Why PLS-SEM is suitable for complex modeling?
An empirical illustration in Big Data Analytics Quality

Shahriar Akter (First Author)
School of Management, Operations & Marketing
Faculty of Business
University of Wollongong, NSW 2522, Australia
E-mail : sakter@uow.edu.au

Samuel Fosso Wamba (Second Author & Corresponding Author)
Toulouse Business School
Toulouse University
20 Boulevard Lascrosses, 31068 Toulouse, France
Email: fossowam@gmail.com

Saifullah Dewan (Third Author)
School of Information Systems & Accounting
Faculty of Business, Government and Law
University of Canberra, ACT 2601 Australia
Email: dmsaif6@yahoo.com.au

Why PLS-SEM is suitable for complex modeling?
An empirical illustration in Big Data Analytics Quality

Abstract

The emergence of multivariate analysis techniques transforms empirical validation of theoretical concepts in social science and business research. In this context, structural equation modeling (SEM) has emerged as a powerful tool to estimate conceptual models linking two or more latent constructs. This paper shows the suitability of the partial least squares (PLS) approach to SEM (PLS-SEM) in estimating a complex model drawing on the philosophy of verisimilitude and the methodology of soft modelling assumptions. The results confirm the utility of PLS-SEM as a promising tool to estimate a complex, hierarchical model in the domain of big data analytics quality (BDAQ).

Keywords

PLS-SEM, big data, big data analytics quality, business value, satisfaction

1. Introduction

Foresight of phenomenon and power over them depend on knowledge of their sequences, and not upon any notion we may have formed respecting their origin or inmost nature.

Mill (1865,p.266)

Quantitative research has made an enormous impact on social science and business research through its positivist epistemological belief since John Stuart Mill and the 19th century experimental researchers. This impact has gained a momentum by satisfying the need for causal modeling and empirical validation of theories to explain complex concepts (Blalock 1964; Bagozzi 1980; Huber and McCann 1982; Sawyer and Peter 1983; Iacobucci and Hopkins 1992; Hair et al. 2012; Stafford 2011). In this line of development, structural equation modeling (SEM) has emerged as a powerful multivariate analysis technique over the last four decades combining the features of the principal components and the regression analysis (Hair, Ringle, and Sarstedt 2012). Between the two approaches in SEM, the covariance based approach (CBSEM) is useful to confirm theoretically established relationships, however, this technique has distributional constraints (multivariate normality of the observed indicators) in estimating a large model (Chin 1998b; Fornell and Bookstein 1982; Hair, Ringle, and Sarstedt 2011). As a result, the focus of CBSEM mostly is on small conceptual models, which results into hindering the development and validation of large, complex models (Chin, Peterson, and Brown 2008).

This study refers a complex model to a large-hierarchical model which consists of 10 or more constructs and 50 or more items (Chin 2010). PLS-SEM (Wold 1975; Lohmöller 1989; Tenenhaus and Tenenhaus 2011) gains prominence in estimating such a complex-hierarchical model by eliminating the ambiguity of incorrect solutions (Becker, Klein, and Wetzels 2012; Wetzels, Odekerken-Schroder, and Van Oppen 2009). However, some researchers have questioned PLS-SEM's rigor (e.g.,Guide Jr and Ketokivi 2015; Rönkkö et al. 2016) despite its well established roots in management information systems (Chin and Gopal 1995; Chin 1998b; Ringle, Sarstedt, and Straub 2012a; Chin, Marcolin, and Newsted 2003; Marcoulides, Chin, and Saunders 2009), strategic management (Sarstedt, Ringle, and Hair 2014; Bentler and Huang 2014) and marketing (Fornell and Bookstein 1982; Chin, Peterson, and Brown 2008; Hair et al. 2012). Thus, in an effort to illuminate the rigor of PLS-SEM in estimating a complex model, we validate a hierarchical, reflective-formative model in the context of big data analytics quality (BDAQ).

Big data has emerged as the new oil, new soil, the next management revolution (McAfee and Brynjolfsson 2012b) and the ultimate force behind "transforming management theory and practice (George, Haas, and Pentland 2014). The extant literature shows that more than 91% of Fortune 1000 companies have embraced big data analytics (Kiron, Prentice, and Ferguson 2014) and achieved 5-6% higher growth in firm performance than competitors (Akter and Wamba 2016). Despite its skyrocketing optimism, many firms still struggle to capitalize on big data analytics (BDA) to derive quality insights. Competitive advantage from BDA is waning as managers grapple to understand the complexity of analytics quality in the emerging data economy (Ransbotham, Kiron, and Prentice 2016). Thus, for empirical illustration, the study first, presents an overview of PLS-SEM applications to estimate a complex model in the context of BDA and prediction oriented analysis (Gefen, Straub, and Rigdon 2011; Rigdon 2014). We define BDA as an integrated approach to collect and process big data in order to provide actionable insights for managerial decision making. Second, the study applies PLS-SEM to develop and validate a complex big data analytics quality (BDAQ) model and its impact on business value (BVAL) and BDA satisfaction (BDAS) in a nomological network. We define BDAQ as user perceived analytics quality model that measures overall excellence or superiority of the BDA platform. Conceptually, this study extends quality modelling in big data using the resource based theory (RBT) and methodologically, it presents the rigor of PLS-SEM as the ultimate tool for complex modeling. The remainder of the paper is organized as follows: Section 2 focuses on literature review and the conceptual model; Section 3 discusses the research methodology; Section 4 highlights the findings based on empirical illustration; Section 5 discusses the techniques for assessing a complex model and finally, section 6 concludes the paper with limitations and future research directions.

2. Literature review

2.1. PLS-SEM

It struck me that it might be possible to estimate models with the same arrow scheme by an appropriate generalization of my LS algorithms for principal components and canonical correlations.

(Wold 1982, p.200)

There is no doubt that PLS-SEM has become very popular for survey research in recent years since its introduction in 1966 by Herman Wold. The development of PLS-SEM is largely driven by its advantages in distributional assumptions, absence of factor indeterminacy and models with more parameters than observations (Dijkstra and Henseler 2015a). The PLS-SEM is regarded as a variance based approach to SEM (Chin, Marcolin, and Newsted 2003; Tenenhaus 2008) becomes appreciated for its ability to estimate both composites and factors (Henseler, Hubona, and Ray 2016). As an alternative to CBSEM, PLS-SEM was developed to estimate complex relationships and emphasize prediction while simultaneously relaxing the demands on data and specification of relationships (Dijkstra 2010; Chin, Peterson, and Brown 2008). Differently from CBSEM, PLS-SEM aims at estimating latent variable proxies (also called latent variable scores) according to a postulated model via an iterative sequence of ordinary least-squares regressions (Wold 1975, 1985). Recent publications on PLS-SEM using hierarchical modeling (Wetzels, Odekerken-Schroder, and Van Oppen 2009; Becker, Klein, and Wetzels 2012; Akter, D'Ambra, and Ray 2010; Akter, D'Ambra, and Ray 2011), consistent PLS (PLSc) for factor models (Dijkstra and Henseler 2015b), heterotrait-monotrait ratio of correlations (HTMT) for discriminant validity (Henseler, Ringle, and Sarstedt 2015), overall model fit using bootstrapping (Dijkstra and Henseler 2015a), interaction effects (Chin, Marcolin, and Newsted 2003; Fassott, Henseler, and Coelho 2016) and model specification (Sarstedt et al. 2016) providing a proof that it is able to define latent variables scores with well-defined statistical relations among them.

