial turmoil leading to the conclusion that a certain percentage of change in price is from a certain change in market volatility. The problem is that the time series might be non-linear and that the percentage of change in price relative to volatility might be representative in a power law distribution – the system might be convex or concave. So the external validity would only be established by increasing the period range to include turmoil and tranquil markets.

There are five experimental design types to consider: (1) controlled experimentation, (2) Solomon four-group, (3) post-test-only control group, (4) effects through time, and (5) factorial design (Frankfort-Nachmias and Nachmias, 2008, pp. 103-107). Frankfort-Nachmias and

Nachmias (2008) proposed that controlled experimentation is the strongest method of establishing causation but external validity can be compromised (p. 103). Frankfort-Nachmias and Nachmias described a controlled experiment as a design that allows for pretesting, post testing, and control-experiment group comparisons (pp. 103-104). In quantitative finance it is possible to construct a controlled experiment. For example, to determine if forecasting accuracy increases, a researcher can compare using constant volatility or time varying volatility. The control group is calculated with constant volatility and the experimental group is calculated with the time varying volatility. The pretest is done by measuring forecasting accuracy on a random sample of assets among the control and experimental groups. Then exposure to a time varying volatility measure in the experimental group is conducted and a new forecast is calculated. In the post test the mean forecast accuracies are compared to determine if differences are statistically significant.

A Solomon four-group design is similar to the classic controlled experiment but there are an additional set of control and experimental groups that are not pretested to determine if pretesting affects the results (Frankfort-Nachmias and Nachmias, 2008, p. 104). This type of research design might be useful in a simulation application in finance, whereby the experiment can be constructed assuming certain initial conditions and determine if these initial conditions affect the post testing results.

A post test only control group design is usually a more practical experiment design because in many cases it might be cost prohibitive to pretest (Frankfort-Nachmias and Nachmias, 2008, p. 106). A post test only control group is the same as a classical experimental design except that no pre testing is conducted on the experimental or control group. In quantitative finance an example would be to compare the performance of forecasting between the standard Black-Scholes pricing model used on the control group of assets and a Levy process pricing model used on the experimental group of assets.

Research designs that determine the effects of time are important, especially in non-linear dynamic systems. Experimental designs used to study effects extended in time that are constructed similarly to the Solomon four-group design but have more post testing is performed for each time period of interest (Frankfort-Nachmias and Nachmias, 2008, p. 107). In quantitative finance effects on a system through time are very important, therefore many research designs should be utilizing additional post testing. For example, in financial time series there are lagged effects within the time series. By looking at x periods out one can determine the multiplicative or decaying effect.

When considering more than one independent variable in a research design the researcher needs to develop a factorial design to determine how each independent variable contributes to the dependent variable (Frankfort-Nachmias and Nachmias, 2008, p. 108). Utilizing a matrix chart to organize the possible combination among the variables can be useful to the researcher. In quantitative finance, researchers are interested in a multitude of variables that affect price, such as drift, volatility, and price jumps. Regressing price with many different variables and determining statistical significance will help describe the financial system better, which in turn perhaps leads to better forecasting.

CROSS-SECTIONAL

A cross-sectional design is usually employed with survey research, whereby random samples of individuals are asked a series of questions (Frankfort-Nachmias and Nachmias, 2008, p. 116). The questions are independent variables and the response by the individuals is the dependent variable. In behavioral finance, a cross-sectional design can be utilized to determine investor behavior or perceptions related to stress. For example, a random sample of investors could be asked a series of questions about risk tolerance to determine if they are an equity or bond investor. Frankfort- Nachmias and Nachmias (2008) considered that cross-section designs are not as powerful of a research tool as traditional experimental designs because before-and-after comparisons are not possible (p. 117).

Some of the survey research methods employed in a cross-sectional design are: (1) mail questionnaire, (2) personal interview, and (3) telephone interview. The advantages of a mailed questionnaire are the following: (1) low cost, (2) reduction in biasing error, greater anonymity, (3) considered answers, and (4) accessibility (Frankfort-Nachmias and Nachmias, 2008, p. 207). Frankfort-Nachmias and Nachmias (2008) suggested that the disadvantages of a mailed questionnaire are: (1) requires simple questions, (2) no opportunity for probing, (3) no control over who fills out the questionnaire, and (4) low response rate (p. 207).

