Check for
updates

Higher-Order PLS-PM Approach for Different Types of Constructs

Corrado Crocetta[1] · Laura Antonucci[1] · Rosanna Cataldo[2] · Roberto Galasso[2] · Maria Gabriella Grassia[2] · Carlo Natale Lauro[3] · Marina Marino[2]

Abstract

Partial least squares path modeling (PLS-PM) has become very popular in recent years, for measuring concepts that depend on different aspects and that are based on different types of relationships. PLS-PM represents a useful tool to explore relationships and to analyze the influence of the different aspects on the complex phenomenon analyzed. In particular, the use of higher-order constructs has allowed researchers to extend the application of PLS-PM to more advanced and complex models. In this work, our attention is focused on higher-order constructs that include reflective or formative relationships. Even if the dispute between formative models and reflective models is not exactly recent, it is still alive in current literature, for the most part within the context of structural equation models. This paper focuses attention on theoretical and mathematical differences between formative and reflective measurement models within the context of the PLS-PM approach. A simulation study is proposed in order to show how these approaches fit well in different modeling situations. The approaches have been compared using empirical application in a sustainability context. The findings from the simulation and the empirical application can help researchers to estimate and to use the higher-order PLS-PM approach in reflective and formative type models.

Keywords Partial least squares-path modeling · Higher-order constructs · Formative constructs · Reflective constructs

1 Introduction

Over the last 30 years, many researchers have focused their attention on measuring the importance of constructs and the nature of the relationships between constructs. The focus of their scientific works, regarding reflective and formative relationships, has been primarily on identification and estimation issues (Blalock 1982; Bollen and Lennox 1991). The choice between formative or reflective models has enjoyed increasing attention in the literature of the recent years (Andreev et al. 2009; Diamantopoulos et al. 2008; Coltman et al.

✉ Corrado Crocetta
  corrado.crocetta@unifg.it

Extended author information available on the last page of the article

2008; Jarvis et al. 2003; MacKenzie et al. 2005; Petter et al. 2007). As Howell et al. (2007) write, "the works of Bollen and Lennox (1991), MacCallum and Browne MacCallum and Browne (1993), Edwards and Bagozzi (2000) and Bollen and Ting (2000) have given conceptual and methodological tools to deal with observable formative and reflective indicators. [...] Researchers in a number of disciplines are now opting to depart from the dominant reflective measurement tradition in the social sciences, choosing instead to develop and use formative measures. Diamantopoulos and Winklhofer (2001) presented guidelines for developing a formative measure that parallels DeVellis's (DeVellis 2016) paradigm for scale development under the reflective model". Particularly, Bollen and Ting (2000) proposed an empirical test to determine whether indicators are more likely to be formative or reflective. Similarly, Jarvis et al. (2003) suggested guidelines for the development of formative measures. The choice between formative and reflective models is also investigated in the PLS-PM approach and in particular in the Higher Order PLS-PM approach (Diamantopoulos and Siguaw 2006; Bollen and Lennox 1991). In PLS-PM, the selection of the construct mode of the Higher-Order approach, reflective or formative, is an important moment for researchers. They can choose among a reflective-reflective, reflective-formative, formative-reflective or formative-formative model (Jarvis et al. 2003; Wetzels et al. 2009). Each type represents the specific relationship between higher-order (HOC) and lower-order (LOC) constructs, as well as the measurement model used in the LOCs (Hair et al. 2017; Cheah et al. 2019). There are already many studies which have compared the approaches reported in the literature in the case of the reflective-formative type of HOCs in PLS-PM. The aim of this paper is to compare the performance of the different approaches available in PLS-PM to estimate the HOCs in different modeling situations through a simulation study, in which we include all types of HOCs in our conceptual framework. The results of the simulation will allow us to formulate guidelines and recommendations for the use of the different approaches for different type higher-order constructs in PLS-PM. The paper is organized as follow: in Sect. 2 we provide some conceptual background to the problem of formative vs reflective constructs; Sect. 3 reports the theoretical principles of the PLS-PM approach and, in detail, Higher-Order PLS-PM; in Sect. 4 the different approaches proposed in the literature are discussed; Sect. 5 presents the results of the simulation study while Sect. 6 reports the main results of the application case study; some concluding remarks will be made in the final Section.