One of the main characteristics of the PLS-SEM is that it is able to estimate a model with a large number of latent variable and indicators even with a small sample size (Chin, Peterson, and Brown 2008). For complex models, PLS-SEM ensures *factor determinacy* by directly estimating latent variable scores, *factor identification* by introducing flexible residual covariance structure and above all, *robust prediction* in the context of small sample size, asymmetric distribution and interdependent observations (Chin 1998a, 1998b; Wetzels, Odekerken-Schroder, and Van Oppen 2009). These distinctive methodological features make PLS-SEM a possible alternative to the more popular CBSEM approaches for complex modeling (Henseler, Ringle, and Sinkovics 2009; Hair et al. 2012). As such, PLS-SEM is more suitable in a complex setting to validate large-hierarchical models by providing robust solutions (Chin, Peterson, and Brown 2008). This is echoed by Chin (2010, p.661), "It is under this backdrop of high complexity that PLS, regardless of whether applied under a strong substantive and theoretical context or limited/exploratory conditions, comes to the fore relative to CBSEM".

2.2. Complex Models and PLS-SEM

Truth, Existence, Knowledge, Causality, Identity, Goodness: these are the principal notions which philosophers examine. Intelligent persons normally have thoughtful and useful lives without pausing to look into these notions and into the connections between them. Once one starts to look into them, it is difficult to stop.

Stuart Hampshire (Pyke 2011)

The extant literature in social science and relevant research philosophies have contributed to the field of complex modeling using the philosophy of verisimilitude (i.e., trust likeness or nearness to the truth). For example, Meehl (1990) states that most models struggle to capture reality and suffer imperfection due to the imbalance between incompleteness and falseness. Whereas falseness represents the contradictions between the research model and the real world, incompleteness focuses on the ability to capture complex reality. Although these two philosophies play an important role in estimating reality, "Most SEM studies seem to focus on the falsity of a model as opposed to its completeness. In part because of algorithmic constraints, few SEM models are very complex (i.e., have a large number of latent variables). Emphasis on model fit tends to restrict researchers to testing relatively elementary models representing either a simplistic theory or a narrow slice of a more complex theoretical domain" (Chin, Peterson, and Brown 2008, p.294). The philosophy of verisimilitude urges to recognize that "scientific theories are never impeccably veridical in all aspects" (Rozeboom 2005, p.1335) and thus, practical theory adjudication should concentrate more on how a research model is true and to what extent it is true rather than whether a research model is true or not.

In exploring CBSEM, Shah & Goldstein (2006) identified an average of 4.4 latent variables and a mean of 14 indicators per model in a review of 93 articles. By comparison, in exploring PLS based models, Ringle et al. (2012a) identified an average of 8.12 latent variables and 27.42 indicators per model in a review of 65 studies published in MIS Quarterly. Hair et al. (2012) identified an average of 29.55 indicators per PLS path model in 204 studies of top 30 marketing journals. These results highlight the suitability of PLS-SEM as a tool to estimate a large, complex model (Chin, Peterson, and Brown 2008). In a similar spirit, Ringle et al. (2012b) comment "…prior studies appearing in scholarly journals (e.g., Reinartz, Haenlein, and Henseler 2009)—including those more critical of the PLS-SEM method (e.g., Lu et al. 2011) —indicate that PLS-SEM overcomes problematic model identification issues and that it is a powerful method to analyze complex models using smaller samples". Dijkstra and Henseler (2015a,p.10) support that PLS-SEM has "the possibility of estimating models having more variables or parameters than Observations". Although few studies in CBSEM focused on developing a large model using small sample size; these models are restricted by 3 items per LV to achieve goodness of fit (e.g., Marsh, Hau, and Wen 2004; Barendse, Oort, and Garst 2010). This constraint is criticized by MacCallum (2000) as it obstructs capturing the complexity of an empirical phenomenon. In this context, Blalock (1979, p.881) states, "reality is sufficiently complex that we will need theories that contain upward of fifty variables if we wish to disentangle the effects of numerous exogenous and endogenous variables on the diversity of dependent variables that interest us". He further adds that there is a natural imbalance between generalizability and parsimony in developing models, so 'parsimony' could be sacrificed in building complex models to describe more diverse settings and populations. In this case, PLS-SEM enjoys certain advantages in estimating complex models because of its flexible iterative algorithm and the soft modeling assumptions. Lohmoller (1989, p.64) comments, "It is not the concepts nor the models more the estimation techniques which are 'soft', only the distributional assumptions". Because of its flexibility in modelling both composites and factors, McDonald (1996) identifies PLS as a sophisticated multivariate analysis platform whereas Hair, Ringle, and Sarstedt (2011) label it as a silver bullet. As such, scholars across disciplines (e.g., Fornell and Bookstein 1982; Hulland 1999; Chin 2010; Hair et al. 2012; Chin, Peterson, and Brown 2008; Sarstedt et al. 2016; Henseler, Hubona, and Ray 2016) put forward PLS-SEM as tool of trade for survey research to capture complexity in models.

Our review examines all empirical studies using PLS-SEM published in *the Production Planning & Control (PPC) journal* from 2010 to 2015 indicates an average number of 5.4 constructs and 33.6 indicators per PLS-SEM model to embrace the complexity in capturing reality. Table 1 also shows that PLS-SEM studies in PPC used an average number of 194.8 samples. Although small sample size is the most frequently cited reason for using PLS-SEM, the review indicates that operations researchers used relatively large sample size which is clearly immune to threats from data inadequacies (Ringle, Sarstedt, and Straub 2012a). Overall, this review reflects the flexibility of PLS-SEM in handling large models with fewer restrictions.

INSERT TABLE 1 HERE

2.3. Big data analytics quality

Big data refers to the massive amount of structured and unstructured data which has four characteristics, that is, volume (i.e., huge quantity), variety (i.e., number, text, image, voice and video), velocity (i.e., speed) and veracity (i.e., reliability of data) (Fosso Wamba et al. 2015; Akter et al. 2016). According to Sanders (2016,p.28), “Big data without analytics is just a massive amount of data. Analytics without big data are simply mathematical and statistical tools and applications”. We define big data analytics (BDA) as an integrated data collection and analysis process to provide solid insights for managerial decision making (Akter and Wamba 2016). Although there is high adoption of BDA in recent years to obtain competitive advantage, many companies face enormous challenges to derive quality insights from data (Ransbotham, Kiron, and Prentice 2016). We define big data analytics quality (BDAQ) as the overall excellence or superiority of BDA platform perceived by its users. BDAQ also refers to the distinctive attribute of the overall analytics platform to produce valuable insights for business (Ji-fan Ren et al. 2016). The extant literature shows that the quality of technology and information determine the extent of business value in big data environment. In this regard, Barton and Court (2012) argue that both technology and information quality work as an ecosystem in producing solid insights for managers. Technology quality refers to the quality of the analytics platform that is reflected in system reliability, system adaptability, system integration and system privacy (Davenport, Barth, and Bean 2012; Nelson, Todd, and Wixom 2005). On the other hand, information quality represents the quality of data driven insights in terms of currency, format, accuracy and completeness (Nelson, Todd, and Wixom 2005). Wixom, Yen, and Relich (2013) show that the quality of technology and information in big data environment influence business value, which refers to the strategic benefits for firms. In addition to business value, scholars (Davenport 2006; McAfee and Brynjolfsson 2012a) also identify that the quality parameters influence user satisfaction, that determines sustainability of the analytics platform.

