Common Biases In Business Research

Sulaiman Mouselli[1], Hiba Massoud[2]

Introduction

In research, bias occurs when an error is introduced into sampling or testing which results in selecting or encouraging one outcome, conclusion, or answer over others. Bias can happen at any phase of research, including study design, methodology selection, data collection, and stating conclusions [1]. Given the significant threats of these biases on the reliability and validity of research conclusions, understanding different types of biases, their consequences, and treatment methods is the corner stone in avoiding such biases and an important step in critically evaluating research. This chapter discusses biases that are common in quantitative research, biases associated with quantitative research and biases that usually occur in quantitative research using qualitative data. It will focus on introducing business researchers to their definitions and sources. The chapter also suggests methods to uncover those biases and provides remedies and ways to deal with such biases.

1. Biases in Quantitative Business Research:

Quantitative research in business is characterized with extensive use of numerical data which could be time-series, cross-sectional, or panel that requires the application of various estimation techniques. This exposes researchers to a unique set of biases when conducting such quantitative type of research in business. Following is a discussion of the definition, sources, consequences and possible remedies of biases that face quantitative researchers.

1.1. Sample size bias:

Researchers usually prefer larger sample size on smaller ones in an attempt to increase the precision with which they can estimate population parameters. Under normal conditions, a larger sample size increases the reliability of estimates, reduces the standard error in t-statistic, and increases the probability of rejecting the null hypothesis of insignificance. However, [2] explains that two risks may arise in the attempt to increase the sample size:

Firstly and mainly, sampling from more than one population. That is, increasing the sample size may result in drawing observations from different population and consequently lead to misleading conclusions. For example, the inclusion of certain sectors in financial markets research may alter the results towards the dominant sectors in the markets which is the financial sector (banking and insurance), the dominant sector in many developing financial markets. Moreover, Researchers are usually prone to include companies from the financial sector in their asset pricing tests despite their special characteristics on the aim to increase the size of their samples. This means that estimates and conclusions drawn from such sample will be erroneous.

Secondly, increasing the size of the sample usually results in additional expenses that outweigh the benefit of increased accuracy of estimates. Moreover, a large biased sample is no better than a small biased sample.

[1]Arab International University, Damascus, Syria, S-mousele@aie.edu.sy, Department of Finance, Faculty of Business Administration

[2]Coventry University, Coventry, UK, ac5148@coventry.ac.uk, School of Economics, Finance and Accounting

1.2. Data mining bias:

Data mining is the practice of extensively searching a data set for statistically significant relationships until a desired pattern is discovered [2]. During such process, large number of hypotheses about a data set are tested in a very short time, usually without a plausible theoretical background, by searching for combinations of variables that might show a correlation. Researchers may also be prone to split their samples into subsamples that confirm a certain desired conclusions.

Given that a large number of hypotheses are tested and many subsamples are tried, it is expected that some hypotheses will appear to be highly statistically significant although they are merely coincidences. Data mining bias most commonly occurs when;

First, researchers have not formed a hypothesis in advance, and are therefore open to any hypothesis suggested by a data. For example, when studying the factors that affect stock returns, researchers through all available variables (accounting and economic) in the model and hope to figure out a number of significant relationships. Then, they try to explain these relationships which usually lack any economic meaningful story.

Second, when researchers narrow the data used in order to increase the probability of the sample accepting or refuting a specific hypothesis. Sample selection needs to be appropriately selected and any exclusion should be clearly justified. For example, splitting the main sample into two subsamples on the basis of the introduction of new accounting standards or regulations is reasonably justified. However, doing the same sub-sampling without such justification is not usually accepted and is a source of great concerns.

Third, when data is constructed on some empirically motivated characteristics instead of theoretically motivated characteristics. [3] find that tests of asset pricing models may yield misleading inferences when tested portfolios are constructed on the basis of the same certain properties used to construct the examined risk factors. [4] use 25 size-book-to-market portfolios to test the explanatory power of their three-factor model. Those portfolios turn to be the standard testing portfolios in any evaluation of asset pricing models. However, these portfolios are constructed on the basis of empirically motivated characteristic of the stock, such as market capitalization and book-to-market ratio rather than a theoretically motivated characteristic, such as dividend yield, or a market-based characteristic, such as industry sector. The results of such tests will be biased toward out-weighting risk factors that are constructed in a similar way to those of the tested portfolios.