2 Formative Versus Reflective Constructs

In the last few years, many articles have focused on the different constructs model, and about theoretical errors resulting from model misspecification (Bollen and Lennox 1991; Diamantopoulos and Winklhofer 2001; Jarvis et al. 2003; MacKenzie et al. 2005; Roy et al. 2012). In a *reflective model* (Edwards and Bagozzi 2000; Diamantopoulos and Winklhofer 2001) the construct is considered as the cause and the indicators its manifestations. For example, as Eboli et al. (2018) claim, "intelligence determines the responses of a subject to a questionnaire designed to assess this aspect and not vice versa". Hence, if the intelligence of a person increases, this will lead to an increase in the number of correct answers to all questions (Simonetto 2012). Thus, the construct determines its indicators (as shown on the left of Fig. 1) and each indicator, being a manifestation of the construct, can be removed if its coefficient is not statistically significant (Bollen and Lennox 1991). In a *formative model*, the indicators determine the latent construct (Bollen and Lennox 1991) (Fig. 1, on
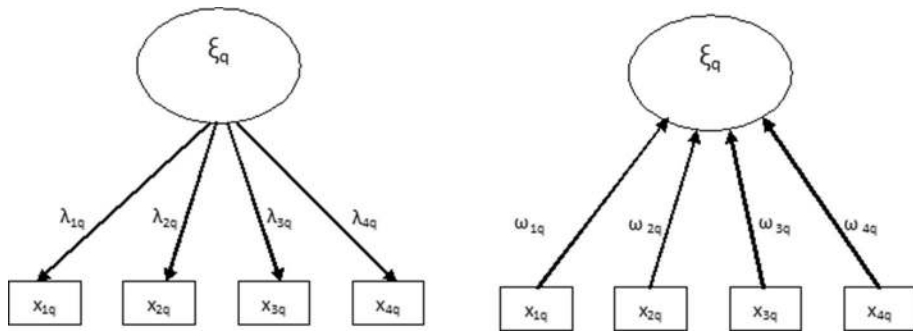
Fig. 1 Reflective and formative constructs

the right). According to MacCallum and Browne (1993), "in many cases indicators could be viewed as causing rather than being caused by the latent variable measured by the indicators". In this kind of model, a single indicator cannot be removed without affecting the definition of the construct. As Mazziotta and Pareto highlight in their works (Mazziotta and Pareto 2013, 2019), "a typical example of formative model is the measurement of well-being of society. This depends on health, income, occupation, services, environment, etc., and not vice versa. Therefore, if any one of these factors improves, well-being will increase (even if the other factors do not change). However, if well-being increases, this will not necessarily be accompanied by an improvement in all the other factors".

Today, formative models are becoming a standard tool in socio-economic research, particularly in the fields of causal modeling and multidimensional evaluation. Although theoretically the debate on the use of formative models is still very heated (there are still unsolved methodological problems encountered when addressing structural equation models comprising formative constructs), in practice these models are applied in many studies (Diamantopoulos et al. 2008). Generally, it is important to understand the nature of indicators, the reflective or formative nature, because an incorrect specification of the latent constructs can undermine the construct content validity, misrepresent a model, and lead to less useful theories for both researchers (Coltman et al. 2008). According to Coltman et al. (2008) there are three theoretical assumptions to consider when deciding whether the measurement model is formative or reflective: the nature of the construct, the direction of causality between the indicators and the latent construct, and the characteristics of the indicators used to measure the construct (Eboli et al. 2018).

In order to determine whether a construct is reflective or formative, the statistical instruments can also be used. As a matter of fact, the constructs are assessed through the theory underpinning each LV (Bagozzi 2007; Chin 1998b) and their validity by means of Cronbach's alpha and Average Variance Extracted (AVE) (Coltman et al. 2008). Additionally, the indicator-construct causality flow can also be checked using Weight-Loading Sign (WLS). A negative WLS indicates a causality issue which means that the indicator-construct link is impossible or reverse (Kock 2013; Wagner 1982), a tale-tale sign of a Simpson paradox instance (Kock 2015; Roni et al. 2015). Roy et al. (2012) claim "Wrongly modeling a reflective model as formative, and vice versa, is known as *model misspecification*". A misspecification in the measurement model impacts on the structural paths of the LV, thus leading to erroneous path coefficients (Jarvis et al. 2003; MacKenzie et al. 2005). For this reason, it is critical to understand when the use of formative or reflective

measurement models is appropriate and how such models should be formulated (Roy et al. 2012).

