10a

MEASURING EARNINGS QUALITY: A PROPOSAL FOR A THEORETICAL FRAMEWORK AND AN EMPIRICAL METHOD

Ana Licerán-Gutiérrez
Departamento de Economía Financiera y Contabilidad
Universidad de Jaén
Paraje Las Lagunillas, s/n. E-23071.

Manuel Cano-Rodríguez
Departamento de Economía Financiera y Contabilidad
Universidad de Jaén
Paraje Las Lagunillas, s/n. E-23071.

Área temática: A.- Información financiera y normalización contable.

Keywords: Earnings dimensions; earnings quality concept; earnings quality measurement; Partial Least Squares (PLS); Structural Equation Model (SEM).

MEASURING EARNINGS QUALITY: A PROPOSAL FOR A THEORETICAL FRAMEWORK AND AN EMPIRICAL METHOD

Abstract

In this paper, we stress the importance of earnings quality conceptualization. Based on Dechow et al.'s(2010) classification of earnings quality indicators, we find that earnings properties are linked to earnings quality in a formative way, whereas investors' responsiveness and other external indicators are reflections of earnings quality. Using a simulation process, we prove that earnings-quality measurement methods in previous research are likely to measure earnings quality with substantial error, except when correlations among its components are highly positive. Estimates by second-generation regression methods (specifically, PLS method) prove to be more robust to low or negative correlations among earnings quality dimensions.

KEYWORDS: Earnings dimensions; earnings quality concept; earnings quality measurement; Partial Least Squares (PLS); Structural Equation Model (SEM).

1- INTRODUCTION

Earnings quality is arguably the most common topic in empirical research on financial accounting (Licerán-Gutiérrez and Cano-Rodríguez, 2017). Despite the vast literature on this topic, the previous research lacks a sound conceptualization of the earnings quality construct.

Conceptualization is the process by which the researcher specifies the exact meaning of the conceptual variables of the model (in our case, earnings quality) by describing the different dimensions of that variable and the indicators that can be used to measure it (Babbie, 2017) as well as the relationship between the conceptual variable and those dimensions and indicators. This description constitutes an auxiliary theory that links the theoretical model with the real world (Edwards and Bagozzi, 2000).

The lack of a clear conceptualization of earnings quality has various implications for empirical research. The first implication is the difficulty of interpreting the empirical results. Previous research has employed a large number of earnings quality proxies that represent different earnings properties. Earnings quality measures have been based on (the absence of) abnormal accruals, the existence of discontinuities in cross-sectional distribution of earnings, predictability or smoothness of reported earnings, value relevance of earnings or book values, accounting conservatism, investors' behaviour, or the opinion of external parties. This large diversity makes it very difficult to interpret the empirical results because it is not clear whether all these metrics are actually measuring the same facet of a concept, different facets of a single concept, different concepts, or whether they are substitutes or complements (Ewert and Wagenhofer, 2011).

Dechow et al. (2010) propose a classification of the empirical indicators of earnings quality intro three groups: earnings quality properties, investors' reactions to earnings quality, and other external indicators of earnings quality. Based on their analysis, the authors conclude that earnings quality is a multidimensional concept that is composed of various properties that are expected to increase the usefulness of earnings for decision making. Additionally, these properties are not substitutes, nor is one property superior to the others for all the decision models (Dechow et al., 2010). Despite this multidimensional nature of earnings quality, the vast majority of papers on earnings quality have adopted a single-dimension approach: They analyse only one of the properties of earnings quality (Licerán-Gutiérrez and Cano-Rodríguez, 2017). Moreover, the few papers that develop a multidimensional measure of earnings quality have not undertaken a conceptualization process that clarifies which dimensions are the components of the construct or the nature of the relationship between the construct and the components. Consequently, these multidimensional measures have been criticized because of the subjectivity in the selection of the proxies and their weights or because of the lack of analyses on the relationship among the different proxies (Leuz and Wysocki, 2016; Licerán-Gutiérrez and Cano-Rodríguez, 2017).

In this paper, we contribute to the research on earnings quality by proposing a formal conceptualization of earnings quality concept, analysing the nature of the relationships between the different components and earnings quality construct. Starting from Dechow et al.'s (2010) classification, we argue that the different types of indicators can be related to earnings quality construct in two different ways: Earnings properties are related to the earnings quality concept in a formative way, whereas investors' behaviour and other external indicators relate to earnings quality in a reflective way. We also discuss the implications of these two different forms of relationships for empirical research.

Finally, we propose the use of second-generation regression models for analysing the relationships between earnings quality and other conceptual variables because second-generation regression models address both the relationship among conceptual variables and their measurement problem. Using a simulation process, we compare the performance of a second-generation regression model (specifically, a partial least squares (PLS) model) with the three approaches most commonly used for measuring earnings quality: single indicator, equally weighted index of various indicators, and common factor scores from a factorial analysis of various earnings properties. Results show that PLS model outperforms the other approaches except when the correlation between earnings quality dimensions approaches +1. As previous research has shown that the correlation between earnings quality dimensions is nearly 0 and can even be negative, results indicate that PLS model is likely to produce more accurate estimates of earnings quality than traditional methods.

The remainder of the paper is structured as follows: section 2 describes our framework for measuring earnings quality. Section 3 evaluates the extant research on earnings quality, comparing it to the framework previously defined. In section 4, we define a parsimonious model of earnings quality and design a simulation process to compare estimation errors of the PLS approach with traditional methods employed in empirical research on earnings quality. Section 5 concludes the paper.

2- A FRAMEWORK FOR MEASURING EARNINGS QUALITY

In this section, we provide a general description of the framework for theory-based empirical research in social sciences. Then, we apply this framework to extant research on earnings quality, showing how extant empirical research on this topic has dealt with conceptual specification of earnings quality, operationalization of the empirical measures, and the analysis of the relationships among conceptual variables.

2.1- A general overview of the framework for theory-based empirical research in social sciences

In broad terms, the goal of theory-based empirical research in the social sciences is to test the adequacy of a theoretical model as compared to the real world. To achieve this goal, empirical researchers follow a process that covers two different levels (Bisbe et al., 2007; Babbie, 2017): the conceptual specification level and the operational level.

In the conceptual specification level, the theoretical model is defined. Theoretical models are sets of relationships among different conceptual variables that formalize the key elements of a theory (Bollen, 2002). This level starts with the conceptualization process in which the researcher identifies and specifies the exact meaning of the conceptual variables of interest.[1] It also involves the description of the different aspects of the concept, known as dimensions, as well as the indicators that will be used to measure the concept (Edwards and Bagozzi, 2000; Bisbe et al., 2007; Babbie, 2017). Subsequently, the set of relationships among the conceptual variables that form the model will be determined (Bisbe et al., 2007).

In the operational level, researchers develop specific procedures that will result in empirical observations that represent the conceptual variables in the real world (Babbie, 2017). Finally, by analysing the relationships among the empirical observations that represent the conceptual variables, researchers indirectly test the extent to which the theoretical model is consistent with the real world (Bisbe et al., 2007).