3. Research Model for Empirical Illustration

Drawing the Resource based Theory (RBT), the study views analytics quality as an analytics resource only if it is rare and costly to imitate (Ray, Muhanna, and Barney 2005). The RBT of BDA argues that BVAL and BDAS depend on the quality of resources that are valuable, rare, inimitable and properly organised. The RBT also focuses on complex connections among the heterogeneous resources, such as system and information quality, to examine BVAL and BDAS. Thus, using RBT as a theoretical foundation, the study examined commonly found dimensions of BDA that influence quality perception. The review identified two primary dimensions of BDAQ, that is, technology quality and information quality (Ji-fan Ren et al. 2016). BDAQ also emerged as a hierarchical construct throughout our review and theoretical exploration, which consists of two primary dimensions and eight subdimensions as shown in Figure 1 (Davenport, Barth, and Bean 2012; Davenport and Harris 2007; McAfee and Brynjolfsson 2012b; Fosso Wamba et al. 2015). Therefore, based on the systematic literature review, this study identifies BDAQ as a complex construct model because of its large number of dimensions and subdimensions under multiple hierarchies (See Figure 1).

INSERT FIGURE 1 HERE

We specify the proposed BDAQ model as a higher-order, reflective-formative model as the first-order dimensions are reflective (Mode A) and the higher-order dimensions are formative (Mode B) (Chin 2010; Ringle, Sarstedt, and Straub 2012b). We define the proposed quality model as a complex model because it involves large number of constructs and indicators under multiple levels and dimensions (Edwards 2001; Jarvis, MacKenzie, and Podsakoff 2003; MacKenzie, Podsakoff, and Jarvis 2005; Law and Wong 1999; Netemeyer, Bearden, and Sharma 2003). As part of embedding the higher-order quality model in a causal network, the study models it with criterion variables, such as, business value and satisfaction. We define 'satisfaction' as the overall attitudinal response by the big data analysts toward BDA and 'business value' as the degree of perceived benefits to the organization at a strategic level, e.g., competitive advantage (Wixom, Yen, and Relich 2013). The impact of BDAQ on BVAL is a dominant concern in big data environment (Wixom, Yen, and Relich 2013). The significance of the association between BDAQ and BVAL was highlighted by the extant literature (Lavalle et al. 2011; Wixom, Yen, and Relich 2013; Ji-fan Ren et al. 2016). Thus, we postulate that:

H1: BDAQ has a significant positive impact on business value (BVAL).

The extant literature identifies that the excellence of BDAQ has a significant positive impact on business value, which ultimately drives satisfaction of big data analytics users (Ji-fan Ren et al. 2016). This study argues that the assessment of BDAQ results in an affective or emotional response, such as BVAL and BDAS. In this regard, Golder et al. (2012) state that "[p]ositive quality disconfirmation increases satisfaction; negative quality disconfirmation decreases satisfaction". Thus, this study explores the link between quality-value-satisfaction and posits that:

H2: Business value (BVAL) has a significant positive impact on user satisfaction (BDAS).

H3: BDAQ has a significant positive impact on satisfaction.

The study identifies BVAL at the heart big data research because it will be directly influenced by BDAQ (Wixom, Yen, and Relich 2013). BVAL is identified as a mediator in the study because, first, BDAQ (predictor) influences BVAL (mediator); second, BVAL influences BDAS and, finally, BDAQ influences BDAS (i.e., the dependent variable) without any influence of the mediator (Baron and Kenny 1986). Thus, the mediating role of BVAL in big data analytics research is important to explore:

H4: Business value, as a mediator, influences the relationship between BDAQ and satisfaction.

4. Methodology

4.1. Data Collection and Sampling

Table 2 presents the operational definitions of all the dimensions and subdimensions of BDAQ. All the scales to measure BDAQ were drawn from prior literature and adapted to suit the context. The study measured all the first-order constructs using 7 point Likert scale (i.e., strongly disagree - strongly agree) except satisfaction, which was measured using a 7 point semantic differential scale (i.e., very dissatisfied -very satisfied). Data were collected from the 302 big data analysts in the US and France using a leading market research firm. Specifically, the sample includes 150 valid responses from the France and 152 from the U.S.

INSERT TABLE 2 HERE

4.2. Data Analysis

The study applied PLS-SEM to estimate a hierarchical, reflective-formative BDAQ construct in order to avoid the limitations of CBSEM regarding improper solutions or empirical under identification (Chin 2010; Wetzels 2009). Due to the soft modeling assumptions, application of PLS-SEM helps in avoiding positively-biased model fit indices for our large-complex model (Chin and Newsted 1999; Hair et al. 2012; Hair, Ringle, and Sarstedt 2011), which represents 13 latent constructs (i.e., 8 first-order + 2 second-order + 1 third-order + 2 outcome constructs) and 82 items (24+24+24+6+4). Indeed, this is a challenging context for CBSEM based studies "due to the algorithmic nature requiring inverting of matrices" (Chin 2010, p.661). Therefore, the study favored PLS-SEM to remove the uncertainty of inadmissible solutions for a large, complex model both in exploratory and confirmatory settings (Hair, Ringle, and Sarstedt 2011; Hulland, Ryan, and Rayner 2010). In this context, Chin (2010, , p.660) states that "[i]t should not be construed that PLS is not appropriate in a confirmatory sense nor in well researched domains". The study uses the approach of repeated indicators suggested by Wold (cf.Lohmöller 1989, p. 130-133), Akter, D'Ambra, and Ray (2011) and Becker, Klein, and Wetzels (2012) in estimating the hierarchical BDAQ model.

5. Findings

5.1. The Measurement Model

The study used SmartPLS 3.0 (Ringle, Wende, and Becker 2015) to estimate the measurement properties of the complex, hierarchical BDAQ model. Specifically, the study applied nonparametric bootstrapping (Efron and Tibshirani 1993; Chin 2010) with 5000 replications to obtain the standard errors of the estimates (Hair et al. 2013) and a path weighting scheme for the inside approximation.

Table 3 presents measurement properties of the first-order model in order to examine reliability, convergent validity and discriminant validity. The key psychometric properties including loadings of manifest variables, Cronbach's alphas, composite reliabilities (CRs) and average variance extracted (AVEs) have confirmed scale reliability (Chin 2010) by successfully meeting the threshold of 0.7, 0.7, 0.8 and 0.5 respectively. The convergent validity was ensured as all the items load much higher on their corresponding constructs than on other constructs. The study also calculated the square root of the AVE in the Table 4 to ensure discriminant validity (Fornell and Larcker 1981). As such, the findings of the measurement model provided adequate evidence of reliability, convergent validity, and discriminant validity. These findings provide the confidence to confirm all the hypothesized relationships of the structural model.

INSERT TABLE 3 HERE

INSERT TABLE 4 HERE

In Table 5, this study shows the findings of the complex-higher-order BDAQ model. The study estimated the third-order BDAQ construct, which consists of 2 second-order formative constructs (technology quality and information quality) representing 24 (3+3+3+3+3+3+3+3) valid items. Since both the second and third-order constructs are formative, thus, we estimated the weights of items of higher-order constructs which are significant at $p < 0.05$. The findings show minimum evidence of collinearity as the variance inflation factor (VIF) of all items was less than 5.

Table 5 shows that the degree of explained variance of the third order BDAQ construct is explained by second-order technology quality (58%) and information quality (49%). Accordingly, second-order constructs are explained by its first-order dimensions, such as information quality is explained by completeness (27%), currency (37%), format (26%) and accuracy (24%). The findings ensure that all the paths are significant at $p < 0.001$ both at the first-order and higher-order level. The study analyzes the implications of these results in the discussion section.

INSERT TABLE 5 HERE

5.2. The Structural Model

This study confirms the nomological validity of BDAQ model by examining its relationship with BVAL and BDAS. In order to assess the nomological validity, the study uses BVAL and BDAS with the hierarchical BDAQ construct. In the main effects model (Figure 2), the findings provide a standardized beta of 0.797, 0.175 and 0.681 respectively from BDAQ-BVAL, BVAL-BDAS and BDAQ-BDAS. Path coefficients are significant at $p < 0.05$, thus support H1, H2 and H3 (see Table 6). These results also confirm the significance of BVAL as a partial mediator between BDAQ and BDAS, which explains about 17% (0.797*0.175 /0.797*0.175+0.681) of the total effect of BDAQ on BDAS.