3 Higher-Order Constructs in PLS-PM

The PLS-PM approach analyzes multiple relationships between a set of blocks of Manifest Variables (MV), assuming that each block of variables is represented by a Latent Variable (LV) or by a theoretical concept and that the relationships between the blocks are established on the basis of the knowledge (theory) of the phenomenon analyzed. PLS-PM is evolving as a statistical modeling technique, and there are several published articles on the method (Bollen 1989; Chin 2010; Joseph et al. 2014). Two very important review papers on the PLS approach to Structural Equation Modeling (SEM) are Chin (1998b) and Tenenhaus et al. (2005). Lately, Lauro et al. (2018) showed recent developments in PLS-PM for the treatment of non-metric data, hierarchical data, longitudinal data and multi-block data. PLS-PM is made up of two elements, the "Measurement Model" (also called the "Outer Model"), which describes the relationships between the MVs and their respective LVs, and the "Structural Model" (also called the "Inner Model"), which describes the relationships between the LVs (Esposito et al. 2010; Roy et al. 2012; Eboli et al. 2018). As Lauro et al. (2018) write, "the PLS-PM approach consists of an iterative algorithm that computes the estimation of the LVs, measured by a set of MVs, and the relationships between them, by means of an interdependent system of equations based on multiple and simple regression. The idea is to determine the scores of the LVs through a process, that, iteratively, computes, first, an outer and, secondly, an inner estimation (Tenenhaus et al. 2005). For this reason, the procedure name is partial (Aria et al. 2018). In recent years, in the context of PLS-PM models, HOCs have become very popular (Edwards 2001; Jarvis et al. 2003; Johnson et al. 2012). According to Chin (1998b); Chin et al. (2003), "HOC Models, also known as *Hierarchical Models*, are explicit representations of multidimensional constructs that exist at a higher level of abstraction and are related to other constructs at a similar level of abstraction completely mediating the influence from or to their underlying dimensions". Law et al. (1998) define "[...] a construct as multidimensional when it consists of a number of interrelated attributes or dimensions and exists in multidimensional domains. These dimensions can be conceptualized under an overall abstraction, and it is theoretically meaningful and parsimonious to use this overall abstraction as a representation of the dimensions." Typically, HOC Models are characterized by the number of levels in the model (often restricted to second-order models) (Rindskopf and Rose 1988) and the different relationships between the HOCs and the LOCs (reflective and formative relationships) (Edwards 2001; Jarvis et al. 2003; Joseph et al. 2014; Wetzels et al. 2009; Becker et al. 2012). As shown by Becker et al. (2012), "a higher (or second)-order construct is a general concept that is either represented (reflective) or constituted (formative) by its dimensions (lower (or first)-order constructs). Therefore, the relation between the higher and lower-order constructs is not a question of causality, but rather a question of the nature of the hierarchical LV, as the higher-order construct (the general concept) does not exist without its lower-order constructs (dimensions). If the higher-order construct is reflective, the general concept is manifested by several specific dimensions, themselves being latent (unobserved). If the higher-order construct is formative, it is a combination of several specific (latent) dimensions in a general concept" (Edwards 2001; Wetzels et al. 2009). As we can see in Fig. 2, there are four main types of HOC Models discussed in the literature
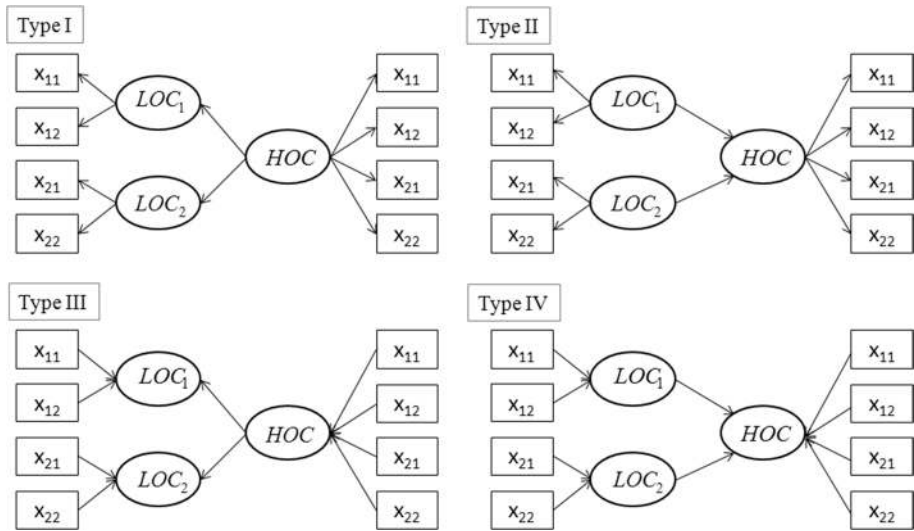
Fig. 2 Types of higher-order constructs

(Jarvis et al. 2003; Wetzels et al. 2009) and used in applications (Johnson et al. 2012). These types of models depend on the relationship among the first-order latent variables and their manifest variables, and the second-order latent variables and the first-order latent variables (Becker et al. 2012).

The Type I is the *Reflective-Reflective Measurement Model*, known as the *Second-Order Construct Type I*, one of the models most frequently applied in SEM among researchers nowadays. The Type II is the *Reflective-Formative Measurement Model Type II*. According to Chin's clarification, the LOCs are selectively measured constructs that do not share a common cause but rather form a general concept that fully mediates the impact on subsequent endogenous variables (Chin 1998b). In recent years, this type of model has become the most widely used in empirical applications, currently being considered by researchers, also thanks to the recent availability of appropriate model software (Roni et al. 2015). Several works describe different methods used to estimate the reflective-formative HOCs (Becker et al. 2012; Ciavolino 2012). The third type is the *Formative-Reflective Measurement Model Type III*, slightly different compared to the Reflective-Formative Type II in the explanation above. In this instance, the HOC is a common concept of several specific formative LOCs. Examples in the empirical literature are rather scarce, but a meaningful application of such a model could be firm performance as a reflective HOC measured by several different indices of a firm performance as formative LOCs (Becker et al. 2012). Finally, the fourth type is represented by the *Formative-Formative Measurement Model Type III*, the least frequently implemented in the SEM. This application is appropriate when both the HOC and LOCs are formative constructs.

Empirical studies reports that HOC PLS-PM, the reflective-formative type and formative-formative type, were most frequently employed. This indicates the predominance of formative type hierarchical LV models, even if clear guidelines on their use are lacking in the literature (Joseph et al. 2014). For PLS-PM, guidelines are mainly available for HOCs with reflective relationships (Lohmöller 2013; Wetzels et al. 2009; Wold 1982), even though Joseph et al. (2014) show that HOCs with reflective relationships in the first-order

and second-order of the hierarchy represent only a minority (20%) of the models applied in *MIS Quarterly*. However, there is a large need for guidelines on the use and modeling of HOCs with formative relationships in PLS-SEM (Becker et al. 2012).