Data mining biases are also apparent in testing the profitability of some technical trading rules in Finance. Researchers tend to continue examining trading rules that worked in the past and ignore those that did not success. In other word, they are using the same dataset to construct trading rules and to test them which clearly prefer one trading rule on others [5].

To test for data mining bias in the trading rule example, [6] and [7] suggest the use of bootstrapping through placing the examined trading rule in a universe of broadly similar trading rules and testing the whole universe of trading rules before reaching a conclusion on the successful trading rule. The bootstrap should be applied to each trading rule through the use of sampling with replacement from the time-series of observed returns of this trading rule. The significance level from the bootstrap test can then be used to examine if superior technical trading rule really exists.

Fourth, when certain estimation method has been adopted to the dataset with neither a clear justification of the selection of this methodology nor a discussion of the advantages and disadvantages of the adopted model and other possible methodologies. Researchers tend to apply Ordinary Least Squares (OLS) in their model estimation and hypothesis testing while ignoring other more robust estimation techniques such as Maximum Likelihood (ML) and General Method of Moments (GMM).

There are a number of warning signs that data mining bias might exist. One of these signs is when too much digging to reach the conclusion is noticed. This usually involves testing numerous variables and trying subsamples until one that appears significant is discovered. Another sign is when no plausible story can be reached from the claimed significant relationship. Furthermore, it could be a sign of a bias when the researcher does not provide enough robustness tests to support her empirical results.

The best way to avoid data mining bias is to clearly justify the sample selection and to provide full rationalization of the model estimation and hypothesis testing methods supported by prior research in addition to "out-of-sample" tests to check whether the apparently significant relationships continue to hold.

The idea of out-of-sample testing is essentially that data not used in model estimation, is used for model testing. In other words data should be divided into two sets; training and test sets. It is useful to divide data, using a portion of the data (the training set) for the development of the model, and reserving a portion of the data (the test set) for testing the model that is built. The principle is that if you build a model on a particular set of data, it will of course test quite well. By dividing the data and using part of it for model development, and testing it on a separate set of data, a more convincing test of the model accuracy is obtained [8] .A relationship observed in the estimation period that is purely the result of data mining, and is therefore spurious, is very unlikely to be repeated for the out-of-sample period. Therefore, models that are the product of data mining are likely to fit very poorly in out-of-sample tests and to show different results.

1.3. Sample-selection bias:

It is the bias that results from the elimination of certain observations (such as firms, funds) or assets from a study due to different reasons. [9] argues that there are two reasons for sample selection bias to arise in practice; "*First, there may be self selection by the individuals or data unit being* *investigated. Second, sample selection decision by analysts or data processors operates in much the same fashion as self selection*".

This bias could also be a database-specific that results from using historical information and suffers from a type of sample-selection bias known as survivorship bias. Survivorship bias results from using a dataset that only includes the survivors over a period, not the full set of firms that were listed over this period [10]. Many databases suffer from this bias especially those that only list firms or funds that currently alive (for example, Compustat for United States accounting information), which means firms that have failed are not included in the database. Consequently, results obtained from the study will be biased towards firms or funds that are successful and may not accurately reflect the true picture.

Sample selection bias is even worse in studies of credit rating and hedge fund returns. If firms that would have received lower credit ratings (because they have weak financials) opt not to request a rating, results of the determinants of credit rating will only be valid for firms with high credit rating, and coefficients will be inconsistent [5]. Similarly, because hedge funds are not required to publicly disclose their performance data, only funds with good performance will choose to disclose their performance. Hence, hedge fund returns will be overestimated and coefficient estimates will be inconsistent.

To control for sample-selection bias in hedge fund studies, [11] suggests two-step procedure that involves first estimating a 0-1 probit model of whether the fund solicits to disclose its performance and second estimating the ordered probit model for the determinants of the performance of hedge funds.