INSERT FIGURE 2 HERE

INSERT TABLE 6 HERE

5.3. An assessment of the PLS-SEM based complex model

This study applied PLS-SEM in estimating the complex, hierarchical research model with mediating effects. Although PLS-SEM successfully validated the research model, this study investigated the significance of model fit, predictive relevance and unobserved heterogeneity to establish further rigor. Model fit is essential to establish conjectures (Tenenhaus et al. 2005; Henseler, Hubona, and Ray 2016), predictive relevance (Q^2) is critical to check the extent of reproduction of observed values and finally, unobserved heterogeneity is important for identifying significant heterogeneity in data which can lead to bias parameter estimates and invalid statistical conclusions (Esposito Vinzi et al. 2008). The findings show that the study achieved an adequate GoF value (> 0.36), standardized root mean square residual (SRMR) (< 0.080) and Q^2 (> 0.50). Furthermore, to determine unobserved heterogeneity, we applied REBUS-PLS which detects two equal size groups in the sample (US vs. French). However, only few slightly significant differences have been observed between model parameters applying to the two detected groups. The presence of unobserved heterogeneity in the data has to be discarded in our case. This may due to the small sample size (i.e., n=151) for each of the detected groups.

6. Discussion

The study answered the key question posed by the research whether PLS-SEM can estimate a complex model. The findings illustrate that PLS-SEM entail the flexibility of soft modelling assumptions in validating a reflective-formative, hierarchical quality model in big data analytics research. According to Jacoby (1978, p.91), "we live in a complex, multivariate world [and that] studying the impact of one or two variables in isolation, would seem……relatively artificial and inconsequential". Thus, there is huge possibility of PLS-SEM based research in complex, predictive settings, such as the big data environment (Barclay, Higgins, and Thompson 1995; Hulland 1999; Lohmöller 1989; Lohmoller 1988; Wold 1980, 1985; Chin 1998a, 2010; Dijkstra 2010).

This study explained in detail the methodological gestalt of complex modeling using PLS-SEM in order to demonstrate why this study is a leap forward. Since the soft modeling assumptions of PLS-SEM facilitate developing complex models both in theoretical and applied research contexts, it has immense potential to capture the complexity of causal modeling. Indeed, PLS-SEM is best suited for complex models especially when the primary objective is prediction, the focus is on explaining variance of large number of variables and the sample size is small (Hulland, Ryan, and Rayner 2010). For example, the complex model in our study has robustly explained variances of 13 latent variables, 82 (24+24+24+6+4) indicators with 302 samples. The application of PLS-SEM makes it possible to extend the theoretical and managerial contributions of the study. Theoretically, the study contributes in several ways. First, the study offers a conceptual framework integrating RBT and BDAQ in order to provide a theoretical synergy to work with big data against the backdrop of analytics studies that show mixed results in business value creation. Second, extending the RBT, the study proposes a quality dominant logic in BDA research with two dimensions (i.e., technology quality and information quality) and eight sub-dimensions (i.e. system reliability, system adaptability, system integration and system privacy as sub-dimensions of technology quality, and completeness, accuracy, format and currency as sub-dimensions of information quality). Third, the study has identified a full, yet tightly entangled, set of dimensions that help predict the quality of BDA and their effects on business value and satisfaction. Finally, the research presents rigour by conceptualizing all the dimensions, developing their scales and estimating the mediating effects of BVAL on BDAQ-BDAS link. Highlighting the importance of mediating effects, Iacobucci (2009, p.673) states "If mediation clarifies the conceptual picture somewhat, with the insertion of just one new construct— the mediator—imagine how much richer the theorizing might be if researchers tried to formulate and test even more complex nomological networks". Practically, the proposed BDAQ model presents practitioners with an instrument for investigating a holistic quality analysis and design of analytics. The results highlight that only having a sound technology platform is not adequate to ensure solid insight from analytics platform. Although firms invest lot of resources to improve analytics platform, sophisticated insights explaining how BDA platform can improve BVAL and BDAS deserve equal attention. The findings clearly show how to tap into BDAQ to influence business outcomes. Practitioners now can have a coordinated focus to ensure the simultaneous quality of technology and information. Overall, these findings provide the blueprint to identify and improve a specific quality dimension of big data analytics at different levles.

7. Limitations and Future Research Directions

Although PLS-SEM is a preferred technique for complex modeling in social science and business research, there are few challenges that need to be addressed in order to establish it as an esoteric method. For example, first, PLS-SEM should have the flexibility of imposing constraints on model coefficients (weights, loadings, path coefficients) in order to specify any information or conjectures available a priori in estimating model parameters (Vinzi et al. 2010). Second, this approach should allow specific treatment of categorical variables, outliers, non-linearity and mutual causality both in measurement and structural models, which can lead toward estimation of interaction and quadratic effects. Third, application of PLS methods should clarify both observed heterogeneity (Sarstedt, Jörg, and Christian 2011) and unobserved heterogeneity departing from the assumption that all individuals act in a similar fashion (Becker et al. 2013). Fourth, PLS studies should pay serious attention to an adequate statistical power and representativeness of data in the context of inferential statistics using small sample size (Marcoulides, Chin, and Saunders 2009). Fifth, PLS models can take into account feedback loops and model fit indices in order to leverage its application as an SEM tool (Henseler, Hubona, and Ray 2016). Finally, future research can explore non-linear effects, parameter bias under Mode A and B, population type and data conditions in the domain of complex modeling (Sarstedt et al. 2016). The limitations mentioned in the study represent exciting avenues for PLS-SEM researchers to establish it as a powerful platform for complex modeling. Drawing on the arguments of Hair et al. (2012; 2011) and Reinartz et al. (2009), Ringle et al. (2012b, p.vii) state that "… PLS-SEM can indeed be a "silver bullet" in certain research situations (e.g., when models are relatively complex and representative sets of data are rather small". In addition to the analysis tool, future research can evaluate the stability of the research model by using objective measures, collecting longitudinal data and recruiting large samples across various industries. Theoretically, the dimensions of the BDAQ model could be extended by adding talent quality due to the ability of data scientists in generating meaningful insights and gaining competitive advantages. Since BDA is transforming operations and enhancing firm performance, future research can also investigate the impact of analytics culture in achieving business outcomes.

8. Conclusion

Overall, our study of complex modeling answers a salient question raised by Chin (2010, p.645), *"...whether the goal is to explain the covariances of a relatively small set of measured items based on a few underlying latent constructs or to focus on the complex interrelationships among a large set of factors that more closely mirrors the study context".* Drawing on the philosophy of verisimilitude (Rozeboom 2005; Meehl 1990), we propose that PLS-SEM may prove highly useful in developing and validating complex models especially when the focus is on embracing completeness (Meehl 1990), capturing reality (Cudeck and Henly 2003) or, reflecting the true parameters in the study. Therefore, the study concludes that PLS-SEM is a modest and realistic technique to establish rigor in complex modeling, which reflects Wold's (1982) viewpoint: *"There is nothing vague or fuzzy about soft modeling; the technical argument is entirely rigorous".*

7. References

Akter, Shahriar, John D'Ambra, and Pradeep Ray. 2011. "Trustworthiness in mHealth information services: an assessment of a hierarchical model with mediating and moderating effects using partial least squares (PLS)." Review of. *Journal of the American Society for Information Science and Technology* 62 (1):100-16.

Akter, Shahriar, John D'Ambra, and Pradeep Ray. 2010. "Service quality of mHealth platforms: development and validation of a hierarchical model using PLS." Review of. *Electronic Markets* 20 (3-4):209-27.