[1] We define conceptual variables as the representation of ideas or abstract concepts that researchers establish to design models (Sarstedt et al., 2016).

2.2- Conceptualization, constructs and indicators

Conceptualization is, therefore, the process by which the exact meaning of the conceptual variables is specified, describing the different dimensions of that variable and the indicators that can be used to measure it (Babbie, 2017). Consequently, the theoretical discussion in theory-based empirical research should describe not only the nature and direction of the relationships among the different conceptual variables but also the specific way in which each of these conceptual variables is measured. This description constitutes an auxiliary theory that links the theoretical model with the real world (Edwards and Bagozzi, 2000). In this sense, Kaplan (1964) differentiates between three methods of obtaining empirical measures for conceptual variables:

(1) Direct observation. When the concept can be simply and directly observed, the researcher can measure it by applying such observation.

(2) Indirect observation. This procedure is applied when the concept is not observed directly but is measured using relatively more subtle or complex methods.

(3) Finally, there are conceptual variables that cannot be observed either directly or indirectly because they are abstract concepts that do not exist in the real world. These conceptual variables are typically called “constructs” or “latent variables”, and they can be defined as theoretical creations used to describe a phenomenon of interest (Edwards and Bagozzi, 2000; Bisbe et al., 2007; Babbie, 2017). Although constructs do not exist as entities in the real world, researchers employ them because they help in the organization, communication, understanding or prediction of real things (Babbie, 2017). To operationalize these constructs, the researcher has to specify what the signs of its presence or absence in the real world are and how to capture these signs using observable variables, called indicators (Bisbe et al., 2007).

In addition to the specification of the indicators or the dimensions, conceptualization also requires that researchers determine the nature and direction of the relationships between the construct and its indicators (Edwards and Bagozzi, 2000).[2] Generally, there are two broad ways in which a construct is related to its indicators: reflective and formative measurement (Sarstedt et al., 2016).

In reflective measurement, indicators are considered error-prone manifestations of the construct in the sense that the presence or absence of that construct produces variations in the value of the indicators (Edwards, 2001). In other words, the variations in the indicators' measures are considered *consequences* of the construct. Mathematically, the relationship between a construct and any of its indicators is usually represented as follows:

$$x_j = a_j \cdot Y + \varepsilon_j, \tag{1}$$

where x_j is each of the observable variables, Y is the construct, a_j is the slope of the relationship between the indicator and the construct, and ε_j is the measurement error.

In formative measurement, on the other hand, indicators are seen as the inherent constitutive facets of the construct, and therefore, the indicators as a group jointly determine the conceptual meaning of the construct (Bisbe et al., 2007). The construct can be then modelled as a linear combination of the indicators plus an error disturbance (Bollen and Bauldry, 2011) that represents those causes of the construct that have neither been discussed in prior literature nor revealed by exploratory research (Sarstedt et al., 2016). A mathematical representation for a causal indicator formative model would be the following:

$$Y = \sum_j w_j \cdot x_j + \xi, \tag{2}$$

where Y is the construct, x_j represents the indicator, w_j is the structural coefficient of each indicator, and ξ is the error disturbance of the model. These formative indicators will have conceptual unity in the sense that they correspond to the different facets of the concept represented by the construct (Bollen and Bauldry, 2011).

The difference between reflective and formative measurement presents important implications for the operationalization process. Thus, in reflective measurement, all the

[2] Some authors believe that both specification of indicators and the description of the nature of their relationship with the construct constitute an auxiliary theory, and therefore, they do not belong to the operational level but to the conceptual level (Edwards and Bagozzi, 2000; Bisbe et al., 2007).

indicators of the same construct are expected to be highly correlated given that all of them are affected by the presence of the same construct (Edwards and Bagozzi, 2000). These indicators can be considered interchangeable, and removing a specific indicator would not alter the conceptual domain of the construct (Jarvis et al., 2003; Bisbe et al., 2007).[3] Consequently, researchers do not need to use all the available indicators because a sample of them can be sufficient for measuring the concept, as long as convergent and discriminant validity tests support its consistency.

Unlike reflective measurements, indicators in formative models are not necessarily highly correlated because they do not share the same causes.[4] More importantly, they cannot be considered interchangeable: The omission of an indicator would imply that one of the constitutive facets of the construct is left out, thereby changing the definition of the construct (Jarvis et al., 2003). Therefore, it will not be enough to use a sample of indicators to measure the concept; instead, a full census of the indicators would be required (Bisbe et al., 2007).

The determination of the reflective or formative nature of the relationships among the conceptual variable and its indicators is a key feature of conceptualization because misspecification of these relationships may have serious consequences for the conclusions that are drawn. Various papers have demonstrated that the application of a reflective (formative) measurement model to a truly formative (reflective) construct leads to inaccurate conclusions about the structural relationships among constructs (Jarvis et al., 2003; MacKenzie et al., 2005; Chang et al., 2016). To help researchers select the proper model, Jarvis et al. (2003) compile a set of decision rules for determining whether a construct is formative or reflective. Table 1 reports these rules.

TABLE 1 HERE

2.3- A proposal for earnings quality conceptualization

[3] The fewer indicators included in the model, however, the lower the reliability of the set of indicators.

[4] Given that all formative indicators influence the same construct, some correlation among them can be expected, but the model does not assume nor require it (Jarvis et al., 2003).

According to the previous framework, research on earnings quality should start with the conceptualization of the conceptual variable, that is, the specification of the exact meaning of earnings quality, its indicators, and the relationships (reflective or formative) between them.

Regarding the exact meaning of earnings quality, neither the Financial Accounting Standards Board (FASB) nor the International Accounting Standard Board (IASB) provide a formal definition of earnings quality within their conceptual frameworks. Using the terminology from the FASB's Statement of Financial Concepts No. 1, Dechow et al. (2010) define earnings quality as follows: "Higher quality earnings provide more information about the features of a firm's financial performance that are relevant to a specific decision made by a specific decision-maker". According to this definition, earnings quality would be a non-observable construct, and therefore, empirical researchers need to specify the methods to operationalize this conceptual variable.

Although the IASB and the FASB do not specify the meaning of earnings quality, they provide a list of qualitative characteristics that are expected to increase the utility of financial information (IASB, 2010). Consistent with this approach, empirical research on earnings quality has developed a wide range of proxies for representing earnings quality, including measures for persistence, predictability, smoothness, accruals quality, value relevance, and conservatism (these proxies have been surveyed, for example, in Schipper and Vincent, 2003; Dechow and Schrand, 2004; Dechow et al., 2010; Licerán-Gutiérrez and Cano-Rodríguez, 2017). Despite the large number of empirical proxies for earnings quality and the vast research on this topic, there is a lack of theoretical explanations about whether these metrics are measuring the same facet of the construct, different facets of a single construct, or different constructs, as well as whether they are substitutes or complements (Ewert and Wagenhofer, 2011). In other words, the nature and direction of the relationships between these proxies and the earnings quality construct has been scarcely studied.

Dechow et al. (2010) classify these proxies into three categories: properties of earnings, investor responsiveness to earnings, and external indicators of earnings misstatements. Next, we discuss the direction of the theoretical relationship between these three categories and earnings quality.