A common bias in sample selection occurs also when a researcher selects the sample from the currently available observations and ignores firms that disappear during the sample period. This type of bias mixes the sample selection bias with look-ahead bias because it only includes firms with available information at a certain recent point of time. One way to avoid this bias is through considering all dead and live cases (firms) during the whole sample period.

1.4. Look-ahead bias:

Look-ahead bias can be defined as "*the use of information in a simulation that would not be available during the time period being simulated, usually resulting in an upward shift of the results*"[12]. This type of bias arises when a researcher includes data that would not have been known or available at the time period being tested. Including such data will compromise the accuracy of the results.

For example, the false assumption that accounting data become available immediately at the end of the financial period is a source of look-ahead bias. A trading rule on the basis of book-to-market ratio cannot be constructed until data on book value is released in the financial statements of the company otherwise the true performance of the trading strategy being tested is falsified. Many studies recognize this fact and choose to construct portfolios at the end of June each year to give investors enough time to know, absorb and construct portfolios on the basis of book-to-market ratio that is usually made available at the end of December each year.

1.5. Time-period bias:

The results of a test will be biased if it is based on a certain specific time period which is usually referred to as time-period specific [2]. Researchers usually face the dilemma of the choice between short and long time periods. A relatively short-time period raises the doubt that the results are only valid during that particular period. On the other hand, a very long time period may fail to uncover structural changes that took place during that period. This type of bias is highly likely to exist as a result of some unique changes during other periods. Such bias could result in having the research results working well during a specific time frame but not working good for extended time periods.

For example, business related research regarding the Syrian external trading activities (exports and imports) trends done before 2011. When those same studies are conducted today for a 10-year timeframe, the conclusions might be quite different. The same could be said about Damascus Securities Exchange performance-related research. Circumstances have changed dramatically after the political crises that researchers could not consider time periods before and after crises homogeneous and therefore researchers should avoid time series overlapping years before and after 2011.

Researchers should ideally choose their studies' time periods carefully. This should be done after understanding the nature of data in its' environment, then studying and testing time frames towards unique effects or circumstances, in order to ensure that the conclusions aren't specific to one unique period or environment.

1.6. Omitted variable bias:

When applying the Ordinary Least Squares (OLS) technique in model estimation, researchers sometimes intentionally or unintentionally drop one or more important explanatory variables. The bias is created when the model compensates for the missing variable through over– or under–estimating the importance of the remaining included explanatory variables.

Omitted variable bias may present in two occasions. On one hand, researchers may exclude explanatory variables from the estimated model in the presence of multicollinearity. On the other hand, researchers may not be aware of certain important explanatory variables and consequently do not include them in the model. However, if the removed variable was relevant in the data-generating process of the dependent variable, an omitted-variable bias will result [5].

Omitted variable bias can be easily detected by the magnitude and significant of the intercept term of the estimated model. A large and highly significant intercept indicates that one or more important explanatory variables are missing from the model. The researcher should give enough care to the theory behind the model and should not be driven only by previous empirical results.

2. Biases in Qualitative Business Research

The following types of biases occur mainly but not exclusively in qualitative business research.

2.1. Sample-size bias

Qualitative research is highly likely to use quantitative sampling techniques when the generalizability is their ultimate goal and therefore could encounter a number of related sampling biases similar to those related to quantitative research. Sampling from more than one population is also a source of bias in qualitative business research. Increasing the sample size may result in including observations from different population and consequently lead to misleading conclusions. For example, in questionnaires to investigate the views of small size population (for instance, the higher management levels) where fewer numbers of observations are available and usually difficult to get hold off, researchers are usually prone to reclassify managerial levels or include lower management levels in the sample. This will bias the results of the study and may lead to different conclusions.

Researcher should bear in mind that there is neither a universal answer on the ideal sample size, nor a universal rule for calculating the size of samples for qualitative research [13]. Therefore, instead of focusing thoroughly and solely on increasing sample size, the researches should be concerned about other important issues in sampling, such as judgment. In sampling, judgments have to be made about some important issues enlightened by the nature and coverage of the study as well as the types of questions to be answered and the way to answer them. These issues includes access, representation, the quality of data and strategies of sampling [14].