Akter, Shahriar, and Samuel Fosso Wamba. 2016. "Big data analytics in E-commerce: a systematic review and agenda for future research." Review of. *Electronic Markets*:1-22. doi: 10.1007/s12525-016-0219-0.

Akter, Shahriar, Samuel Fosso Wamba, Angappa Gunasekaran, Rameshwar Dubey, and Stephen J. Childe. 2016. "How to improve firm performance using big data analytics capability and business strategy alignment?" Review of. *International Journal of Production Economics* 182:113-31. doi: http://dx.doi.org/10.1016/j.ijpe.2016.08.018.

Bagozzi, R. P. 1980. *Causal models in marketing*: Wiley New York.

Barclay, Donald, Christopher Higgins, and Ronald Thompson. 1995. "The partial least squares (PLS) approach to causal modeling: personal computer adoption and use as an illustration." Review of. *Technology Studies* 2 (2):285-309.

Barendse, MT, FJ Oort, and G JA Garst. 2010. "Using restricted factor analysis with latent moderated structures to detect uniform and nonuniform measurement bias; a simulation study." Review of. *AStA Advances in Statistical Analysis* 94 (2):117-27.

Baron, R. M., and D. A. Kenny. 1986. "The moderator-mediator variable distinction in social psychological research: conceptual, strategic and statistical considerations." Review of. *Journal of personality and social psychology* 51 (6):1173–82.

Barton, D, and D Court. 2012. "Making advanced analytics work for you." Review of. *Harvard business review* 90 (10):78.

Becker, Jan-Michael, Kristina Klein, and Martin Wetzels. 2012. "Hierarchical latent variable models in PLS-SEM: guidelines for using reflective-formative type models." Review of. *Long Range Planning* 45 (5):359-94.

Becker, Jan-Michael, Arun Rai, Christian M Ringle, and Franziska Völckner. 2013. "Discovering unobserved heterogeneity in structural equation models to avert validity threats." Review of. *MIS Quarterly* 37 (3):665-94.

Bentler, Peter M., and Wenjing Huang. 2014. "On Components, Latent Variables, PLS and Simple Methods: Reactions to Rigdon's Rethinking of PLS." Review of. *Long Range Planning* 47 (3):138-45. doi: http://dx.doi.org/10.1016/j.lrp.2014.02.005.

Blalock, Hubert M. 1964. "Causal inferences in nonexperimental research." Review of.

———. 1979. "The presidential address: Measurement and conceptualization problems: The major obstacle to integrating theory and research." Review of. *American Sociological Review* 44 (6):881-94.

Chin, Wynne W. 1998a. "Commentary: Issues and opinion on structural equation modeling." Review of. *MIS Quarterly*.

———. 1998b. "The partial least squares approach for structural equation modeling." Review of.

———. 2010. "How to write up and report PLS analyses." Review of. *Handbook of partial least squares*:655-90.

Chin, Wynne W, and Abhijit Gopal. 1995. "Adoption intention in GSS: relative importance of beliefs." Review of. *ACM SigMIS Database* 26 (2-3):42-64.

Chin, Wynne W, Barbara L Marcolin, and Peter R Newsted. 2003. "A partial least squares latent variable modeling approach for measuring interaction effects: Results from a Monte Carlo simulation study and an electronic-mail emotion/adoption study." Review of. *Information Systems Research* 14 (2):189-217.

Chin, Wynne W, and Peter R Newsted. 1999. "Structural equation modeling analysis with small samples using partial least squares." Review of. *Statistical strategies for small sample research* 2:307-42.

Chin, Wynne W, Robert A Peterson, and Steven P Brown. 2008. "Structural equation modeling in marketing: some practical reminders." Review of. *The Journal of Marketing Theory and Practice* 16 (4):287-98.

Cudeck, Robert, and Susan J Henly. 2003. "A realistic perspective on pattern representation in growth data: comment on Bauer and Curran (2003)." Review of.

Davenport, T. H. 2006. "Competing on Analytics." Review of. *Harvard business review* 84 (1):98-107.

Davenport, Thomas H, Paul Barth, and Randy Bean. 2012. "How 'Big Data'is Different." Review of. *MIT Sloan Management Review* 54 (1):43-6.

Davenport, Thomas H, and Jeanne G Harris. 2007. *Competing on analytics: the new science of winning*: Harvard Business School Press.

Dijkstra, Theo K. 2010. "Latent variables and indices: Herman Wold's basic design and partial least squares." In *Handbook of partial least squares*, 23-46. Springer.

Dijkstra, Theo K, and Jörg Henseler. 2015a. "Consistent and asymptotically normal PLS estimators for linear structural equations." Review of. *Computational statistics & data analysis* 81:10-23.

———. 2015b. "Consistent partial least squares path modeling." Review of. *MIS quarterly= Management information systems quarterly* 39 (2):297-316.

Edwards, Jeffrey R. 2001. "Multidimensional constructs in organizational behavior research: An integrative analytical framework." Review of. *Organizational Research Methods* 4 (2):144-92.

Efron, Bradley, and Robert Tibshirani. 1993. *An introduction to the bootstrap*. Vol. 57: Chapman & Hall/CRC.

Esposito Vinzi, Vincenzo, Laura Trinchera, Silvia Squillacciotti, and Michel Tenenhaus. 2008. "REBUS-PLS: A response-based procedure for detecting unit segments in PLS path modelling." Review of. *Applied Stochastic Models in Business and Industry* 24 (5):439-58.

Fassott, Georg, Jörg Henseler, and Pedro S. Coelho. 2016. "Testing moderating effects in PLS path models with composite variables." Review of. *Industrial Management & Data Systems* 116 (9):1887-900. doi: 10.1108/imds-06-2016-0248.

Fornell, Claes, and Fred L Bookstein. 1982. "Two structural equation models: LISREL and PLS applied to consumer exit-voice theory." Review of. *Journal of Marketing Research*:440-52.

Fornell, Claes, and David F Larcker. 1981. "Evaluating structural equation models with unobservable variables and measurement error." Review of. *Journal of Marketing Research*:39-50.

Fosso Wamba, Samuel, Shahriar Akter, Andrew Edwards, Geoffrey Chopin, and Denis Gnanzou. 2015. "How ‘big data’ can make big impact: Findings from a systematic review and a longitudinal case study." Review of. *International Journal of Production Economics* 165 (0):234-46. doi: http://dx.doi.org/10.1016/j.ijpe.2014.12.031.

Gefen, David, Detmar W Straub, and Edward E Rigdon. 2011. "An update and extension to SEM guidelines for admnistrative and social science research." Review of. *Management Information Systems Quarterly* 35 (2):iii-xiv.

George, Gerard, Martine R Haas, and Alex Pentland. 2014. "Big data and management." Review of. *Academy of Management Journal* 57 (2):321-6.

Golder, Peter N, Debanjan Mitra, and Christine Moorman. 2012. "What is quality? an integrative framework of processes and states." Review of. *Journal of marketing* 76 (4):1-23.

Guide Jr, V. Daniel R., and Mikko Ketokivi. 2015. "Notes from the Editors: Redefining some methodological criteria for the journal." Review of. *Journal of Operations Management* 37:v-viii. doi: http://dx.doi.org/10.1016/S0272-6963(15)00056-X.

Hair, Joe F, G Tomas M Hult, Christian Ringle, and Marko Sarstedt. 2013. *A primer on partial least squares structural equation modeling (PLS-SEM)*: SAGE Publications, Incorporated.

Hair, Joe F, Christian M Ringle, and Marko Sarstedt. 2011. "PLS-SEM: Indeed a silver bullet." Review of. *The Journal of Marketing Theory and Practice* 19 (2):139-52.