4 Estimation of Higher-Order Constructs in PLS-PM

Within the frame of PLS-PM, two main approaches have been developed in the literature in order to estimate the parameters of HOCs: the Repeated Indicator Approach Lohmöller (2013) and the Two Step Approach (Joseph et al. 2014; Wetzels et al. 2009). Recently, other two approaches have been presented: the Mixed Two Step Approach and the PLS Components Regression Approach (Cataldo et al. 2017). All these approaches will be briefly described in this section. For a more detailed discussion and step-by-step illustration, see Becker et al. (2012), Lohmöller (2013), Joseph et al. (2014), Lauro et al. (2018) and Cataldo et al. (2017). The Repeated Indicators Approach is the first and the most popular approach: the indicators of the LOCs are used as the MVs of the HOC (Wilson and Henseler 2007) (Fig. 3). The Two Step Approach, as described by Cataldo et al. (2017), "consists of two phases: first, the LV scores of the LOCs are computed without the HOC (Rajala and Westerlund 2010); then, the PLS-PM analysis is performed using the computed scores as indicators of the HOCs. The implementation is not performed through a single PLS run; this implies that any Second-Order Construct, investigated in stage two, is not taken into account when estimating the LV scores in stage one" (Fig. 3).

The Mixed Two Step Approach begins with the implementation of the PLS-PM using the indicators of the LOCs as the MVs of the HOC, such as the Repeated Indicators Approach. In this way, the algorithm gives the scores of the LOCs. Then, the scores of the blocks become indicators of the HOC, and the PLS-PM algorithm is run again (Fig. 4). At last, the PLS Component Regression Approach is described by Cataldo et al. (2017) as consisting of three different steps (Fig. 5): "firstly, a HOC is formed of all the MVs of the LOCs; then, PLS-Regression is applied in order to obtain h components for each block; once h components have been obtained, they represent the MVs of the HOC and the PLS-PM algorithm is performed".

In the literature there are already works that compare these approaches within the reflective-formative type of HOC in PLS-PM. In particular (Becker et al. 2012) compared the Repeated Indicators Approach and the Two Step Approach using a simulation
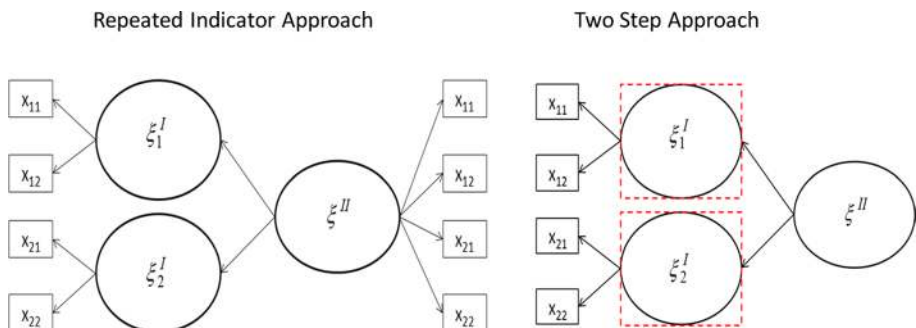
Fig. 3 Model building: the repeated indicator approach and the two step approach
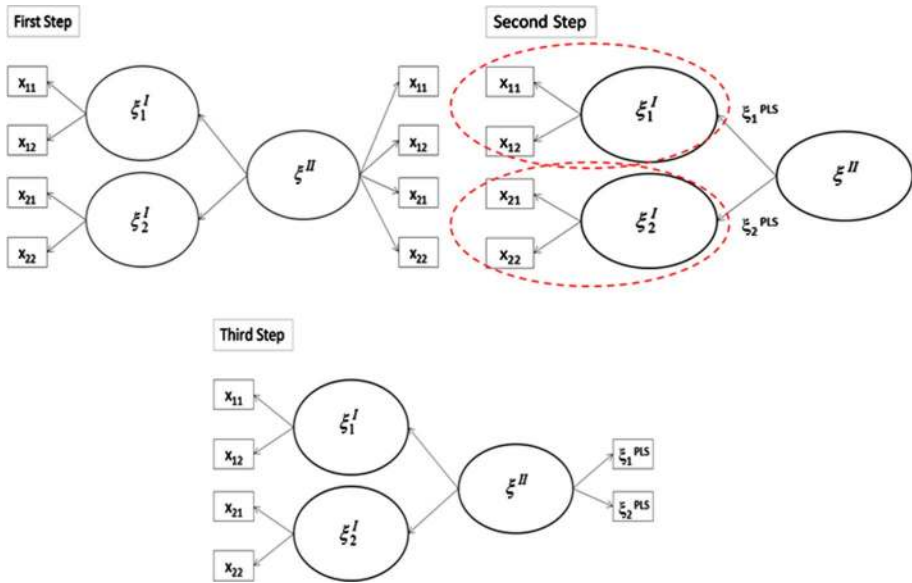
Fig. 4 The mixed two step approach
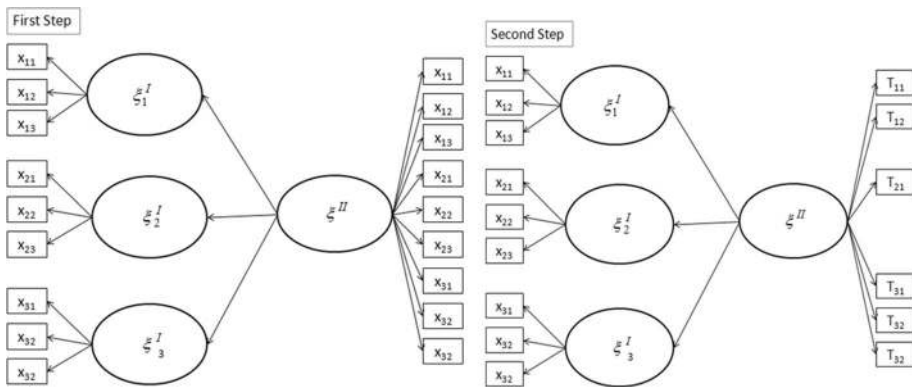
Fig. 5 Model building: the partial least squares component regression approach