Determining the size of qualitative research sample should be made in the light of a number of criteria such as the research purpose, question and design, the population size and its heterogeneity or homogeneity, the confidence level and confidence interval required, the accuracy required, the likely response rate, and the type and number of variables used for statistical tests. Some guides to test sample adequacy are available. [15] Provides a formula that gives an indication of the sample size adequacy that could be used for both quantitative and qualitative research. This formula, however, could only be used as an indicator given that the researcher has some good information and understanding about the research population. The sample size indicator formula allows the researcher to form an opinion on the adequacy of his sample size by comparing it to n being the required sample size as calculated using the occurrence of the condition (state) P and the maximum error required E, as:

$$n = \frac{P \times (100 - P)}{E^2}$$

Another way of testing sample size adequacy for generalizability purposes; is the ratio of entries (observations) to independent variables. This is particularly useful when quantitative techniques are used to analyze qualitative data. In essence, this ratio should be greater than 15 entries (but preferably greater than 20) for each independent variable. This means that there should be 20 twenty entries (observations) for each independent variable. For example, if the study involves 8 independent variables the sample size should be greater than 160 observations. Whereas, the higher the ratios of entries to independent variables the lower the risk of making the results too specific to the sample "overfitting" as this decreases the generalizability of the results [16]. However, when generalizability is not an ultimate goal for a qualitative research it could be useful to study in detail small samples. In such cases the two techniques suggested above become irrelevant and it is then up to the researchers to try to justify and amend the size of their sample to meet their research goals.

2.2. Sample-selection bias

Representativeness is a very important factor to consider when selecting the study sample as this leads to stronger sample external validity and increases the confidence in the generalizability of the survey results to the targeted population.. Therefore, describing the relationships between samples and their populations is very significant. In order to convey such a relationship, the researcher must be able to describe it in terms of characteristics that are common to both the sample and its parent population [17]. Again, this requires the researcher to know a lot about the population. Such available information about both the characteristics of the sample and its population allows examining the representativeness of the sample and possible biases using tests such as the t-test. If a bias is detected, this could be due to many factors such as sample design and coverage, and non-response.

It is also common that researchers are unable to reach all the members of a target population for a number of other reasons such as being unable to get the needed permission to access the data, in addition to time and resources limitations. In such cases the researcher must identify that portion of the population which is accessible. However, difficulties might lead researchers sometimes to use non-probability sampling such as choosing the respondents because of them being easy to access or have a links with some already accessed observations or respondents such as surveying customer satisfaction or investors decisions by using a friends' network or work-based networks.. This is called convenience sampling. Researchers may also use snowball sampling based on the reference of some respondents. These sampling processes are highly likely to result in having biased samples which include specific types of respondents (respondents sharing similar characteristics) and exclude others [18]. Furthermore, such processes give well socially connected respondents higher chances of being selected compared to others who are less socially connected [19].

Researchers are advised to plan their sampling and to identify the kind of sampling strategy (qualitative) they require (probability, non-probability, or mixed methods sample). When using probability sampling, all members of the population have equal chances to being included; the choice to include them in the sample is purely random and based on chance. The aim is for generalizability and wide representation and this lead to less risk of bias in the sample. For non-probability sampling, researchers should understand their samples and need to be aware how the consequences of samples' shortages could affect the generalizability of their research's results.

2.3. Measurement bias

Researchers need to be careful when using data that are collected for specific reasons or when changes introduced to the way data are collected [20]. Data that are recorded erroneously on purpose suffer from deliberate distortion problem that is very common when using secondary data sources such as corporations' reports and their related records. For instance, the turnover rate of part time or hourly based employers might be intentionally ignored in the reports to enhance such indicative ratio. Similarly, bad environmental practices might deliberately be allocated smaller paragraphs providing minimum details in the social responsibility reports compared to good environmental practice in an attempt to signal better performance by manipulating the readers' opinions. Deliberate distortion is more pronounced when data is collected for specific reason and/or for the interest of a particular group. Another good example for this type of bias might occur when consumer satisfaction surveys are prepared by employees who tend to underestimate and give low credit to customers' undesirable observations and notes in an attempt to show better service quality levels to their supervisors, management teams and stockholders.