Hair, Joe F, Marko Sarstedt, Christian M Ringle, and Jeannette A Mena. 2012. "An assessment of the use of partial least squares structural equation modeling in marketing research." Review of. *Journal of the Academy of Marketing Science* 40 (3):414-33.

Hair, Joseph F, Christian M Ringle, and Marko Sarstedt. 2012. "Editorial-Partial Least Squares: The Better Approach to Structural Equation Modeling?" Review of. *Long Range Planning* 45 (5-6):312-9.

Henseler, Jörg, Geoffrey Hubona, and Pauline Ash Ray. 2016. "Using PLS path modeling in new technology research: updated guidelines." Review of. *Industrial Management & Data Systems* 116 (1):2-20.

Henseler, Jörg, Christian M Ringle, and Marko Sarstedt. 2015. "A new criterion for assessing discriminant validity in variance-based structural equation modeling." Review of. *Journal of the Academy of Marketing Science* 43 (1):115-35.

Henseler, Jörg, Christian Ringle, and Rudolf Sinkovics. 2009. "The use of partial least squares path modeling in international marketing." Review of. *Advances in International Marketing (AIM)* 20:277-320.

Huber, Joel, and John McCann. 1982. "The impact of inferential beliefs on product evaluations." Review of. *Journal of Marketing Research*:324-33.

Hulland, John. 1999. "Use of partial least squares (PLS) in strategic management research: a review of four recent studies." Review of. *Strategic management journal* 20 (2):195-204.

Hulland, John, Michael J Ryan, and Robert K Rayner. 2010. "Modeling customer satisfaction: a comparative performance evaluation of covariance structure analysis versus partial least squares." In *Handbook of partial least squares*, 307-25. Springer.

Iacobucci, Dawn. 2009. "Everything you always wanted to know about SEM (structural equations modeling) but were afraid to ask." Review of. *Journal of Consumer Psychology* 19 (4):673-80.

Iacobucci, Dawn, and Nigel Hopkins. 1992. "Modeling dyadic interactions and networks in marketing." Review of. *Journal of Marketing Research*:5-17.

Jacoby, Jacob. 1978. "Consumer research: a state of the art review." Review of. *the Journal of Marketing*:87-96.

Jarvis, Cheryl Burke, Scott B MacKenzie, and Philip M Podsakoff. 2003. "A critical review of construct indicators and measurement model misspecification in marketing and consumer research." Review of. *Journal of consumer research* 30 (2):199-218.

Ji-fan Ren, Steven, Samuel Fosso Wamba, Shahriar Akter, Rameshwar Dubey, and Stephen J. Childe. 2016. "Modelling quality dynamics, business value and firm performance in a big data analytics environment." Review of. *International Journal of Production Research*:1-16. doi: 10.1080/00207543.2016.1154209.

Kiron, David, Pamela Kirk Prentice, and Renee Boucher Ferguson. 2014. "The analytics mandate." Review of. *MIT Sloan Management Review* 55 (4):1-25.

Lavalle, S., E. Lesser, R. Shockley, M. S. Hopkins, and N. Kruschwitz. 2011. "Big Data, Analytics and the Path From Insights to Value." Review of. *MIT Sloan Management Review* 52 (2):21-32.

Law, Kenneth S, and Chi-Sum Wong. 1999. "Multidimensional constructs M structural equation analysis: An illustration using the job perception and job satisfaction constructs." Review of. *Journal of Management* 25 (2):143-60.

Lohmoller, Jan-Bernd. 1988. "The PLS program system: Latent variables path analysis with partial least squares estimation." Review of. *Multivariate Behavioral Research* 23 (1):125-7.

Lohmöller, Jan-Bernd. 1989. *Latent variable path modeling with partial least squares*: Physica-Verlag Heidelberg.

Lu, Irene RR, Ernest Kwan, D Roland Thomas, and Marzena Cedzynski. 2011. "Two new methods for estimating structural equation models: An illustration and a comparison with two established methods." Review of. *International Journal of Research in Marketing* 28 (3):258-68.

MacCallum, Robert C, and James T Austin. 2000. "Applications of structural equation modeling in psychological research." Review of. *Annual review of psychology* 51 (1):201-26.

MacKenzie, Scott B, Philip M Podsakoff, and Cheryl Burke Jarvis. 2005. "The problem of measurement model misspecification in behavioral and organizational research and some recommended solutions." Review of. *Journal of Applied Psychology* 90 (4):710-29.

Marcoulides, George A, Wynne W Chin, and Carol Saunders. 2009. "A critical look at partial least squares modeling." Review of. *MIS Quarterly* 33 (1):171-5.

Marsh, Herbert W, Kit-Tai Hau, and Zhonglin Wen. 2004. "In search of golden rules: Comment on hypothesis-testing approaches to setting cutoff values for fit indexes and dangers in overgeneralizing Hu and Bentler's (1999) findings." Review of. *Structural Equation Modeling* 11 (3):320-41.

McAfee, A., and E. Brynjolfsson. 2012a. "Big Data: The Management Revolution." Review of. *Harvard business review* 90 (10):60-8.

McAfee, Andrew, and Erik Brynjolfsson. 2012b. "Big data: the management revolution." Review of. *Harvard business review* (90):60-6, 8, 128.

McDonald, Roderick P. 1996. "Path analysis with composite variables." Review of. *Multivariate Behavioral Research* 31 (2):239-70.

Meehl, Paul E. 1990. "Appraising and amending theories: The strategy of Lakatosian defense and two principles that warrant it." Review of. *Psychological Inquiry* 1 (2):108-41.

Mill, John Stuart. 1865. *Auguste Comte and Positivism*: N Trubner & Co.

Nelson, R Ryan, Peter A Todd, and Barbara H Wixom. 2005. "Antecedents of information and system quality: an empirical examination within the context of data warehousing." Review of. *Journal of Management Information Systems* 21 (4):199-235.

Netemeyer, Richard G, William O Bearden, and Subhash Sharma. 2003. *Scaling procedures: Issues and applications*: SAGE Publications, Incorporated.

Pyke, Steve. 2011. *Philosophers*: Oxford University Press.

Ransbotham, Sam, David Kiron, and Pamela Kirk Prentice. 2016. "Beyond the Hype: The Hard Work Behind Analytics Success." Review of. *MIT Sloan Management Review* 57 (3).

Ray, Gautam, Waleed A Muhanna, and Jay B Barney. 2005. "Information technology and the performance of the customer service process: A resource-based analysis." Review of. *MIS Quarterly*:625-52.

Reinartz, Werner, Michael Haenlein, and Jörg Henseler. 2009. "An empirical comparison of the efficacy of covariance-based and variance-based SEM." Review of. *International Journal of Research in Marketing* 26 (4):332-44.

Rigdon, Edward E. 2014. "Rethinking partial least squares path modeling: breaking chains and forging ahead." Review of. *Long Range Planning* 47 (3):161-7.

Ringle, Christian M, Marko Sarstedt, and Detmar W Straub. 2012a. "Editor's comments: a critical look at the use of PLS-SEM in MIS quarterly." Review of. *MIS Quarterly* 36 (1):iii-xiv.

Ringle, Christian M, Sven Wende, and Alexander Will. 2005. "SmartPLS 2.0 (beta)." In.: Hamburg, Germany.

Ringle, Christian, Marko Sarstedt, and Detmar Straub. 2012b. "A critical look at the use of PLS-SEM in MIS quarterly." Review of. *MIS Quarterly (MISQ)* 36 (1).

Ringle, CM, S Wende, and JM Becker. 2015. "Smart PLS 3. Boenningstedt: SmartPLS GmbH." In.