study and an empirical application in strategic human resource management context. Many authors Joseph et al. (2014), Wetzels et al. (2009), Cheah et al. (2019) agree that the Two Step Approach has the advantage of estimating a more parsimonious model and proves suitable for the estimation of Second-Order Constructs since it produces estimates that are better than those obtained through the Repeated Indicators Approach. However, the Two Step Approach presents some important limitations related to the components of each block, such as the fact that only one component is chosen for each block, and this has a strong representative power but a weak predictive power in the analysis of the HOC. For these reason, lately, the two other approaches, the Mixed Two Step Approach and the PLS Components Regression Approach, have been proposed in

order to overcome these drawbacks. Since the aim of PLS-PM is to estimate the relationships between the LVs, these two approaches provide components that are at the same time representative of their blocks and predictive of the Second-Order Construct (Cataldo et al. 2017). Moreover, the PLS Component Regression Approach overcomes the problem of the number of components of the First-Order Constructs, giving the possibility of choosing the number of components to be extracted manually or according to a specific criterion (Cataldo et al. 2017). Cataldo et al. (2017) compared all four approaches through a simulation study with only one type of HOCs, particularly the reflective-formative type of HOCs and their findings suggest that the Mixed Two Step and PLS Component Regression Approaches are always the best choices, in terms of the bias and Mean Squared Error (MSE) of the estimates, when the researcher aims at studying the formative relationships of the structural model with constructs measured reflectively by their indicators. Starting from the simulation study of Cataldo et al. (2017), in this paper we propose an extended version of that simulation study, the approaches have been compared with all type categories reported by (Jarvis et al. 2003).

5 Simulation Study

The objective of this paper is to analyze and compare, within the same simulation design, the performance of the different approaches available in PLS-PM to estimate HOCs: the Repeated Indicators Approach, the Two Step Approach, the Mixed Two Step Approach and the PLS Component Regression Approach. In the literature there are works that apply simulation study for comparing the performance of the first two approaches for modeling HOCs, for example (Ciavolino and Nitti 2013; Becker et al. 2012). In this simulation work, we have included all the types of HOCs discussed in the literature (Jarvis et al. 2003) in our conceptual framework, using different sample sizes, in order to understand the effect of the sample dimension on the type of HOC. The performances have been evaluated in terms of the prediction accuracy, the estimate bias and the efficiency of the considered approach. The Monte Carlo simulation is used to compare the performance of these approaches, through the R language package. The data generation process is consistent with the procedure described by Paxton et al. (2001) for a Monte Carlo SEM study. In this simulation the same structure of the model and the parameters of the population has been adopted such as in the Cataldo et al. work (Cataldo et al. 2017). Firstly, we defined the structure of the model and the parameters of the population. Then, we generated randomly the Second-Order LV and, given the parameters and the error terms, we estimated the First-Order LVs. According to the outer parameters and error terms, finally, we generated the First and Second-Order MVs. The underlying population model used for the simulation consisted of one Second-Order LV (denoted by ξ^{II}) and four First-Order LVs (denoted by $\xi_1^I, \xi_2^I, \xi_3^I,$ and ξ_4^I), each of them formed of five MVs. The approach performances have been compared on the basis of the different sample size (n = 100, 250, 500, 1000). In the study design we considered 500 replications for each condition. Obviously, the Second-Order LV, in terms of the number of items, differed according to the estimation approach used: for the Repeated Indicators Approach it was formed of all the MVs of the First-Order constructs; for the Two Step and Mixed Two Step Approaches, it corresponded to the number of First-Order LVs; while for the PLS Component Regression Approach, the numerosity of the

block depended on the number of components of the First-Order dimension extracted by the PLS Regression.

The starting point was the generation of the First-Order LVs ξ_i^I as random variables $\xi_q^I \sim N(0, 1)$. The data generated were rescaled in the interval [1, 100].