One way to deal with this deliberate distortion problem is through cross-check verification of data [21]. That is to triangulate the findings of main study with one or more independent data sources. Consequently, researchers and reviewers will be more confident of the data used in the research and less doubtful of the conclusions they reach especially if they are inconsistent with previous results.

A change to the way data are collected is a major source of measurement bias. Once changes to data collection process occur whether because of a altering members of the data collection team or modifying the procedures used to collect data, there is a possibility to reach different conclusions. This is a very serious issue for data based on questionnaires which are adjusted from time to time such as Corruption Perception Index issued by Transparency International and for longitudinal surveys such as the Price Indexes where researchers focus on trends development [20]. However, government-sponsored sources are more likely to announce such changes than within-company sources. Hence, researchers should be warned of such measurement bias and need to take them into account when reaching conclusions or commenting on them.

2.4. Interviewer bias

This type of bias occurs when the comments, tone, style of language, or non-verbal behavior (such as, facial expressions, body language, and manner of dress) of the interviewer affect how interviewees react when questions are posed by this interviewer [20]. Social status, race, and gender can also produce bias. Additionally, bias happens when the interviewer unconsciously enforces his/her own believes in the way he/she are imposing the questions. When interview questions are biased, they influence interviewees' answers. Leading questions, misunderstood questions, and unanswerable questions are examples of biased questions. Questions' order could also be one source of bias. Furthermore, bias will also occur in the way the researcher interpret responses. Interviewer bias challenges the validity and reliability of research when the researchers do not start by building trust with their respondents or seen as lacking confidence or credibility by the interviewees.

Some of the interviewer biases are unavoidable, but many of them could be reduced or controlled. For example, the interviewers' body language and gestures should not be excessive. The way they look and talk should also be as neutral as possible. They should not give opinions while collecting date and should try to avoid all sources of question bias.

2.5. Interviewee or response bias

Interviewee bias is a major issue in in-depth and semi-structured interviews as this is likely to affect the quality of data collected by researchers. This type of bias occurs when the respondents consciously or unconsciously misrepresent the truth [22]. The Interviewee may initially accept to take part in an interview but may still be sensitive or uncomfortable when it comes to the unstructured exploration of particular topics. Hence, they intentionally mislead researchers by giving false answers in order to hide their lack of knowledge or to avoid embarrassment. Providing falsified answers could also be stimulated when interviewees are faced by sensitive questions or those asking for critical information that they do not have the authority to reveal, or do not want to discuss with the interviewer. . Such misleading behavior of providing untrue answers for some of the questions will result in providing a mix of true and false parts of an incomplete picture the interviewer might be trying to draw regarding the situation, incident, or organization they are trying to interview about and study.

Another aspect in Interviewee bias is the type, characteristic, nature, positions and grade of the respondents who accept to take part of the research as interviewees [20]. Given the fact that interviews are time-consuming and could require the interviewees to undertake some follow up before and after the interview (to arrange and then check contents), there will usually be a reduction of the willingness to be interviewed of those who initially agreed to take part in the research. This issue needs to be taken into account through proper sampling.

Different respondents' circumstances, intentions, or un-intentions lead to different types of interviewee biases. For example, consistent biased answers may be noticed when respondents try to maintain consistency in their answers. This leads all their next answers to be influenced by their earlier answers. Although the first answer and some other following answers might be true, some of the following answer statements might be untrue. Bias could be caused by a dominant respondent; this type of bias is likely to happen during focus groups where some interviewees could influence the opinions and answers of the others by dominating talks time, vocalizing their experience and knowledge, positions, attractiveness, etc. Overestimation is another bias that might be associated with focus groups. This type of bias occurs when respondents overstate their reactions, opinions, or answers because of being influenced by the group's thoughts, actions, and expressions. Some interviewee bias types could be related to psychological status of the interviewees. For example, extreme feelings bias could occur if an interviewee is in an extreme mood state. The moods of such interviewees could affect their cognitions, and their answers could reflect their moods. Angry people tend to provide angry or pessimistic answers. Busy or work-stressed people may provide short and brief answers. In the same regards, it is also possible that some interviewees may get angry with the interviewer or the interview settings. An opposite case, however, may be noticed when an interviewee chooses his/her answers to satisfy the moderator, sponsor, or interviewer. Respondents may also tend to provide socially acceptable answers that are not true causing social acceptance/desirable bias.