Rönkkö, Mikko, Cameron N McIntosh, John Antonakis, and Jeffrey R Edwards. 2016. "Partial least squares path modeling: Time for some serious second thoughts." Review of. *Journal of Operations Management*.

Rozeboom, Wm W. 2005. "Meehl on metatheory." Review of. *Journal of Clinical Psychology* 61 (10):1317-54.

Sanders, Nada R. 2016. "How to Use Big Data to Drive Your Supply Chain." Review of. *California Management Review* 58 (3):26-48.

Sarstedt, Marko, Joseph F Hair, Christian M Ringle, Kai O Thiele, and Siegfried P Gudergan. 2016. "Estimation issues with PLS and CBSEM: Where the bias lies!" Review of. *Journal of Business Research* 69 (10):3998-4010.

Sarstedt, Marko, Henseler Jörg, and M. Ringle Christian. 2011. "Multigroup Analysis in Partial Least Squares (PLS) Path Modeling: Alternative Methods and Empirical Results." In *Measurement and Research Methods in International Marketing*, 195-218. Emerald Group Publishing Limited.

Sarstedt, Marko, Christian M. Ringle, and Joseph F. Hair. 2014. "PLS-SEM: Looking Back and Moving Forward." Review of. *Long Range Planning* 47 (3):132-7. doi: http://dx.doi.org/10.1016/j.lrp.2014.02.008.

Sawyer, Alan G, and J Paul Peter. 1983. "The significance of statistical significance tests in marketing research." Review of. *Journal of Marketing Research*:122-33.

Shah, Rachna, and Susan Meyer Goldstein. 2006. "Use of structural equation modeling in operations management research: Looking back and forward." Review of. *Journal of Operations Management* 24 (2):148-69.

Stafford, T. F. 2011. "Special Research Commentary Series on Advanced Methodological Thinking for Quantitative Research." Review of. *Management Information Systems Quarterly* 35 (2):xv-xvi.

Tenenhaus, Arthur, and Michel Tenenhaus. 2011. "Regularized generalized canonical correlation analysis." Review of. *Psychometrika* 76 (2):257-84.

Tenenhaus, Michel. 2008. "Component-based structural equation modelling." Review of. *Total Quality Management* 19 (7-8):871-86.

Tenenhaus, Michel, Vincenzo Esposito Vinzi, Yves-Marie Chatelin, and Carlo Lauro. 2005. "PLS path modeling." Review of. *Computational statistics & data analysis* 48 (1):159-205.

Vinzi, Vincenzo Esposito, Wynne W Chin, Jörg Henseler, and Huiwen Wang. 2010. *Handbook of partial least squares: Concepts, methods and applications*: Springer.

Wetzels, Martin, Gaby Odekerken-Schroder, and Claudia Van Oppen. 2009. "Using PLS path modeling for assessing hierarchical construct models: guidelines and empirical illustration." Review of. *MIS Quarterly* 33 (1):177.

Wixom, B. H., B. Yen, and M. Relich. 2013. "Maximizing value from business analytics." Review of. *MIS Quarterly Executive* 12:111-23.

Wold, Herman. 1975. "Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach." Review of. *Perspectives in Probability and Statistics, In Honor of MS Bartlett*:117-44.

———. 1980. "Model construction and evaluation when theoretical knowledge is scarce." In *Evaluation of econometric models*, 47-74. Academic Press.

———. 1982. "Models for knowledge." In *The making of statisticians*, 189-212. Springer.

———. 1985. "Partial least squares." Review of. *Encyclopedia of statistical sciences*.

List of Tables

Table 1. Model parameters using PLS-SEM in PPC (2010-2015)

Studies	Sample size	Constructs	Items	Context
(Clegg, Gholami, and Omurgonulsen 2013)	183	5	30	Quality management and performance
(Ren et al. 2015)	110	6	22	Inter-organisational value co-creation in supply chain
(Lee et al. 2015)	119	3	16	Greening the supplier, environmental performance and competitive advantage.
(Kotzab et al. 2015)	274	4	38	Supply chain management
(Ahmad and Mehmood 2016)	288	9	62	Enterprise systems and performance of future city logistics
Average	194.8	5.4	33.6	

Table 2. Construct and definitions

Construct and definitions	**Sources**
BDA technology quality is defined as systems reliability, system adaptability, system integration, and system privacy. *System reliability* refers to the degree to which the BDA is reliable over time; *System adaptability* refers to degree to which the BDA can adapt to a variety of user needs and changing conditions; *system integration* refers to the ability to integrate various sources of data to produce meaningful insights; and finally, *system privacy* refers to the degree to which the BDA system is safe and protects user information.	(Nelson, Todd, and Wixom 2005); (Parasuraman, Zeithaml, and Berry 2005)
BDA Information quality is defined as the *completeness*, *accuracy*, *format*, and *currency* of information produced by BDA. *Completeness* indicates the extent to which the user perceives that BDA provide all the necessary information; *accuracy* focuses on the perceived correctness of information; *format* refers to the perception of how well the information is presented; and, finally, *currency* refers to the user's perception of the extent to which the information is up to date.	(Wixom and Todd 2005)
BDA Business value is defined as the strategic value refers to the degree of perceived benefits to the organization at a strategic level, e.g., competitive advantage.	(Gregor et al. 2006)
BDA satisfaction refers to the users' affect with (or, feelings about) BDAQ.	(Spreng, MacKenzie, and Olshavsky 1996),

Table 3. Psychometric Properties for First-order Constructs

Dimensions	Subdimensions	Items	Loadings	Alpha	CR	AVE
Technology Quality	System Reliability	The system operates reliably for the analytics. The system performs reliably for the analytics. The operation of the system is dependable for the analytics.	0.928 0.933 0.935	0.952	0. 952	0.868
	System Adaptability	The system can be adapted to meet a variety of analytics needs. The system can flexibly adjust to new demands or conditions during analytics. The system is flexible in addressing needs as they arise during the analytics.	0.907 0.918 0.897	0.933	0.933	.823
	System Integration	The system effectively integrates data from different areas of the company. The system pulls together data that used to come from different places in the company. The system effectively combines different types of data from all areas of the company.	0.923 0.908 0.938	0.945	0.945	0.852
	System Privacy	The system protects information about personal issues. This system protects information about personal identity. The system offers a meaningful guarantee that it will not share private information.	0.912 0.942 0.926	0.948	0. 984	0.859
Information quality	Completeness	The business analytics used: ____ provides a complete set of information. ____ produces comprehensive information. ____ provides all the information needed.	0.885 0.895 0.832	0.903	0.904	0.759
	Currency	____ provides the most recent information. ____ produces the most current information. ____ always provides up-to-date information.	0.919 0.776 0.883	0.932	0.932	0.821
	Format	The information provided by the analytics is ____ well formatted. The information provided by the analytics is ____ well laid out. The information provided by the analytics is ____ clearly presented on the screen.	0.936 0.933 0.928	0.952	0.952	0.869
	Accuracy	The business analytics used: ____ produces correct information. ____ provides few errors in the information. ____ provides accurate information.	0.913 0.886 0.919	0.894	0.896	0.742
BDA satisfaction (BDAS)		I am satisfied with my use of BDA service. I am contented with my use of BDA service. I am pleased with my use of BDA service. I am delighted with my use of BDA service.	0.896 0.879 0.835 0.890	0.929	0.929	0.766
Business value (BVAL)		The BDA used by the firm: Creates competitive advantage. Aligns analytics with business strategy. Establishes useful links with other organizations. Enables quicker response to change. Improves customer relations. Provides better products or services to customers.	0.847 0.859 0.813 0.909 0.834 0.795	0.937	0.937	0.712

*items eliminated due to low factor loadings or cross loadings.