Properties of earnings are those characteristics that are expected to affect the usefulness of reported earnings in the decision-making process. Earnings management, earnings smoothing, earnings predictability or conservatism are typical examples of such characteristics (Licerán-Gutiérrez and Cano-Rodríguez, 2017). A common feature of these characteristics is that they are jointly determined by the fundamental performance of the company, the ability of the accounting system to measure such performance, and the manager's decisions regarding the accounting system (Dechow et al., 2010). In other words, managers make accounting choices that affect these properties, thereby increasing or decreasing the earnings quality level. According to this, Jarvis et al.'s conditions for justifying the modelling of a construct in a formative fashion prevail in the case of the properties of earnings: The direction of the relationship is from the measures to the construct (changes in the properties cause changes in earnings quality); earnings properties are not substitutable nor interchangeable, and they are not highly correlated (Dechow et al., 2010); finally, the different properties do not share the same nomological network or have the same antecedents and/or consequences. In summary, the relationship between earnings properties and earnings quality appears to be formative.

The other two categories of earnings quality proxies defined by Dechow et al. (2010) are the investor responsiveness to earnings and the external indicators of earnings misstatements. Investor responsiveness to earnings constitutes the measures of the influence of earnings on equity investors' decisions, typically representing the relationship between accounting earnings and market returns.[5] The measures included in the third category (external indicators of earnings misstatements) are indicators of the existence of problems with the quality of earnings issued by an external party. These measures are SEC Accounting and

[5] The typical proxies for investor responsiveness are earnings response coefficient (ERC) and R^2 from the earnings-return model (Dechow et al., 2010)

Auditing Enforcement Releases, restatements, and reported internal control procedure deficiencies (Dechow et al., 2010).

These two categories of earnings quality can be considered consequences of the observation of the earnings quality level by an external party (investors, the SEC, or auditors). There would be a reflective relationship, therefore, among the earnings quality construct and the investor responsiveness and external indicators measures.

In summary, earnings quality can be considered a non-observable construct. Dechow et al. (2010) identify three categories of measures that have been previously used as proxies of earnings quality (earnings properties, investors' responsiveness, and external indicators of earnings misstatements). We provide a theoretical analysis of the relationships between the measures of these three categories and the earnings quality construct. Our conclusions are graphically represented in figure 1.

FIGURE 1 HERE

Earnings properties are related to earnings quality in a formative way, whereas investors' responsiveness to earnings and external indicators of misstatements are reflections of earnings quality. This difference is relevant because, as previous research has demonstrated, applying a reflective (formative) measurement model to a truly formative (reflective) construct strongly affects the results and can lead to erroneous conclusions (Jarvis et al., 2003; MacKenzie et al., 2005; Chang et al., 2016).

Consequently, the formative approach is more suitable for measuring earnings quality through earnings properties. The use of this approach requires a census of all the dimensions that define the construct because removing one of those dimensions would alter the conceptual domain of the construct (Bisbe et al., 2007).

When measuring earnings quality using investors' responsiveness to earnings or external indicators of earnings misstatements, the researcher should adopt a reflective approach. This implies that all the indicators are reflections of the same construct, and consequently, it is not necessary to use a full census of indicators.

The next step after defining the model for measuring earnings quality is to determine the relationships between the dimensions of earnings quality and the indicators used to operationalize them. Independent of the formative (earnings properties) or reflective (investors' responsiveness or external indicators of misstatements) approaches used, we consider that the different empirical indicators that have been previously used in accounting research for each dimension of earnings quality can be considered reflections of the underlying construct.

The reflective relationship between dimensions and indicators can be observed for the reflective dimensions as well.

3- AN EVALUATION OF THE EXTANT EMPIRICAL RESEARCH ON EARNINGS QUALITY

In this section, we compare the methods that empirical researchers have commonly applied for measuring earnings quality with the framework described in the previous section.

A remarkable feature of the empirical research on earnings quality is that, despite the consensus on its multidimensional nature, most of the empirical works are based on single-indicator analyses (Licerán-Gutiérrez and Cano-Rodríguez, 2017). That is, most previous empirical works that have measured earnings quality have represented it using only one empirical proxy that is assumed to fairly represent the construct. More specifically, most empirical works on earnings quality have relied on indicators of one of the properties of earnings described by Dechow et al. (2010), particularly earnings management (Licerán-Gutiérrez and Cano-Rodríguez, 2017).

The use of a single indicator for measuring the earnings quality construct produces two problems that can lead to misspecification: conceptual misspecification and noisy measures problems.

Regarding the first problem, we have discussed the fact that the properties of earnings are related to earnings quality in a formative way. As the proper measurement of a construct using a formative model requires the use of a full census of its dimensions to measure earnings quality using earnings properties (Bisbe et al., 2007), given that the omission of any dimension can imply a relevant change in the definition of the construct (Jarvis et al., 2003), a full census of these properties is also required. Therefore, by using a single instrument, researchers are incurring a conceptual misspecification of the construct, as they are not including the other dimensions of earnings quality. The observed relationships, therefore, cannot be considered relationships between earnings quality and other variables but, at best, relationships between those other variables and the specific dimension of earnings quality that has been measured through the instrument.

On the other hand, instruments are usually empirical observations that represent the underlying dimension with an error. By using only one of these instruments, the authors are employing noisy estimations of the dimension. Moreover, as the dimension represented by the instrument is not observable, the size of the measurement error is unknown.

Although the vast majority of papers on earnings quality have used only one instrument, a few studies have tried to develop a multidimensional measure of earnings quality (Licerán-Gutiérrez and Cano-Rodríguez, 2017). These papers have followed two different approaches for developing a multidimensional measure of earnings quality. The most common approach (Bhattacharya et al., 2003; Leuz et al., 2003; Biddle and Hilary, 2006; Burgstahler et al., 2006; Doupnik, 2008; VanTendeloo and Vanstraelen, 2008; Beatty et al., 2010; Boulton et al., 2011; Gaio and Raposo, 2011; Healy et al., 2014; Jung et al., 2014) consists of the computation of various indicators for the different properties of earnings, combining them into an equal-weighted index of the ranking values of these indicators. The second approach consists of the estimation of the common factor scores from a factor analysis of several proxies of the different properties of earnings (Francis et al., 2008; Bhattacharya et al., 2012).

Despite the effort of these papers to develop multidimensional measures of earnings, none of them presents a formal theoretical discussion on how the developed measures are related to earnings quality construct. Consequently, these methods present several flaws (Leuz and Wysocki, 2016; Licerán-Gutiérrez and Cano-Rodríguez, 2017) related to the subjectivity in the selection of the proxies and their weights or the lack of analyses of the relationship among the proxies. Next, we discuss how these flaws can affect to the two aforementioned approaches.