Another feeling related bias could arise during interviews related to politics in non-democratic countries where respondents are reluctant to give their opinions regarding political issues, parties, or public figures because of the fear of possible consequences and/or punishments. In similar reactions, respondents would rather not talk in response to questions tackling sensitive subjects. Respondents may also choose to give false answers to hide secrets, causing sensitivity bias. Finally, some bias could be a result of errors where interviewees forget or have fade memories.

The researcher must be aware of the possible interviewee biases and respond according to the type of bias expected. Some advices are choosing a suitable time and place for the interview, checking the mood state of the interviewees, and building trust with them are essential. Judging the answers of the respondents helps in assessing their truthfulness or exaggeration. In such cases, researchers should ask for clarification and challenge the respondents in a friendly way without revealing too much about themselves. Projective techniques or indirect questions are sometimes helpful, especially with sensitive topics. Managing people during interviews is very important, especially during focus groups as it involves more than five respondents and could involve different types of bias. Dominant respondents should be kept in check and other respondents should be allocated equal talk time.

2.6. Observer bias

This bias may occur when observers alter the outcomes of a study due to the way they interact with subjects in the study and in what observers choose to see [20]. This bias happens because an observer cannot detach him/ herself from the social phenomena he/ she is trying to study, or avoid depending on his /her previous experience, culture, attitudes, beliefs, feelings, views, state of mind, reference, error, and personality when he /she tries to interpret and analyze the study data or report its results. The conscious and subconscious are at work and researchers are human. Hence, although observer bias cannot be completely eliminated, researchers should be aware of the consequences of such possible bias on the reliability of their data collection, analyses and reporting. They should acknowledge it s and seek ways to control it. For instance, if a researcher believes that firm performance is not related to the percentage of female representation on the board of directors, then he will tend to formulate his hypothesis to reflect his believes and tend to interpret his findings accordingly.

There are two main methods to avoid such bias. The first method is self-verification where the observer asks himself questions about his conclusions. For instance, he can ask himself: did the answers received really mean that? And what other possible interpretations that could be put forward. The second method is informant verification. If observers rely on informal discussions to reach conclusions, she can write her conclusions from these discussions and ask her informants to verify the content and may be suggest alternative conclusions.

2.7. Transformation bias

As seen earlier, both quantitative and qualitative methods of research try to achieve the highest possible levels of reliability and validity of their measures of social reality away from biases. In qualitative research methods, the measures are considered valid when they produce the right (non-falsified) answers and will considered reliable, when they are trustworthy in that they produce consistent and replicated answers when the same or other similar measures are used. The application of these positivist measurement concepts to qualitative research is deemed inappropriate by a number of scholars. They criticized and denigrated the process of collecting, and analyzing qualitative data as well as the way they apply to reach conclusions for the high likelihood of them being subjective. Qualitative researchers try to take steps to reduce errors and bias, examples of such steps explained above.

The choice of qualitative or quantitative methods or data, however, is then a matter of what is suitable and appropriate for the topic being searched and the research question being asked (Johnson and Harris, 2002). Many quantitative research studies gather and analyze data of qualitative nature that contributes to addressing specific research questions. Some other research questions may require mixed research methods. Quantitative analysis of qualitative data is common in management and accounting research. In order to quantitative analyze qualitative data;

qualitative data needs to be transformed into quantitative formats. A number of available methods could be used to transform qualitative data into quantitative formats. These methods include scoring, ranking, and codifying.

Qualitative data used in quantitative research is likely to face two different categories of research bias sources. The first category of bias sources is associated with its' qualitative nature. Therefore, it could be vulnerable to bias of qualitative research, such as measurement bias, interviewer bias, response bias, observer's bias, or other bias associated with qualitative data, as explained earlier. The second category of bias sources is associated with the process of transforming qualitative data into quantitative formats through scoring, ranking, or codifying.