Personal interviews have the following advantages: (1) flexibility, (2) control of the interview situation, (3) high response rate, and (4) collection of supplemental information (Frankfort-Nachmias and Nachmias, 2008, p. 218). But Frankfort-Nachmias and Nachmias (2008) warn that the researcher must realize the disadvantages of personal interviews are: (1) high cost, (2) interview bias, and (3) lack of anonymity, which all are strengths in a mailed survey (p. 219). In terms of advantages of a telephone interview, Frankfort-Nachmias and Nachmias proposed that there are moderate costs, speed, high response rate, and quality (p. 223). But the disadvantages of a telephone interview are the reluctance to discuss sensitive topics, the possibility of broken-off interviews, and the lack of supplemental information (p. 223).

Since cross-sectional designs uses surveys a researcher needs to consider the types of questions to ask. There are three types of question structures: (1) closed-ended, (2) open-ended, and (3) contingency questions (Frankfort-Nachmias and Nachmias, 2008, p. 233). Closed-ended questions are ones that an interviewee is asked to choose the most appropriate response to a question from a presented list. Open-ended questions are questions when the interviewee answers a question with no presented list to choose from. Usually open-ended questions are answered in sentence or essay format, whereby the interviewee's answers are captured in full. Frankfort-Nachmias and Nachmias (2008) described contingency questions as a special type of closed-ended question, whereby a filter question is initiated and if the interviewee passes the filter question then the interviewer will initiate the contingency questions (p. 235).

When considering a survey the interviewer needs to determine the question format. There are many different formats, such as ratings, matrix questions, and ranking types (Frankfort-Nachmias and Nachmias, 2008, pp. 236-238). But beyond the format of the questions the interviewer needs to determine the sequencing, whereby Frankfort-Nachmias and Nachmias

(2008) suggested funneling and inverted funneling sequencing (pp. 238-239). In terms of funnel sequencing Frankfort-Nachmias and Nachmias describes the questioning process as a progressively narrowing of scope based on the answers of the previous question (p. 238). Inverted funnel sequencing is the opposite of funnel sequencing, whereby specific questions are progressively broader (p. 239).

Cross-sectional designs are good for behavioral finance research, but are not adequate for quantitative finance because the research questions involved in quantitative finance are not survey oriented – the research questions usually are more experimental in design. For example, if a researcher was interested in risk tolerance of an investor a cross sectional design utilizing a survey with funneling sequencing might be useful to narrow down the specific behavioral characteristics for high risk tolerance.

Quasi-Experimental

A quasi-experimental design allow for random selection from the population but does not require the random assignment to comparison groups, whereby reducing the internal validity compared to classical experimental designs (Frankfort-Nachmias and Nachmias, 2008, p. 118). Frankfort-Nachmias and Nachmias (2008) proposed four types of quasi-experimental designs: (1) contrasted-groups; (2) planned variation; (3) panels and time-series; and (4) control-series designs (pp. 118-127). In Frankfort-Nachmias and Nachmias' book contrasted-groups are quasi-experimental designs that compare categorical groups, such as region or political affiliation (p. 118). A simple example of a contrasted-groups design used in behavioral finance would be categorical groups that represent metropolitan regions and compare the mean percentage of savings of each categorical group, which allows for a statistical comparison to be made. Pretesting and post testing can be performed on comparison groups in a contrasted-groups research design (p. 120).

Planned variation designs are where a researcher has designed an experiment where individuals are exposed at different levels of the independent variable to determine causal effects (Frankfort-Nachmias and Nachmias, 2008, p. 122). As Frankfort-Nachmias and Nachmias (2008) suggested, planned variation designs have a tendency to weaken internal validity relative classical experiments because the comparison groups are not randomly assembled leading to possibility of biased conclusions on causal effects. In a behavioral finance application, a planned variation design could be used to determine if a causal effect exists between a required margin in a futures account and volatility of market prices would increase risk aversion for investors.

Panels and time-series designs are used to determine changes in the dependent variable along time (Frankfort-Nachmias and Nachmias, 2008, p. 123). Panel studies are more rigorous because the same comparison group is examined at different time intervals, but Frankfort-Nachmias and Nachmias (2008) warns that this type of design is difficult in social science experiments because individuals assigned to a comparison group need to be observed over a certain time period – suggesting the problem of individuals dropping out of the research study (p. 124). When considering time series designs Frankfort-Nachmias and Nachmias proposed that a typical time series has multiple pretesting and post testing observations (p. 124). Time series designs are common in economics and finance. Econometrics is a field of study in economics, whereby regression analysis is performed on large datasets that are constructed through time. For example, an economic policy can be measured to determine efficacy using time series designs.

Control-series designs are time-series designs that can compare different groups, whereby causal effects can be established (Frankfort-Nachmias and Nachmias, 2008, p. 128). For example, comparing Michigan and the average of the other States on income level over time can be a control-series design experiment. This sort of design can be useful to economists and policy makers. Frankfort-Nachmias and Nachmias (2008) suggested that the major advantage with control-series designs is that history, maturation, and testing effects are shared by all comparison groups allowing for control of these characteristics (p. 128).