The first approach (equal-weighted index) employs a formative measurement of earnings quality, as earnings quality is measured through the aggregation of the values of the indicators that are proxies for the various facets of the construct. In formative measurement, however, all facets must be included in the analysis, as the absence of any of them can alter the definition of the construct (Jarvis et al., 2003). Consequently, the proper implementation of this approach would require that authors use at least one indicator for each earnings quality dimension. Most of the papers that have used this approach, however, do not include indicators for all the earnings quality properties.[6] Moreover, none of the papers present any formal theoretical analysis of the expected relationships between the selected indicators and earnings quality construct. Because of this lack of analysis, indicators relating in a reflective way with the construct may be included as components of the index, thereby producing model misspecification.

A second problem with this approach is the subjectivity associated with indicator weights. The common approach is to consider that all proxies have the same weight in the index, but there is no guarantee that all dimensions have the same influence on the earnings quality construct. Additionally, as the papers do not discuss whether the proxies measure the same or different dimensions (Ewert and Wagenhofer, 2011), it is common for the authors to include a different number of indicators for each dimension, thereby over or underweighting a dimension if its number of indicators is higher or lower than the number of included indicators for the other dimensions.

Regarding the second approach (common factor scores), it would be compatible with a reflective measurement of earnings quality, as earnings quality is captured by the common variance of all indicators. This method would be adequate if the indicators are related to earnings quality in a reflective way. The empirical papers that have followed this approach, however, have used indicators that represent the different properties of earnings, which are

[6] According to Licerán-Gutiérrez and Cano-Rodríguez (2017), the only work that used proxies for all properties of earnings is Gaio and Raposo (2011).

related to earnings quality in a formative way, producing model misstatement that can lead to biased estimates (Jarvis et al., 2003; MacKenzie et al., 2005; Chang et al., 2016).

In summary, the review of the extant research on earnings quality shows the absence of a sound conceptual specification of the earnings quality construct in most of previous papers on the topic. The vast majority of the papers do not address the measurement problem of earnings quality, using only one indicator for one facet of the construct. These papers, therefore, do not take into account the influence of the other dimensions of the same construct or the interactions among them. Moreover, there is no analysis of the importance of a given dimension in the construct or the potential measurement error when representing the dimension by its specific indicator.

Although a few studies try to develop a multidimensional measure of earnings quality, the results of the implemented approaches are likely to be influenced by several limitations because of the absence of a sound conceptual specification. In other words, it is necessary to specify the dimensions that conform to the construct and the way in which those dimensions are related to the construct. By specifying these issues, problems such as the absence of some dimensions in a formative approach or the mix of dimensions that are related in both formative and reflective approaches can be avoided. Additionally, it is necessary to indicate which dimension is represented by each indicator in order to ensure that the researcher is including indicators for all the dimensions to be measured and to avoid underweighting or overweighting certain dimensions because of the number of included indicators.

Finally, previous research has relied mainly on first-generation regression models (mainly ordinary least squares). These regression models are, however, limited when it is necessary to test complicated relationships in a single statistical test (Gefen et al., 2000; Nitzl, 2016). The use of first-generation methods, therefore, provides a simplified vision of the relationships among earnings quality and its causes or consequences. These models are then too limited to consider issues such as the appropriate weights of the dimensions in the formation of earnings quality construct, the relationships among them, or the influence of the measurement error. These issues, however, can be controlled using second-generation regression models such as covariance- or variance-based structural equation models.

4- A COMPARISON OF THE PERFORMANCE OF THE METHODS EMPLOYED TO MEASURE EARNINGS QUALITY

In this section, we conduct a simulation process to compare the estimates of the different approaches that have been previously used to measure earnings quality. In addition, we also add the estimates from a second order regression method, specifically, a PLS method.[7]

4.1- Description of the simulation process

In this simulation process, we estimate the influence of a non-observable earnings quality construct (noted by *EQ*) on a dependent variable (noted as *DEPENDENT*). The earnings quality construct *EQ* is formed by two dimensions (EQ_1 and EQ_2), with a formative relationship between the construct and its dimensions. We consider three available instruments for estimating each dimension (eq_{11}, eq_{12}, and eq_{13} for EQ_1 and eq_{21}, eq_{22}, and eq_{23} for EQ_2). Figure 2 presents the relationships of the simulated model.

FIGURE 2 HERE

We first simulate the values of the two earnings dimensions EQ_1 and EQ_2. We define them as normal variables with 0 mean and standard deviation equal to 1; *r* is the value of the Pearson correlation coefficient between these two random variables. To assess the influence of the correlation between the two earnings components on the estimates of the different approaches, we repeated the simulation process for different values of parameter *r*, ranging from −0.9 to +0.9 with an increase of 0.1.

After simulating the values of the two earnings components, we computed the value of earnings quality construct according to the following equation:

[7] We prefer the PLS method – which is a variance-based structural equation model – over the covariance-based structural equation model because PLS method is particularly suitable for composite-based modeling (Bagozzi, 2011; Becker et al., 2013) and because, as previously discussed, the relationship between earnings properties and earnings quality construct can be considered formative.

$$EQ = b_1 \cdot EQ_1 + b_2 \cdot EQ_2 + \varepsilon_1.$$

EQ represents the earnings quality construct, where EQ_1 and EQ_2 are the two dimensions that form the construct. b_1 and b_2 are the parameters indicating the weight of each dimension. We assume that the researcher does not know the values of these weights, so we randomly generated the values for parameter b_1 from a uniform distribution between 0 and 1, computing parameter b_2 as $1-b_1$. The error term of the equation (ε_1) is also generated from a normal distribution with zero mean and is uncorrelated with all the other random variables. We also assume that the standard deviation of this error is not observable, so we generated a random variable distributed uniformly between 0.1 and 0.5 as the standard deviation of the measurement error.

After computing the values of *EQ*, we standardized this variable by subtracting its mean and dividing by its standard deviation to get a variable with null mean and standard deviation equal to 1.

We then computed the values of the dependent variable according to the following equation:

$$DEPENDENT = a \cdot EQ_{st} + \varepsilon_2,$$

where *DEPENDENT* denotes the values of the dependent variable; *a* is the coefficient of the linear relationship between the dependent variable and the standardized values of the earnings quality construct; the error term (ε_2) is simulated from another normal standard variable with 0 mean and independent from any other random variable. The standard deviation of this error is computed to make the standard deviation of *DEPENDENT* equal to 1. Additionally, we set the value of parameter *a* as 1.[8]

Then, for each earnings quality dimension (EQ_1 and EQ_2), we simulated three indicator variables (eq_{11}, eq_{12}, and eq_{13} as indicators of EQ_1 and eq_{21}, eq_{22}, and eq_{23} as indicators of EQ_2). We computed these indicators according to the following equation:

$$eq_{ij} = \delta_{it} \cdot EQ_i + \varepsilon_{ij}.,$$

[8] We repeated the simulation process with different values for parameter *a* (specifically, 0.5 and 0.1). Results (untabulated) were not qualitatively different from those reported.

where eq_{ij} represents each indicator; EQ_i is the earnings quality component represented by that indicator; parameter δ_{it} represents the relationship between the indicator and the component; and ε_{ij} is the error term, which is generated from a normal distribution with a zero mean and is uncorrelated with all the other variables. The standard deviation of this error is computed to make the standard deviation of the indicator equal to 1. Additionally, we assume that the researcher does not know the exact relationship between the indicators and the components (δ_{it}), so the values of δ_{it} were randomly generated from a uniform distribution between 0.1 and 0.5