Table 4. Mean, Standard Deviation (SD) and correlations of the latent variables for the first order constructs*

Constructs	*Mean*	*SD*	*SYRE*	*SYAD*	*SYIN*	*SYPR*	*COMP*	*CURR*	*FORM*	*ACCU*	*BVAL*	*BDAS*
System Reliability (SYRE)	*4.894*	*1.008*	*0.932**									
System Adaptability (SYAD)	*4.858*	*1.145*	*0.441*	*0.907**								
System Integration (SYIN)	5.045	1.139	0.343	0.451	*0.923**							
System Privacy (SYPR)	5.138	1.167	0.503	0.563	0.426	*0.927**						
Completeness (COMP)	4.772	1.117	0.477	0.559	*0.553*	*0.565*	*0.871**					
Currency (CURR)	5.084	1.081	0.684	0.563	*0.659*	*0.523*	*0.601*	*0.906**				
Format (FORM)	5.073	1.127	0.540	0.516	0.429	0.578	*0.532*	*0.658*	*0.932**			
Accuracy (ACCU)	4.997	1.047	0.523	0.465	0.585	0.487	*0.513*	*0.664*	*0.534*	*0.861**		
Business Value (BVAL)	5.035	1.018	*0.446*	*0.441*	*0.559*	*0.518*	*0.538*	*0.621*	*0.542*	*0.510*	*0.844**	
BDA satisfaction (BDAS)	4.897	1.022	*0.558*	*0.533*	*0.381*	*0.591*	*0.590*	*0.601*	*0.537*	*0.529*	*0.414*	*0.875**

*square root of AVE on the diagonal

Table 5: Assessment of the higher-order, reflective-formative model

Third-order Formative construct	Relationships with second-order dimensions	β	t-stat
BDAQ	Technology quality Information quality	0.583 0.487	8.278 7.122
Second-order Formative constructs	**Relationships with first-order dimensions**	**β**	**t-stat**
Technology quality	System reliability System adaptability System integration System privacy	0.394 0.274 0.307 0.158	4.909 3.289 3.827 2.493
Information quality	Completeness Currency Format Accuracy	0.268 0.374 0.259 0.235	4.086 4.388 3.078 2.978

Table 6. Results of the structural model

Paths	Path coefficients	Standard error	*t* statistic
BDAQ → BVAL	**0.797**	**0.028**	**28.504**
BVAL → BDAS	**0.175**	**0.073**	**2.403**
BDAQ → BDAS	**0.681**	**0.065**	**10.454**

List of Figures

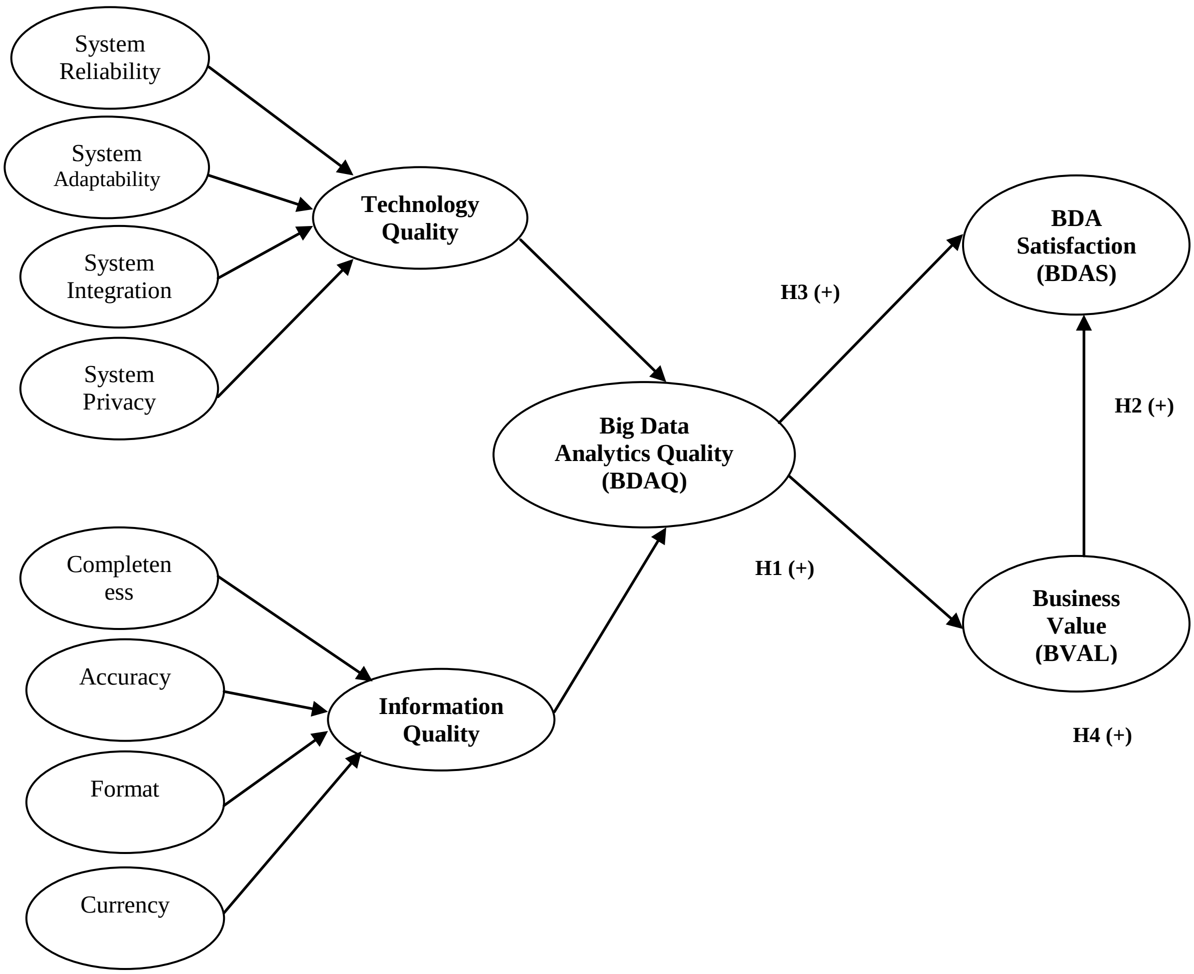

Figure 1. Big Data Analytics Quality Model

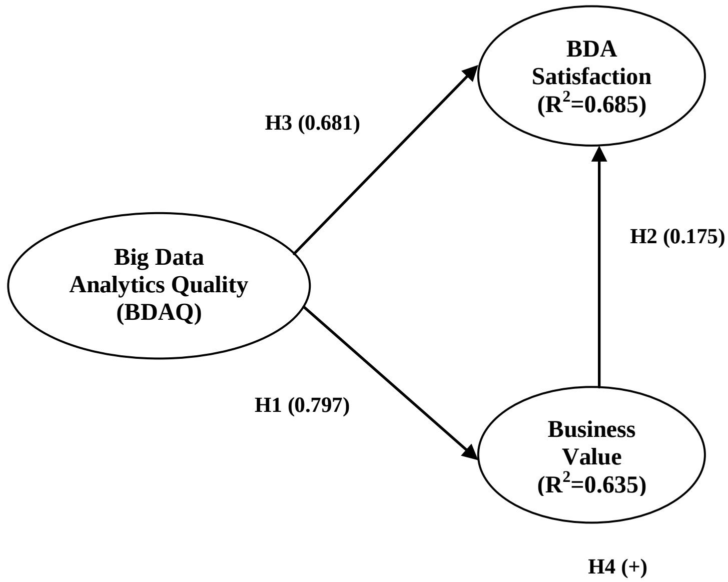

Figure 2. Structural Model