Pre-Experiment

Pre-experimental is a very weak inferential tool for researchers. Pre-experimental designs do not allow for random assignment to comparison groups and experimental manipulation; therefore internal and external validity is compromised (Frankfort-Nachmias and Nachmias, 2008, p. 131). Frankfort-Nachmias and Nachmias (2008) proposed that the one advantage to a pre-experimental design was that it allowed researchers to gather information when other research methods were not available, whereby the research would conclude that further studies would need to be conducted (p. 133). When considering the continuum of inquiry pre-experimental designs would be more qualitative and less quantitative. The disadvantage of pre-experimental designs articulated by Frankfort-Nachmias and Nachmias is that internal and external validity are extremely week (p. 133). A one-shot case study is a type of pre-experimental design, which is similar to a phenomenological study were a single group or event is investigated (p. 131). Due to the lack of a significant sample size, validity of causation is compromised. For example, in quantitative finance, we might start an investigation on the causes of illiquid markets by asking a hedge fund manager how they reacted to the financial crisis of 2008. Per-experimental designs could be used to pilot a research direction.

Sample Design

There are two types of sampling: probability and non-probability sampling (Frankfort-Nachmias and Nachmias, 2008, p. 167). Frankfort-Nachmias and Nachmias (2008) considered a probability sample as to be able to specify the probability of a sample representing the population, whereas a non-probability sample is not being able to set a probability of a sample to the population (p. 167). From Frankfort-Nachmias and Nachmias there are four types of probability sample designs: (1) simple random sample, (2) systematic samples, (3) stratified samples, and (4) cluster samples (p. 169). In simple random sampling each of the sampling units has equal and known nonzero probability for assignment (p. 169). In quantitative finance we could randomly select equities out of the S&P 500 index for sample unit assignment. Systematic sampling is when the first sample is randomly selected but subsequent samples are selected at a set criterion, e.g. every 5^{th} individual. In our S&P 500 index example, we arrange the index in alphabetical order and randomly select on equity and then select every 5^{th} equity thereafter. A researcher should keep in mind that systematic sampling might introduce sample bias, but it is more convenient then simple random sampling.

Stratified samples are used to ensure that minority groups are represented in the sampling process (Frankfort-Nachmias and Nachmias, 2008, p. 171). It is possible that a minority group might not be sampled adequately through a simple random sampling or a systematic sampling because of the sporadic nature of the diffusion in the population. The problem with stratified sampling is that the population composition needs to be known to adequately have all concerned groups represented based on the true population distribution. In economics stratified sampling might be utilized to sample disposable incomes of various racial groups in a country to ensure no sample bias to certain racial groups exists. Frankfort-Nachmias and Nachmias (2008) described cluster sampling as a method used in large-scale research studies due to its low costs to initiate (p. 173). As Frankfort-Nachmias and Nachmias stated a large group is randomly selected and further subdivision sampling are conducted within the large group (p. 173). It is possible to use cluster sampling in economics though the following example. If a researcher was interested in understanding the changes in disposable income in a state and relative to the sales tax rate a cluster sample of a particular county within in a State can be randomly sampled and subdivided.

In terms of non-probability sampling there are three main types: (1) convenience samples, (2) purposive samples, and (3) quota samples (Frankfort-Nachmias and Nachmias, 2008, p. 168). Frankfort-Nachmias and Nachmias (2008) concluded that convenience sampling is utilized when a quick and easy to assemble sample is needed (p. 168). In finance, if a researcher needs to conduct a quick survey of employment a researcher could convenience sample her surroundings. Frankfort-Nachmias and Nachmias suggested that purposive samples are constructed using subjective means that the researcher believes represents the population composition (p. 168). It does seem that purposive sampling is more representative of the population than convenience sampling but the subjectivity needs to be considered to understand the possibility of sample bias. A quota sampling is making a conscious effort to sample to a known ratio in the population (p. 168). Quota sampling seems similar to a stratified sampling but quota sampling is still considered a non-probability sample design because other factors that are represented in the population might not be in the quota sampling technique, e.g. race or age – when the quota considered in the population is gender.

ECONOMETRICS

Introduction to econometrics

Econometrics is a very important tool to analyze economic and quantitative finance themes. Econometrics involves regression models to describe the economic or financial dynamics. There are single factor and multi factor regressions to describe these economic and financial environments. For example, equation 1 is a simple regression.

$$GDP = \beta_0 + \beta_1 Productivity + u \tag{1}$$

GDP is gross domestic production, β_0 is the intercept, β_1 is the parameter for productivity, and u is the error term. β_1 is the change in GDP with a unit change in productivity. If $\beta_1 = .7$ then GDP would increase by .7 with one more productive unit.