We use this simulation process to compare the following four approaches to estimating earnings quality:

1. As the most extensively used approach in previous literature has been the selection of a single proxy, we selected only one of the indicators (specifically, eq_{11}) as the first earnings quality proxy.
2. The second approach is an equally weighted index, the most common approach in previous literature to multidimensional measures of earnings quality. Consistent with previous research, we computed this index as the average decile ranking of the six indicators for each observation. Given that the other earnings quality variables are distributed normally with mean of 0 and standard deviation of 1, and to avoid the potential bias effect of the metrics of the latent variables (Aguirre-Urreta and Marakas, 2012; Chang et al., 2016), we standardized the values of the index by subtracting its mean and dividing the result by its standard deviation.
3. In the third approach, we computed earnings quality as the common factor scores from the factorial analysis of the six earnings components, consistent with previous papers that used this method (Francis et al., 2008; Bhattacharya et al., 2012).
4. Finally, we estimated PLS regression, considering that earnings quality is a non-observable construct that is formatively related to two other non-observable

constructs (EQ_1 and EQ_2) and considering that the relationship between the constructs and the indicators is reflective.

For each variable, we generated 20,000 observations and estimated the coefficients for the linear relationships between the dependent variable and each of the four earnings quality proxies (single indicator, average decile ranking index, common factor scores and estimation from a PLS model). Then, we computed the error for each variable as the squared value of the difference between 1 – the actual value of parameter *a* – and the estimated coefficient. We iterated this process 1,000 times and computed the average estimation error for each earnings quality measure.

4.2- Simulation results

Table 2 reports the average quadratic errors for each earnings quality measure and for the different values of the correlation between the two components. Figure 3 graphically shows the comparison among the quadratic errors of all measures.

TABLE 2 AND FIGURE 3 HERE

Results corroborate that the correlation between the components of earnings quality construct has a key impact on the accuracy of the estimates, as the average estimation error for all the earnings quality measures, with the single exception of the PLS model estimate, is strongly influenced by the correlation between the two components.

The highest estimation errors are observed for r=–0.9, where the common factor scores and the single indicator are the two measures that exhibit the greatest errors (0.985 and 0.901, respectively), followed by the decile ranking index (0.675). The PLS model estimate also exhibits its highest estimation error when r is set at –0.9, but its magnitude is considerably smaller than that of the other variables (0.156).

As the correlation approaches 0, the errors of single indicator and decile ranking index decrease drastically. The error for common factor scores, however, remains high for correlation coefficients lower than –0.1, exhibiting the worst performance of all variables when the correlations between the components are assumed to be negative. The PLS model estimate also exhibits a reduction in its estimation error as r tends to zero, but this reduction is much less steep than that observed for the other variables, showing that this measure is much more robust to the differences in the correlation between the components of the construct.

As the correlation becomes more positive, the estimation errors of all models continue to decrease but with different rhythm. Thus, single indicator reduction is similar to that observed for the negative values of the correlation. This maintained reduction implies that this variable exhibits the smallest estimation error of all methods for correlation coefficients near +1 (0.068 for r=+0.9, 0.092 for r=+0.8), although it exhibits the highest errors for correlations between +0.1 and +0.5.

The reduction in the error for decile ranking index, however, becomes smoother than that observed for the negative correlations. Consequently, its estimation errors are only lower than single indicator for values of r between +0.1 and +0.5, and its errors are the highest for correlations higher than +0.5.

Common factor scores exhibit a large decrease in the error between r=–0.1 and r=+0.1, maintaining the reduction in the estimation error for correlations higher than +0.1 but in a smoother way. This reduction makes common factor scores the second-best option for positive correlations between 0.1 and 0.5, as this approach was outperformed only by PLS regression; for stronger correlations, common factor scores exhibit similar or smaller errors than PLS model estimate, and common factor scores are outperformed only by single indicator for correlations greater than 0.8.

Finally, errors of the PLS method remain stable for positive values of the correlation between the components. It renders the smallest estimation errors for correlations below 0.7, and although it is outperformed by both single indicator and common factor scores for correlations higher than 0.8, the difference in the magnitude of the estimation errors of these three variables is not very relevant.

In summary, results prove that methods previously used in empirical research on earnings quality (single indicator, multivariable index, and common factor scores) are likely to produce biased estimations of the relationship between the earnings quality construct and a dependent variable. The size of this bias is expected to be higher if the correlation between the earnings quality dimensions is near 0 or if it becomes negative. PLS method, however, produces estimation errors that are much less sensitive to the correlation between the two components. These errors are smaller than those of the other three methods for values of the correlation that are near 0 or negative and are higher only for highly positive values of *r.*[9]

In the specific case of earnings quality, however, research on the correlation between earnings quality properties is scant, and their results are often mixed (Licerán-Gutiérrez and Cano-Rodríguez, 2017). Dechow et al. (2010) performed an extensive analysis of the correlations between the different earnings quality indicators they reviewed, finding that, although significant statistically, the size of the correlations between the indicators was not economically significant, finding negative correlations in some cases. Additionally, Ewert and Wagenhofer (2015) studied the relationships among different earnings quality measures in a theoretical model, finding that these measures can react in opposite directions depending on the circumstances. Based on these results, it seems reasonable to expect that the correlation between earnings properties is not strong enough to favour the single indicator or common factor scores approach over the PLS model approach.

5- CONCLUDING REMARKS

Despite the abundant research on earnings quality, conceptualization of the earnings quality construct has received very little attention. In this paper, we stress the importance of this conceptualization process and the need for an analysis of how empirical measures used to represent earnings quality are expected to be related with that construct.

Based on Dechow et al. (2010), we discuss the nature of the relationship between the different indicators of earnings quality that have been employed in previous empirical research and earnings quality construct. Applying Jarvis et al.'s (Jarvis et al., 2003) decision rules, we conclude that earnings properties identified by Dechow et al. (2010) can be considered dimensions of earnings quality construct, with a formative relationship among these properties

[9] It could be argued, however, that a highly positive correlation between the two components would make the difference between them much subtler, and it could even be considered that they are not different but rather the same components.

and earnings quality, whereas investors' responsiveness to earnings and external indicators of misstatements can be considered reflections of earnings quality. The implications of this differentiation for empirical research on earnings quality are multiple. Thus, if a researcher aims to measure earnings quality through earnings properties, a formative approach should be used. The empirical implications of this type of relationship are that a full census of the dimensions that define earnings quality is necessary. If the researcher aims to measure earnings quality using investors' responsiveness to earnings or external indicators of earnings misstatements, the adopted approach should be a reflective one. In this case, a full census of indicators is not necessary, as all of them are reflections of the same construct. Finally, this classification would also be helpful for avoiding the use of reflective measures in a formative approach or formative measures in a reflective approach, as this misspecification is likely to produce important biases in the estimates.