Transforming qualitative data into quantitative formats involves data transformation bias. Such bias is related to the level of researcher's personal judgment exerted during the transformation process. The judgment itself decreases when the transformation technique is straightforward and increases when the technique gets complicated, this depends on the type and the level of informativeness of the qualitative data to be transformed. For example, it is simple to code a response of yes/no answer into binary variable. However, it is quite complicated to code a text (or contents) of a piece of writing into various categories to capture the interests of the writer and their importance to her (for example, disclosure quality and corporate governance quality). The more complicated the coding, the more the researcher's judgment applied, the higher the likelihood of transformation bias.

To minimize such data transformation bias, it is important for the researcher to develop tools such as; classification schemes, check lists, selection criteria, decision rules, and coding protocols. These help to make replicable and valid inferences from data according to their context.

3. References and Bibliography

[1] C. J.Pannucci and E. G.Wilkins, "Identifying and avoiding bias in research," *Plast. Reconstr. Surg.*, vol. 126, no. 2, pp. 619–625, 2011.

[2] B. Shajani, *11th Hour Guide for 2015 Level I CFA*. 2015.

[3] A. Lo and A. MacKinlay, "Data-snooping biases in tests of financial asset pricing models," *Rev. Financ. Stud.*, 1990.

[4] E. Fama and K. French, "Common risk factors in the returns on stocks and bonds," *J. financ. econ.*, 1993.

[5] C. Brooks, *Introductory Econometrics For Finance*, Third., vol. 1. New York: Cambridge University Press, 2014.

[6] R. Sullivan, A. Timmermann, and H. White, "Data-snooping, technical trading rule performance, and the bootstrap," *J. Finance*, 1999.

[7] H. White, "A reality check for data snooping," *Econometrica*, 2000.

[8] Olson, David L. and D. Delen, *Advanced data mining techniques*. Springer Science & Business Media2008 ,.

[9] J. Heckman, "Sample selection Bias as Specification Error," *Econometrica*, vol. 47, no. 1, pp. 153–162, 1979.

[10] E. Gilbert and D. Strugnell, "Does survivorship bias really matter? An empirical investigation into its effects on the mean reversion of share returns on the JSE (1984–2007)," *Invest. Anal. J.*, pp. 1–19, 2015.

[11] J. Heckman, "Sample selection bias as a specification error," *Econom. J. Econom. Soc.*, 1979.

[12] P. Daniel, G., Sornette, D., & Woehrmann, "Look-ahead benchmark bias in portfolio performance evaluation. ," *J. Portf. Manag.*, vol. 36, no. 1, pp. 121–130, 2009.

[13] F. J. Fowler, *Survey Research Methods*, 2nd ed. London: Sage Publications Ltd, 2002.

[14] L. Cohen, L. Manion, and K. Morrison, *Research Methods in Education*, 5th ed. New York: Routledgeflamer, 2003.

[15] A. Easterby-Smith, M., Thorpe, R. and Lowe, *Management Research: An Introduction*, 1st ed. London: Sage Publications, 2002.

[16] W. C. Hair, J.F., Anderson, R.E., Tatham, R.L. and Black, *Multivariate Data Analysis*, 5th ed. Prentice Hall, 1998.

[17] A. Oppenheim, *Questionnaire Design, Interviewing and Attitude Measurement*. London: Pinter, 1992.

[18] S. Lucas, "Beyond the existence proof: ontological conditions, epistemological implications, and in-depth interview research," *Qual. Quant.*, 2014.

[19] S. Berg, "Snowball Sampling–I," in *Encyclopedia of Statistical Sciences*, and B. V. Samuel Kotz, Campbell Read, N. Balakrishnan, Ed. Hoboken, NJ: John Wiley and Sons, Inc, 2006, pp. 7817–7821.

[20] A. Saunders, M., Lewis, P., & Thornhill, *Research methods for business students*. Financial Times, 2007.

[21] G. L. Patzer, *Using Secondary Data in Market Research: United States and World-wide*. Westport: CT: Quorum Books.

[22] M. Sreejesh, S., Mohapatra, S., & Anusree, *Business research methods*. Springer International Publishing AG., 2014.