The ordinary least squares (OLS) method is the typical method used to produce the regression model. The OLS method is used for estimating residuals and a regression line is fitted with the dataset with the least error. There are certain assumptions that need to be met to make sure the OLS is unbiased. Biased OLS conditions reduce the internal validity of the regression analysis. There are basically four assumptions for unbiased OLS regressions for a single factor linear regression: (1) linear in parameters, (2) random sampling, (3) sample variation in the explanatory variable, and (4) zero conditional mean (Wooldridge, 2009, pp. 47-49). Wooldridge (2009) defined that linear in parameters is when the dependent variable is related to the independent variable and the error term in a population (p. 47). Random sampling is axiomatic and needs no further explanation. Wooldridge explained that the sample outcomes on the independent variable are not all the same (p. 48). Lastly, the expectation of the error term given any independent variable is zero (p. 49).

It is important to also consider the homoskedasticity, where the variance of the error term given the independent variable is a constant (Wooldridge, 2009, pp. 47-49). In summary, the homoskedastic assumption and the expected value of the error term of zero are key assumptions to assure that the regression model's internal validity is not compromised.

The assumptions for a multiple factor linear regression are the following: (1) linear in parameters, (2) random sampling, (3) no perfect collinearity, and (4) and zero conditional mean for the error term (Wooldridge, 2009, pp. 84-87). Woodridge (2009) stated that under the assumptions for a multiple factor linear regression a model can be derived using unbiased OLS estimators (p. 88). Woodridge mentions that if an independent variable is correlated to the error term then the independent variable is considered an endogenous explanatory variable (p. 88). But if the independent variable is not correlated to the error term then the independent variable is considered an exogenous explanatory variable (p. 88). A researcher should be careful in including irrelevant variables in a regression model. Over specifying the model does not have an effect on the unbiased coefficients but it can have effects on the variance of the OLS estimators (p. 89). When an under specified model is used the researcher should also beware of accuracy of the regression. Under specified models can create a biased OLS estimator (p. 91). This degree of bias is determined by the correlation of variables and the sign of the omitted coefficient (p. 91).

In terms of the variance of the OLS estimators, a researcher needs to understand the accuracy of the OLS estimator to determine validity of the regression model. In multiple regressions there is the assumption of homoskedasticity for the error term relative to any of the explanatory variables (Wooldridge, 2009, p. 95). Wooldridge (2009) calculated the size of the sample variance to determine the precision of an OLS estimator, e.g. large values mean less precise estimators (p 95). Wooldridge considered that the components for a sample variance are the standard deviation, total sum of the squares, and R-square (p. 95). The larger standard deviation means larger variance for the OLS estimators (p. 95). The larger the total sum of the squares the smaller the sample variance and the more accurate the OLS estimator – when all others components are constant (p. 96). When R-squared is large the sample variance is larger – when all other components are constant (p. 97).

There is a tradeoff between bias and variance when considering including a variable in a regression model (Wooldridge, 2009, p. 99). Wooldridge (2009) had the following rules when two independent variables are uncorrelated with each other (p. 100).

- When the population coefficient for the second term does not equal zero then the coefficient for the first term for the simple regression is biased and the coefficient for the first term from the multi regression is unbiased with the sample variance for the simple regression being less than the sample variance of the multi regression for the first term's coefficient.
- When the population coefficient for the second term equals zero then the sample and multi regression coefficient for the first term are both unbiased with the sample variance for the simple regression being less than the sample variance for the multi regression for the first coefficient.

In many multiple regression analyses OLS is used but the researcher should consider statistical methods that control for over or under specifying a regression model. Some methods for doing this are stepwise routines, whereby a regression model is built up one independent variable at a time to see if the overall statistical significance increases and that the individual independent variable are relevant. Some software packages can also rearrange the stepwise order of a multiple regression to determine which explanatory variables are pertinent – order of the stepwise routine can bias the statistical results.

Simultaneous equation models

Simultaneous equation models (SEM) are important in understand complex dynamic systems (Wooldridge , 2009, p. 547). For example, since supply and demand can be modeled in equilibrium then the supply equation can equal the demand equation. Wooldridge (2009) defined the observable variable as the explanatory variable but the error term as the unobservable variable, whereby the explanatory variables and the error terms are uncorrelated (p. 547). When the explanatory variable is correlated to the error term then this leads to bias in the OLS estimator. Wooldridge considered that simultaneity bias occurs when the dependent variable defines one of the explanatory variables in another equation and is correlated to the error term of that second equation (p. 552).