Finally, we propose the use of second-generation regression methods to improve the estimates of the relationships between earnings quality construct and other variables. To demonstrate the suitability of such methods for empirical research on earnings quality, we have compared the performance of the three most widely used methods for measuring earnings quality (single indicator, decile ranking index and common factor scores) with the results of a second-generation regression method, namely, a PLS model. The results show that the three traditional methods produce estimation errors that are clearly higher than those generated by the PLS model, except when the correlations between the dimensions of earnings quality are near +1. The scant evidence on the correlations between the different earnings properties, however, indicates that these correlations are more likely to be near 0 or negative. The use of a PLS model, therefore, is advised to produce more accurate estimates.

REFERENCES

Aguirre-Urreta, M. I., & Marakas, G. M. (2012). Revisiting Bias Due to Construct Misspecificarion: Different Results from Considering Coefficients in Standardized Form. *MIS Quarterly*, *36*(1), 123–138.

Babbie, E. (2017). *The basics of social research (7th ed.)*. (Cengage, Ed.) (7th ed.). Boston, United States.

Bagozzi, R. (2011). Measurement and Meaning in Information Systems and Organizational Research: Methodological and Philosophical Foundations. *Management Information Systems Quarterly*, *35*(2), 261–292.

Beatty, A., Liao, S., & Weber, J. (2010). Financial Reporting Quality, Private Information, Monitoring, and the Lease-versus-Buy Decision. *The Accounting Review*, *85*(4), 1215–1238.

Becker, J. M., Rai, A., & Rigdon, E. E. (2013). Predictive Validity and Formative Measurement in Structural Equation Modeling: Embracing Practical Relevance. *Thirty Fourth International Conference on Information Systems*, 1–19.

Bhattacharya, N., Ecker, F., Olsson, P. M., & Schipper, K. (2012). Direct and mediated associations among earnings quality, information asymmetry, and the cost of equity. *The Accounting Review*, *87*(2), 449–482.

Bhattacharya, U., Daouk, H., & Welker, M. (2003). The world price of earnings opacity. *The Accounting Review*, *78*(3), 641–678.

Biddle, G. C., & Gilles Hilary. (2006). Accounting Quality and Firm-Level Capital Investment. *The Accounting Review*, *81*(5), 963–982.

Bisbe, J., Batista-Foguet, J.-M., & Chenhall, R. (2007). Defining management accounting constructs: A methodological note on the risks of conceptual misspecification. *Accounting, Organizations and Society*, *32*(7–8), 789–820.

Bollen, K. A. (2002). Latent variables in psychology and the social sciences. *Annual Review of Psychology*, *53*, 605–34.

Bollen, K. A., & Bauldry, S. (2011). Three Cs in measurement models: causal indicators, composite indicators, and covariates. *Psychological Methods*, *16*(3), 265–84.

Boulton, T. J., Smart, S. B., & Zutter, C. J. (2011). Earnings Quality and International IPO Underpricing. *The Accounting Review*, *86*(2), 483–505.

Burgstahler, D. C., Hail, L., & Leuz, C. (2006). The Importance of Reporting Incentives: Earnings Management in European Private and Public Firms. *The Accounting Review*, *81*(5), 983–1016.

Chang, W., Franke, G. R., & Lee, N. (2016). Comparing reflective and formative measures: New insights from relevant simulations. *Journal of Business Research*, *69*(8), 3177–3185.

Dechow, P., Ge, W., & Schrand, C. (2010). Understanding earnings quality: A review of the proxies, their determinants and their consequences. *Journal of Accounting & Economics*, *50*(2/3), 344–401.

Dechow, P., & Schrand, C. (2004). Earnings quality. *The Research Foundation of CFA Institute*.

Doupnik, T. S. (2008). Influence of Culture on Earnings Management: A Note. *Abacus*, *44*(3), 317–340.

Edwards, J. R. (2001). Multidimensional Constructs in Organizational Behavior Research: An Integrative Analytical Framework. *Organizational Research Methods*, *4*(2), 144–192.

Edwards, J. R., & Bagozzi, R. P. (2000). On the nature and direction of relationships between constructs and measures. *Psychological Methods*, *5*(2), 155–74.

Ewert, R., & Wagenhofer, A. (2011). *Earnings Quality Metrics and What They Measure* (Working Paper). University of Graz.

Ewert R. and A. Wagenhofer (2015), Economic Relations among Earnings Quality Measures, *Abacus*, 51 (3), 311–55.

Francis, J., Nanda, D., & Olsson, P. (2008). Voluntary disclosure, earnings quality and cost of capital. *Journal of Accounting Research*, *46*(1), 53–99.

Gaio, C., & Raposo, C. (2011). Earnings quality and firm valuation: International evidence. *Accounting and Finance*, *51*(2), 467–499.

Gefen, D., Straub, D. W., & Boudreau, M.-C. (2000). Structural Equation Modeling and Regression : Guidelines for Research Practice. *Communications of the Association for Information Systems*, *4*, 7.

Healy, P. M., Serafeim, G., Srinivasan, S., & Yu, G. (2014). Market competition, earnings management, and persistence in accounting profitability around the world. *Review of Accounting Studies*, *19*, 1281–1308.

International Accounting Standards Board (IASB) (2010). *The Conceptual Framework for Financial Reporting*. London.

Jarvis, C. B., MacKenzie, S. B., & Podsakoff, P. M. (2003). A Critical Review of Construct Indicators and Measurement Model Misspecification in Marketing and Consumer Research. *Journal of Consumer Research*, *30*(2), 199–218.

Jung, B., Lee, W.-J., & Weber, D. P. (2014). Financial reporting quality and labor investment efficiency. *Contemporary Accounting Research*, *31*(4), 1047–1076.

Kaplan, A. (1964). *The conduct of inquiry: methodology for behavioral science*. Chandler Pub. Co.

Leuz, C., Nanda, D., & Wysocki, P. D. (2003). Earnings management and investor protection: an international comparison. *Journal of Financial Economics*, *69*(3), 505–527.

Leuz, C., & Wysocki, P. D. (2016). The Economics of Disclosure and Financial Reporting Regulation: Evidence and Suggestions for Future Research. *Journal of Accounting Research*, *54*(2), 525–622.

Licerán-Gutiérrez, A., & Cano-Rodríguez, M. (2017). A review on the multidimensional analysis of earnings qualilty. In *EAA Annual Congress* (pp. 1–28). Valencia.

MacKenzie, S. B., Podsakoff, P. M., & Jarvis, C. B. (2005). The Problem of Measurement Model Misspecification in Behavioral and Organizational Research and Some Recommended Solutions. *Journal of Applied Psychology*, *90*(4), 710–730.

Nitzl, C. (2016). The use of partial least squares structural equation modelling (PLS-SEM) in management accounting research: Directions for future theory development. *Journal of Accounting Literature*, *37*, 19–35.

Sarstedt, M., Hair, J. F., Ringle, C. M., Thiele, K. O., & Gudergan, S. P. (2016). Estimation issues with PLS and CBSEM: Where the bias lies! *Journal of Business Research*, *69*(10), 3998–4010.

Schipper, K., & Vincent, L. (2003). Earnings quality. *Accounting Horizons*, *17*, 97–110.