For the formative structural model, the Second-Order Construct ξ_j^{II} was computed as the product of ξ_q^I by the path coefficient vector β_{qj} with the addition of an error component ζ_j according to the Eq. (1):

$$\xi_j^{II} = \sum_{(q:\xi_q^I \to \xi_j^{II})} \beta_{qj}\xi_q^I + \zeta_j \tag{1}$$

For the reflective structural model, the First-Order construct ξ_j^I was computed as the product of ξ_q^{II} by the path coefficient vector β_{qj} with the addition of an error component ζ_q according to the Eq. (2):

$$\xi_j^I = \sum_{(q:\xi_j^{II} \to \xi_q^I)} \beta_{qj}\xi_j^{II} + \zeta_q \tag{2}$$

The path coefficient vector (β) of the structural model was assumed to have elements equal to 0.7.

For the formative measurement model, the LV was supposed to be generated by its own MVs, following the Eq. (3):

$$\xi_q^I = \sum_{p=1}^{P_q} \omega_{pq}x_{pq} + \delta_q \tag{3}$$

For the reflective measurement model, the MVs are generated starting from the LVs, given the lambda coefficients, following the Eq. (4):

$$x_{nq} = \xi_q^I * (\lambda^I)_k^{-1} + \delta_{nq} \tag{4}$$

where the error term was distributed as a continuous uniform: $\delta \sim U(-1, 1)$. We estimated each approach with a centroid inner weighting scheme.

In order to assess and compare the methods estimated we used the Relative Bias (*RB*) and Standard Deviation (*StD*) following the indications in (Cataldo et al. 2017).

The *RB* was computed as:

$$RB = \frac{1}{n}\sum_{i=1}^{n}\frac{(\hat{\theta}_i - \theta)}{\theta} \qquad i = 1, 2, \ldots, 500 \tag{5}$$

where n represents the number of replications in the simulation, $\hat{\theta}_i$ is the parameter estimate for each replication and θ is the corresponding population parameter. The formula is equivalent to the mean *RB* (Reinartz et al. 2002).

The *StD* was computed as:

$$StD = \sqrt{\frac{1}{n}\sum_{i=1}^{n}(\hat{\theta}_i - E(\hat{\theta}))^2} \qquad i = 1, 2, \ldots, 500 \tag{6}$$

where $E(\hat{\theta})$ is the mean of the estimates across the 500 simulated datasets. A positive RB indicates an overestimation of the true parameter, a negative RB an underestimation. Instead StD index provides information on the efficiency of the estimates. For each scheme it was decided to compare the communality index which is always calculable with each type of relationship both for the structural model and for the measurement model. Furthermore, for the reflective models, the communality was equal to the AVE net of a constant, whereas for the training models it approximated the redundancy index. Table 1 reports the simulation results relating to the communality, relative bias and standard deviation. The results are grouped according to the estimation approach used, sample size and type of higher-order construct.

The estimated community was always significant for all approaches except for the Repeated Indicators Approach that had very low values. Considering the value of the indicator, as can be seen in Fig. 6, it was always higher using the PLS-Regression algorithm, both as the number of samples increased and with each type of model ratio. With the Mixed Two Step approach, instead, it was always lower than with the PLS-Regression but however higher than with the Two Step Approach. Only for the reflective-reflective type, did the communality index of the two approaches have the same value. In Table 1 we also note how, as the number of the sample increased, the variability of the estimate was lower for the PLS-Regression than for the other approaches.

In the "Appendix, Tables 5, 6, 7 and 8" report the simulation results relating to the coefficients (computed as the average of the 500 replications), relative bias and standard deviation. The results are grouped according to the estimation approach used and sample size. For each combination, the path coefficients, relative bias and standard deviation of the four parameters are reported. As can be seen in Tables 5, 6, 7 and 8 , the estimated paths did not differ much for all the sample sizes and for all the approaches and frameworks used. The path estimates were always significant but, while remaining low, the variability of the estimates was lower when using the Mixed Two Step and PLS-Regression Approaches. Finally, the relative bias of the path coefficients are reported in detail in Tables 9, 10, 11 and 12 grouped for each framework (see the "Appendix"). The Two Step Approach heavily underestimated all the path coefficients linking the First-Order Construct with the Second-Order LV in all sample sizes. Looking at the new methods proposed, we can see that for relative small samples (n=100; n=250) the Mixed Two Step Approach worked best, producing estimates near to zero, while the methods had the same performance for large samples (n=500; n=1000), giving an equivalent accuracy.

The Mixed Two Step and PLS-Regression approaches demonstrated a greater accuracy in predicting the higher level construct, since the communality index was higher than that of the Two Step Approach. The difference is remarkable for all sample size. Therefore, these approaches were also the best option for predicting the Second-Order LV. Overall, the results show that these two methods were always the best choice, in terms of the bias and MSE of the estimates.