The identified equation in a two-equation system is the equation that can be estimated given a random sample on the explanatory variables (Wooldridge, 2009, p. 553). Wooldridge (2009) stated that when the explanatory variables are exogenous and they are different in each equation then this condition is considered an exclusion restriction on the model (p. 554). Wooldridge also stated that when solving a system of simultaneous equations one is to determine the rank condition and order condition (pp. 554-555). The easy way to remember the identifier equation is to see if there are any exogenous variables not in an equation but is located in other equations, whereby the equation missing the exogenous variable is the identified equation (p. 554). The order condition is determined by counting the endogenous and exogenous variables, whereby there is at least as many excluded exogenous variables as included endogenous variables in the structural equation (p. 524).

SEMs with time series are excellent for defining a country's economy (Wooldridge, 2009, p. 560). Wooldridge (2009) used the example that the consumption of a country can be modeled with income, tax receipts, interest rate, investment, and government spending (p. 560). Other complex economic systems can also me model and are used in dynamic stochastic equilibrium models (DSGE). DSGE is a system of equations that help model macroeconomic conditions, which is an advanced topic in econometric model. Sometimes simultaneous equations have lagged explanatory variables, whereby a dependent variable is a function with a previous time period explanatory variable. Wooldridge suggested that lagged variables can be considered endogenous or exogenous to the error term of the dependent variable function, but typically lagged variables are considered endogenous – correlated to the error term (p. 562).

Shown in equation 2 through 4 is an example of a three-equation system (Wooldridge, 2009, p. 559).

$$y_1 = \alpha_{12}y_2 + \alpha_{13}y_3 + \beta_{11}z_1 + u_1 \tag{2}$$

$$y_2 = \alpha_{21}y_1 + \beta_{21}z_1 + \beta_{22}z_2 + \beta_{23}z_3 + u_2 \tag{3}$$

$$y_3 = \alpha_{32}y_2 + \beta_{31}z_1 + \beta_{32}z_2 + \beta_{33}z_3 + \beta_{34}z_4 + u_3 \tag{4}$$

The z explanatory variables are exogenous and the y variables are endogenous to the error term u. The first subscript defines the equation number and the second subscript defines the variable number. Also the α parameters are considered endogenous and the β parameters are exogenous. Equation 1 is the identified equation because z_2, z_3, and z_4 are excluded; therefore it meets the order condition.

Panel data can be used in a SEM system. Panel data are longitudinal, in which there are many cross sections of data throughout time. Therefore, SEM can be good at modeling systems that are affected throughout time. The use of dummy variables is also possible with SEM systems. Dummy variable are used to determine the explanatory affects of non-numerical variables, e.g. gender or nationality.

Due to the use of logarithms in SEM systems an explanation on their meaning should be investigated. See table 1 for interpreting the logs in a regression model (Wooldridge, 2009, p. 46). Logarithms in a regression allow for non linear relationships in a regression model.

Table 1

Model	Dependent Variable	Independent Variable	Interpretation of Beta
Level-Level	Y	X	Δy=β Δx
Level-Log	Y	log(x)	Δy= (β/100)% Δx
Log-Level	log(y)	X	%Δy=(100β) Δx
Log-Log	log(y)	log(x)	%Δy=β% Δx

Note: Interpreting the logs in a regression model.

In a level-level model the dependent and independent variables do not have logs, therefore the unit change in the dependent variable is caused by a unit change in the independent variable. For a log-log level then the percentage change in the dependent variable is caused by a percentage change in the independent variable. Wooldridge (2009) considered the beta of the independent variable as the elasticity of y with respect to x in a log-log model (p. 46). Wooldridge also defined 100 x beta as the semi-elasticity of y with respect x for a log-level model (p. 45).

Advanced time series topics

An infinite distributed lag (IDL) model is when the dependent variable is defined by current and all past values of the independent variables, see equation 5 (Wooldridge, 2009, p. 624). This is similar to the lagged functions described in the previous economic section of this Breadth component, lagged independent variables extended to the indefinite past. Wooldridge (2009) suggested that due to there being an infinite amount of lagged coefficients there is some restrictions on estimating the model (p. 626). Wooldridge considered using a Koyck distribution lag model to estimate the coefficients in the IDL model (p. 626). A coefficient only depends on two parameters which are shown in equation 6 (p. 626).

$$y_t = \alpha + \delta_0 z_t + \delta_1 z_{t-1} + \delta_2 z_{t-2} + \ \ldots + \ \mu_t \qquad (5)$$

$$\delta_j = \gamma \rho^j, |\rho| < 1, \ \ j = 0,1,2, \ldots \qquad (6)$$

A standard model with a lagged dependent variable can be obtained from an IDL model (p. 627). The result of the simplified IDL model is shown in equation 7 (p. 628).

$$y_t = \alpha_0 + \gamma_1 z_{t1} + \ldots + \gamma_k z_{tk} + \rho y_{t-1} + v_t \qquad (7)$$

As can be seen in equation 7, the dependent variable is equal to the contemporaneous independent variable and a lagged dependent variable - with coefficients governing the contemporaneous and lagged variables respectively.