VanTendeloo, B., & Vanstraelen, A. (2008). Earnings Management and Audit Quality in Europe: Evidence from the Private Client Segment Market. *European Accounting Review*, *17*(3), 447–469.

TABLE 1. DECISION RULES FOR DIFFERENTIATING BETWEEN FORMATIVE AND REFLECTIVE MODELS (Jarvis et al., 2003)

	Formative model	Reflective model
Criterion 1. Direction of causality	• From items to construct • Indicators define characteristics of the construct • Changes in the indicators produce changes in the construct • Changes in the construct should not produce changes in the indicators	• From construct to item • Indicators are manifestations of the construct • Changes in the indicators should not produce changes in the construct • Changes in the construct produce changes in the indicators
Criterion 2. Interchangeability of indicators	• Indicators are not interchangeable • Indicators need not have the same or similar content or share a common theme • Dropping one indicator alters the conceptual domain of the construct	• Indicators are interchangeable • Indicators should have the same or similar content and share a common theme • Dropping one indicator does not affect the conceptual domain of the construct
Criterion 3. Covariation among the indicators	• It is not necessary for indicators to covary with each other • Changes in one indicator are not necessarily associated with changes in the other indicators	• Indicators are expected to covary with each other • Changes in one indicator are associated with changes in the other indicators
Criterion 4. Nomological network of construct indicators	• Nomological network may differ across indicators • Indicators are not required to have the same antecedents and consequences	• Same nomological network for all the indicators • Indicators are required to have the same antecedents and consequences

TABLE 2. RESULTS OF THE SIMULATION PROCESS

r	*Single indicator*	*Decile ranking index*	*Common factor scores*	*PLS model*
−0.9	0.901	0.675	0.985	0.156
−0.8	0.840	0.549	0.945	0.146
−0.7	0.782	0.480	0.978	0.134
−0.6	0.709	0.431	0.985	0.133
−0.5	0.642	0.381	0.944	0.132
−0.4	0.568	0.352	0.951	0.130
−0.3	0.550	0.322	0.932	0.128
−0.2	0.467	0.298	0.878	0.123
−0.1	0.458	0.270	0.775	0.122
0.0	0.379	0.253	0.457	0.122
+0.1	0.338	0.235	0.228	0.123
+0.2	0.302	0.216	0.184	0.120
+0.3	0.265	0.200	0.164	0.115
+0.4	0.228	0.190	0.154	0.117
+0.5	0.186	0.178	0.141	0.118
+0.6	0.147	0.169	0.132	0.118
+0.7	0.123	0.159	0.123	0.117
+0.8	0.092	0.148	0.111	0.113
+0.9	0.068	0.141	0.103	0.112