6 Application Case Study

The approaches have been compared also in an empirical application in a sustainability context. Global Sustainability was conceived as a Third-Order Construct affecting the Second-Order dimensions, which in turn shaped the First-Order LVs, underlying

Table 1 Communality value for each approach

Approach	Sample size		Reflective formative	Formative reflective	Reflective reflective	Formative formative
Repeated indicators	100	Comm	0.3131	0.3417	0.3668	0.3339
		RB	0.0071	− 0.0124	0.0025	− 0.0087
		StD	0.0303	0.0414	0.0314	0.0328
	250	Comm	0.4577	0.3598	0.3669	0.4491
		RB	− 0.0007	− 0.0039	0.0000	− 0.0008
		StD	0.0199	0.0214	0.0195	0.0212
	500	Comm	0.3199	0.3959	0.4044	0.3117
		RB	0.0008	− 0.0007	0.0011	− 0.0018
		StD	0.0130	0.0146	0.0146	0.0149
	1000	Comm	0.4031	0.3278	0.3470	0.3880
		RB	− 0.0001	− 0.0005	0.0004	0.0002
		StD	0.0104	0.0103	0.0091	0.0110
Two step	100	Comm	0.4570	0.6002	0.6023	0.5620
		RB	0.0059	− 0.0048	− 0.0009	− 0.0256
		StD	0.0438	0.0353	0.0412	0.0481
	250	Comm	0.6263	0.6216	0.6226	0.6022
		RB	− 0.0009	− 0.0005	− 0.0019	− 0.0029
		StD	0.0220	0.0240	0.0231	0.0252
	500	Comm	0.5535	0.6698	0.6701	0.5413
		RB	0.0010	0.0007	0.0007	− 0.0022
		StD	0.0168	0.0151	0.0154	0.0191
	1000	Comm	0.6246	0.6109	0.6116	0.6113
		RB	0.0001	− 0.0006	0.0002	− 0.0011
		StD	0.0108	0.0121	0.0113	0.0120

Table 1 (continued)

Approach	Sample size		Reflective formative	Formative reflective	Reflective reflective	Formative formative
Mixed two-step	100	Comm	0.6115	0.6635	0.6161	0.6909
		RB	0.0091	0.0228	0.0049	0.0200
		StD	0.0354	0.0278	0.0393	0.0324
	250	Comm	0.7190	0.6509	0.6281	0.7472
		RB	− 0.0006	0.0092	0.0010	0.0073
		StD	0.0192	0.0229	0.0226	0.0198
	500	Comm	0.5832	0.7011	0.6766	0.6114
		RB	0.0013	0.0041	0.0013	0.0066
		StD	0.0160	0.0139	0.0148	0.0159
	1000	Comm	0.6679	0.6903	0.6291	0.7102
		RB	0.0007	0.0015	0.0010	0.0015
		StD	0.0112	0.0112	0.0107	0.0111

Table 1 (continued)

Approach	Sample size		Reflective formative	Formative reflective	Reflective reflective	Formative formative
Pls-regression	100	Comm	0.9578	0.9535	0.9627	0.9558
		RB	− 0.0405	− 0.0426	− 0.0423	− 0.0483
		StD	0.0717	0.0663	0.0584	0.0677
	250	Comm	0.9914	0.9701	0.9721	0.9896
		RB	− 0.0040	− 0.0039	− 0.0040	− 0.0052
		StD	0.0203	0.0126	0.0118	0.0233
	500	Comm	0.9527	0.9842	0.9852	0.9511
		RB	− 0.0028	− 0.0010	− 0.0007	− 0.0035
		StD	0.0096	0.0034	0.0032	0.0125
	1000	Comm	0.9839	0.9670	0.9709	0.9832
		RB	− 0.0005	− 0.0013	− 0.0005	− 0.0005
		StD	0.0025	0.0072	0.0040	0.0025
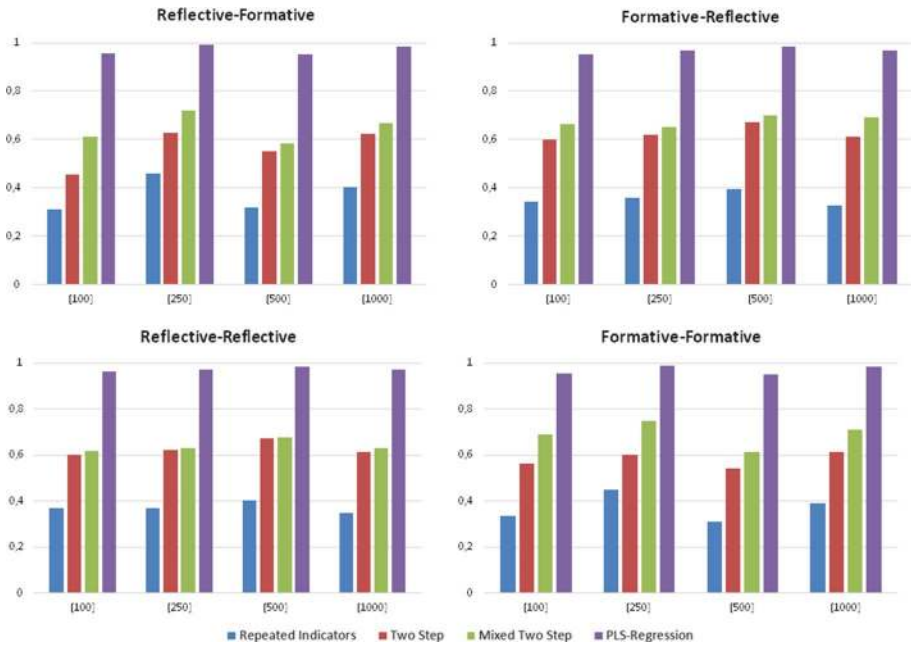
Fig. 6 Communality comparison