The Koyck distribution lag model is considered a special case and a more general definition of an IDL model can be obtained through a rational distribution lag (RDL) model (Wooldridge, 2009, p. 628). The RDL model is simply adding a lagged exogenous variable, which was only a contemporaneous exogenous variable in the Koyck distribution lag model - coefficients are defined the same as in the Koyck distribution lag model (p. 628). Reference equation 7 for the RDL model mathematical definition, where for equation 8:

$$y_t = \alpha_0 + \gamma_0 z_t + \rho y_{t-1} + \gamma_1 z_{t-1} + v_t \qquad (8)$$

When dealing with time series analysis testing for unit roots is important for a researcher or practitioner. A unit root test determines if the time series has drift or not. See equation 9 for an AR(1) model testing for a unit root (Wooldridge, 2009, p. 630). AR(1) is an autoregressive model.

$$y_t = \alpha + \rho y_{t-1} + e_t, \; t = 1,2, \ldots \qquad (9)$$

If $\alpha=0$ and $\rho=1$ then the time series is a random walk without drift, but when $\alpha\neq0$ and $\rho=1$ the time series is a random walk with drift (Wooldridge, 2009, p. 631). Wooldridge (2009) considered testing the unit root with the Dickey-Fuller test or the augmented Dickey-Fuller test (pp. 631-633). The main different between the normal Dickey-Fuller test and the augmented Dickey-Fuller test is that the latter test contains lagged changes in y, whereas the former test only has a lagged y. When the Dickey-Fuller test or augmented Dickey-Fuller test is statistically significant then the time series is non-stationary.

STOCHASTIC MODELING

Monte Carlo simulations are very important analytical tools in quantitative finance. Since non-linear dynamic systems are very difficult to determine through an analytical solution a researcher needs to rely on numerical instruments, such as Monte Carlo simulations. Monte Carlo simulations used in quantitative finance are basically a computer simulation that determines an evolutionary path of price or some other important metric. For example, researchers can use Monte Carlo simulations to determine the probability of a certain range of prices for a particular asset. The idea is to use stochastic variables that are integrated in the Monte Carlo simulation and re-run the simulation with much iteration. The researcher would then understand the possible ranges of prices and determine from the simulation an average ending price of a probability cone for a range of prices.

Stochastic processes have random characteristics. A researcher needs to understand that randomization in computers might actually be clumpy and not uniform, especially when generating random vectors. For example, in MATLAB the random generator will generate a degree of lumpiness in a two dimensional matrix. If a researcher were to random generate from this random vector for their stochastic process the Monte Carlo simulation result would be slightly biased, which weakness internal validity. There are four other common methods to solve the lumpiness problem for random vector generation: (1) Halton, (2) Faure, (3) Sobol, and (4) Hypercube (Huynh, Lai, and Soumare, 2008, p. 107). Each one of these alternatives reduces the degree of lumpiness. But does a researcher really need to adjust their stochastic process to correct the lumpiness when modeling in MATLAB? First let's examine the four other random generators. Huynh, Lai, and Soumare (2008) used the Halton sequence, which is derived from the Van Der Corput sequence, and compared it to the MATLAB generator and found that the Halton sequence substantially reduced lumpiness in a two dimensional matrix (p. 94). Huynh, Lai, and Soumare tested the Halton sequence by using the Van Der Corput sequence, which uses different bases for each dimension, e.g. base 2 for dimension one and base 3 for dimension two (p. 93). The base 2 and base Van Der Corput sequences are the following respectively (p. 92):

0, 1/2, 1/4, 3/4, 1/8, 5/8, 3/8, 7/8, 1/16, 9/16, 3/16, 11/16, 5/16, 13/16, …
0, 1/3, 2/3, 1/9, 4/9, 7/9, 2/9, 5/9, 8/9, 1/27, 10/27, 19/27, 2/27, …

Even though the Halton does improve the lumpiness of the random vector generator compared to MATLAB the Halton sequence would need many dimensions to fill a two dimensional matrix to each homogeneity (p. 94).