FIGURE 1

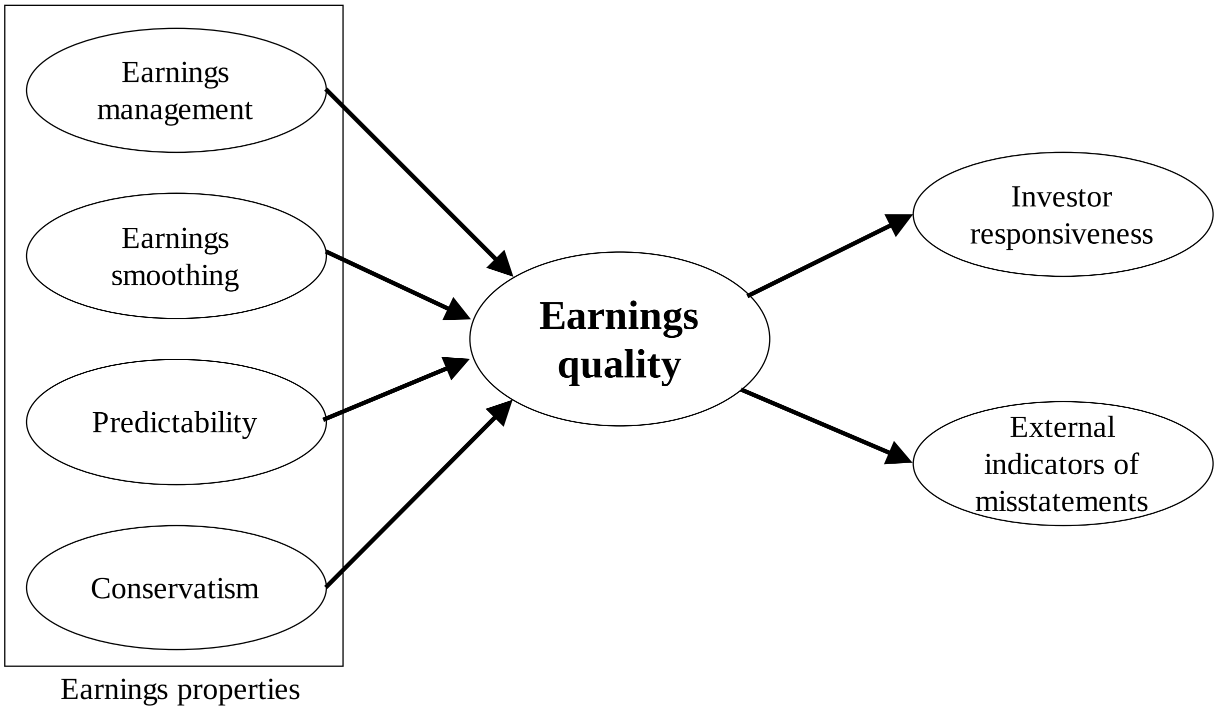
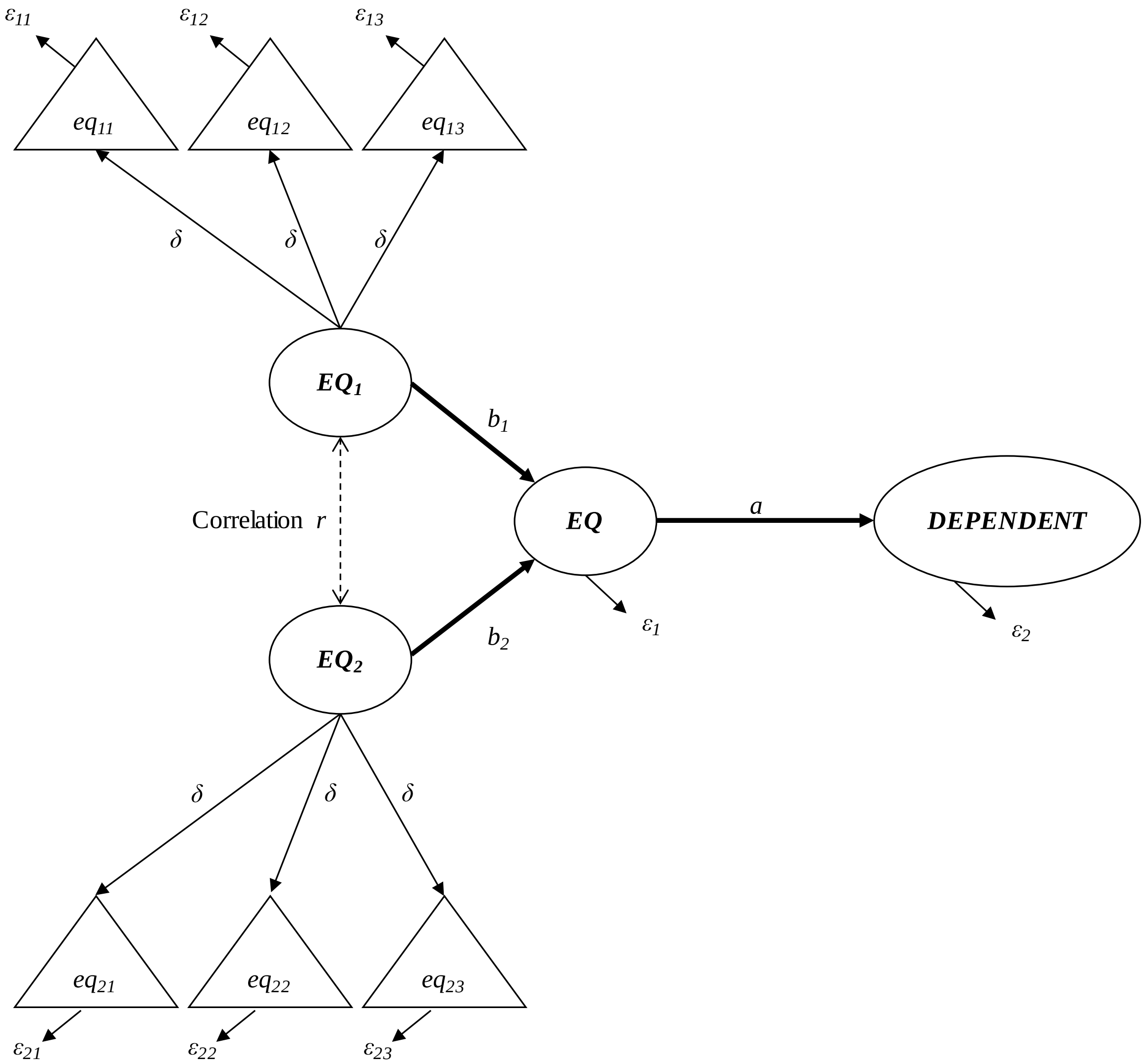
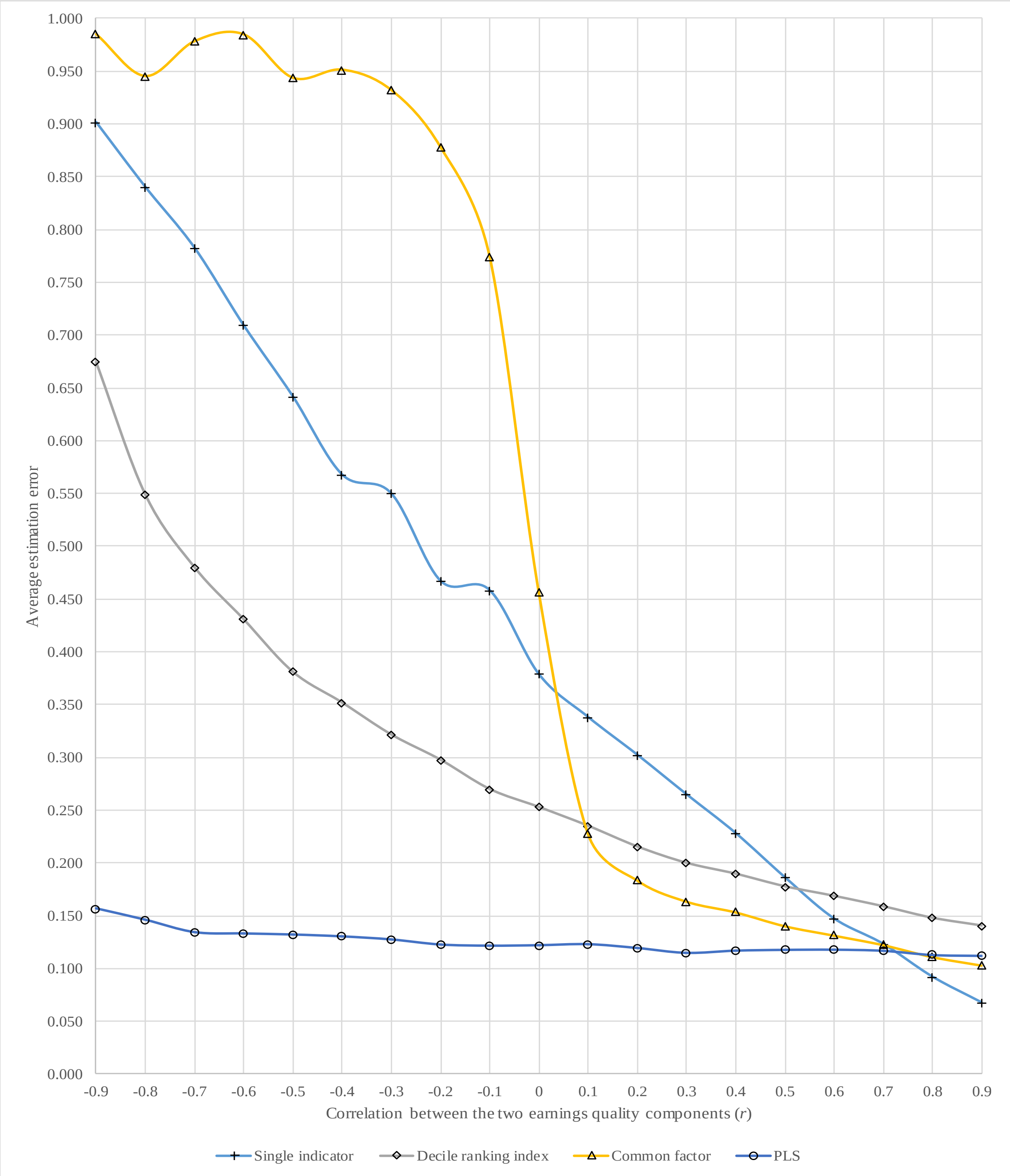

Earnings management
Earnings smoothing
Predictability
Conservatism
Earnings properties
Earnings quality
Investor responsiveness
External indicators of misstatements

FIGURE 2



ε11
ε12
ε13
eq11
eq12
eq13
δ
δ
δ
EQ1
b1
Correlation r
EQ
a
DEPENDENT
ε1
ε2
b2
EQ2
δ
δ
δ
eq21
eq22
eq23
ε21
ε22
ε23

FIGURE 3


1.000
0.950
0.900
0.850
0.800
0.750
0.700
0.650
0.600
0.550
0.500
0.450
0.400
0.350
0.300
0.250
0.200
0.150
0.100
0.050
0.000
Average estimation error
-0.9
-0.8
-0.7
-0.6
-0.5
-0.4
-0.3
-0.2
-0.1
0
0.1
0.2
0.3
0.4
0.5
0.6
0.7
0.8
0.9
Correlation between the two earnings quality components (r)
Single indicator
Decile ranking index
Common factor
PLS