The Faure is similar to the Halton sequence but differs by using the same base for all dimensions and permutation of the terms is implemented (Huynh, Lai, and Soumare, 2008, p. 95). In Huynh, Lai, and Soumare (2008) found that with 1000 terms homogeneity was reached with a base of 31, whereas not in a Halton sequence homogeneity was not reach even with a base of 103 (pp. 94-97). Huynh, Lai, and Soumare warned that in a Faure sequence calculation performance diminishes at about the 25th dimension (p. 97). The Faure sequence is determined by using the smallest prime number greater than the number of dimensions in the simulation (p. 96). From a qualitative perspective the lumpiness in graphing a base 2 and 3 Halton sequence and graphing a base 3 Faure sequence is about the same – an improvement from the MATLAB random generation.

Sobol sequances are also built from Van Der Corput sequences but involves binary forms of terms (Huynh, Lai, and Soumare, 2008, p. 97). Sobol sequences are quite involved and require an algorithm to generate. Huynh, Lai, and Soumare (2008) suggest the follow basic process (p. 99):

1. Select a primitive polynomial for a certain dimension.

2. Use the initial values (M) based on the type of polynomial used.

3. Determine position direction (k) for the decomposition of n term in base 2.

4. Compute a new M by recurrence if needed with the chosen polynomial in 1.

5. Compute values for y_{n+1}.

6. Compute x_{n+1} using

$$x_{n+1} = y_{n+1}/2^{m_{n+1}}.$$

7. Repeat step 3 through 6 to generate following sequence.

It was found the Sobol Sequence does improve homogeneity compared to the MATLAB random generator process in three dimensions and computationally performs better than Halton and Faure sequences in higher dimensions (Huynh, Lai, and Soumare, 2008, p. 107).

Hypercube sampling differs from the previous sequencing methods by being completely random, whereas the other sequencing methods were deterministic through the use of the Van Der Corput sequences (Huynh, Lai, and Soumare, 2008, p. 107). Huynh, Lai, and Soumare (2008) suggested a three step process to implement a hypercube sampling: (1) divide each dimension into N sections; (2) generate uniform random variable and randomly distribute it in the first section; and (3) repeat the step 2 for the remaining sections (p. 101). Huynh, Lai, and Soumare showed that the hypercube sampling has similar distribution characteristics in low dimensions as in MATLAB and that the homogeneity of the hypercube sampling is slightly more (p. 102).

How does the different sampling techniques compare to the MATLAB generation? Should research implement different strategies for sequence sampling to improve their stochastic simulations? It was found that when trying to estimate the volume of a unit cube that MATLAB, Halton, Faure, Sobol, and hypercube methods perform differently and reach convergence at different number of iterations (Huynh, Lai, and Soumare, 2008, p. 103). Huynh, Lai, and Soumare (2008) found that Halton, Faure, and Sobol converge faster than MATLAB and the hypercube in three dimensions (p. 103). But Huynh, Lai, and Soumare did find that Halton and Faure sequencing did lose efficiency faster than the Sobol sequencing, whereas the Hypercube sampling method did not lose efficiency with a more uniform distribution to MATLAB's uniform random generator (p. 105). In conclusion, for estimating a three dimensional problem Sobol performs the best in terms of convergence time.

In an option-like problem MATLAB, Halton, Faure, Sobol, and Hypercubes were compared to determine which method performs best, whereby this problem is still three dimensional. It was found that Halton, Faure, and Sobol mostly performed better in terms of standard error relative to MATLAB for all number of simulations, e.g. 200 though 1000 simulations (Huynh, Lai, and Soumare, 2008, p. 107). Huynh, Lai, and Soumare (2008) found that after 1000 simulations that MATLAB performed substantially worse than all other sampling methods in terms of standard error (p. 107). In conclusion, for three dimensional random sampling for stochastic processes the research needs to weight computational time, convergence efficiency, and the number of iterations to perform for accurate results. In addition, Hypercube sampling seems to perform similar to MATAB in terms of standard error, suggesting not much difference in homogeneity in large number of simulation iterations. In most applications it seems that MATLAB's uniform random generator is adequate for Monte Carlo simulation utilizing a stochastic process when number of iteration are adequate, but this might result in more computation time to generate the Monte Carlo simulation. If a researcher desires to be more computationally efficient and converge faster in their research study then Sobol sequencing might be a better method to implement in the Monte Carlo simulation.

CONCLUSION

In this section a discussion of the strengths and limitations of some research design methods pertaining to experimental, cross-sectional, quasi-experimental, pre-expermental, sample design, econometric models, and stochastic simulations.

Experimental designs have four typical design components: comparison, manipulation, control, and generalization. Each one of these components was elaborated upon. There are five experimental design types to consider: (1) controlled experimentation, (2) Solomon four-group, (3) post-test-only control group, (4) effects through time, and (5) factorial design. Leading to the conclusion that the controlled experimentation design type is the most powerful in establishing causation but external validity might be compromised.

Cross-sectional design is usually employed with survey research, whereby a random sample of individuals is asked a series of questions. For example, a random sample of investors could be asked a series of questions about risk tolerance to determine if they are an equity or bond investor. Cross-section designs are considered not as powerful of a research tool as traditional experimental designs because before-and after comparisons are not possible. Some of the survey research methods employed in a cross-sectional design are: (1) mail questionnaire, (2) personal interview, and (3) telephone interview.

A Quasi-experimental design allows for random selection from the population but does not require the random assignment to comparison groups, thereby reducing the internal validity compared to classical experimental designs. There are four types of quasi-experimental designs: (1) contrasted-groups; (2) planned variation; (3) panels and time-series; and (4) control-series designs. Frankfort-Nachmias and Nachmias (2008) suggested that the major advantage with control-series designs is that history, maturation, and testing effects are shared by all comparison groups allowing for control of these characteristics.

Pre-experimental is a very weak inferential tool for researchers. Pre-experimental designs do not allow for random assignment to comparison groups and experimental manipulation; therefore internal and external validity is compromised. When considering the continuum of inquiry pre-experimental designs would be more qualitative and less quantitative. The disadvantage of pre-experimental designs articulated by Frankfort-Nachmias and Nachmias (2009) is that internal and external validity is extremely weak.

There are two types of sampling: probability and non-probability sampling. There are four types of probability sample designs: (1) simple random sample, (2) systematic samples, (3) stratified samples, and (4) cluster samples. Each one of these four types of sample designs was expounded upon.

Econometrics is a very important tool to analyze economic and quantitative finance themes. Econometrics involves regression models to describe the economic or financial dynamics. A discussion of the ordinary least squares method and how to use it in econometric regressions was put forth in this Breadth component. In many multiple regression analyses the ordinary least squares method is used but the researcher should consider statistical methods that control for over or under specifying a regression model. Some methods for doing this are stepwise routines, whereby a regression model is built up one independent variable at a time to see if the overall statistical significance increases and that the individual independent variables are relevant.

Simultaneous equations models are important in understanding complex dynamic system. For example, since supply and demand can be modeled in equilibrium then the supply equation can equal the demand equation. Simultaneous equation models with time series are excellent for defining a country's economy. Other complex economic systems can also be modeled and are used in dynamic stochastic equilibrium models. Dynamic stochastic equilibrium models are a system of equations that help model macroeconomic conditions, which is an advanced topic in econometric modeling.

Advanced topics in times series analysis was also investigated in the Breadth component. A discussion of infinite distubuted lag models and rational distribution lag models were put forward. The use of the normal Dickey-Fuller test and the augmented Dickey-Fuller test was discussed when testing for unit roots for stationary time series.

Monte Carlo simulations are very important analytical tools in quantitative finance. Since non-linear dynamic systems are very difficult to determine through an analytical solution a researcher needs to rely on numerical instruments, such as Monte Carlo simulations. Stochastic processes have random characteristics. A researcher needs to understand that randomization in computers might actually be clumpy and not uniform, especially when generating random vectors. There are four other common methods to solve the lumpiness problem for random vector generation: (1) Halton, (2) Faure, (3) Sobol, and (4) Hypercube. Each one of these random methods were explored and expounded upon.

In this Breadth component different types of research designs have been evaluated and sub themes seemed to have emerged. The researcher needs to be cognizant of the limitation of a research design and make sure that the appropriate statistical tests are performed on time series. For example, when using the Black-Scholes model for pricing options there are certain assumptions that need to be made with the time series investigated. Those assumptions are constant volatility, normality, no transaction costs, constant drift, and no autocorrelation. It was shown in this Breadth component on how to test for most of these assumptions statistically.

In the Depth component an examination of current research on research methods and designs for the field of quantitative finance is put forward. Research methods are compared and contrasted on how to analyze hedging and price methods for financial assets. An investigation and evaluation of recent trends with research designs for the use in quantitative finance to develop and establish hedging and pricing techniques will be conducted.

Reference List

Frankfort-Nachmias, C. & Nachmias, D. (2008). *Research methods in social sciences* (7th ed.). New York: Worth Publishers.

Huynh, H. T., Lai, V. S., Soumare, I. (2008). *Stochatic simulation and applications in finance with MATLAB programs*. West Sussex, England: John Wiley & Sons Ltd.

Wooldridge, Jeffery. M. (2009). *Introductory econometrics: a modern approach* (4th ed.). Ohio: South-Western Cengage Learning.