specific aspects of the Second-Order dimensions. There were 17 First-Order LVs Sustainable Development Goals (SDGs) with, in total, 169 elementary indicators (MVs) (Cataldo et al. 2020). For the sake of simplicity and for illustrative purposes only, the analysis focuses on the social area of Sustainability and its related goals with 54 elementary indicators. This decision was made in order to facilitate, particularly, the implementation of the different approaches. This section reports the main results of the

Fig. 7 The social area of sustainability and its goals
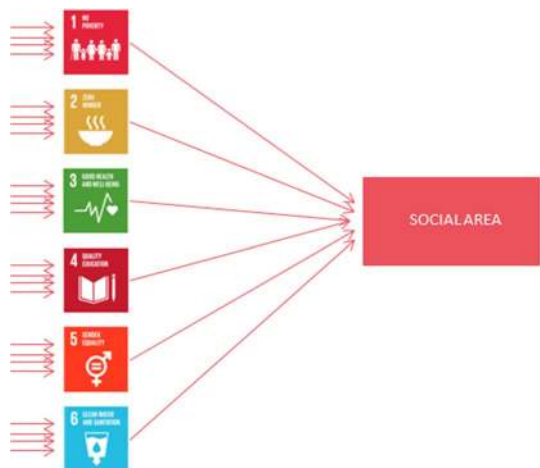
Higher-Order PLS-PM analysis, with a special consideration of the evaluation of the measurement model and the structural model.

The application case focuses on the Type II category reported by Jarvis et al. (2003), a model resulting from the combination of formative LOC and formative HOC (Fig. 7).

The official data derive from the database of the United Nations "Sustainable Development Goals". The analysis was developed with reference to the European Community countries, in total 28. The data are related to the triennium 2015-2017, in particular to the year 2017 (however, for some variables, because of the lack of data for that year, the previous years of this triennium were taken as the reference). Considering that each EIs has different units and values, for the purposes of comparison all the units were first normalized to a value between 0 and 1, where 0 was the value assigned to the least sustainable country while 1 was assigned to the most sustainable for each EI. The Sanchez "plspm" package in the R programming language (Sanchez 2013) was used in order to perform the PLS-PM analysis involving the formative indicators and the centroid scheme for the inner estimation.

Table 2 reports the main quality measurements of the four model approaches.

The Communality measure of the Mixed Two Step approach and the PLS-Regression approach was higher than that of the Repeated Approach and the Two Step Approach. Therefore, the amount of variability of the MVs captured by the social area construct of SDGs using these two methods was higher than when the two classic methods known in the literature were adopted. It is important to note that there is a difference in the use of the scores of the First-Order dimensions. In order to assess the significance of the path coefficients, Table 3 reports the value and significance of the structural coefficients linking the First-Order dimensions to the Second-Order construct.

The model estimated with the Repeated Indicator approach and the Two Step Approach some not significant paths. In all the approaches the goal "Good Health and Well-Being" Goal proves to be most influential among all the factors. This dimension proves not to be significant for the Two Step Approach precisely because it is related to the way it is calculated.

The quality of the model (Tenenhaus et al. 2004) (Table 4) is quite high in all approaches, except the Repeated Indicators Approach, but slightly higher if we estimate the components with the Mixed Two Step Approach and PLS Component Regression Approach. Therefore, the application case focused on the Type II category reported by Jarvis et al. (2003) demonstrated how well the approaches fit the construct in terms of communality and path significance. We can see that the PLS-Regression approach works better than the other approaches reported in the literature.

Table 2 Communality index of the social area of Sustainability for each approach

	Repeated approach	Two step approach	Mixed two step approach	PLS-regression approach
Communality	0.230	0.415	0.631	0.813

Table 3 Path Coefficients and t-statistics for each approach (the not significant path coefficients for each method are marked in bold)

	Repeated approach	Two step approach	Mixed two step approach	PLS-R approach
No poverty				
Path	0.08	0.09	0.11	0.09
T-value	1.61	3.91	2.23	8.72
Zero hunger				
Path	0.06	0.03	0.04	0.06
T-value	1.51	**1.34**	2.41	5.03
Good health and well-being				
Path	0.56	**0.60**	0.57	0.58
T-value	6.58	**0.15**	2.32	2.43
Quality education				
Path	0.16	0.17	0.14	0.17
T-value	2.54	5.84	2.43	9.45
Gender equality				
Path	**0.09**	**0.01**	0.06	0.07
T-value	**1.28**	**0.15**	1.95	3.22
Clean water and sanitation				
Path	0.18	0.22	0.19	0.13
T-value	2.72	6.75	2.61	6.68

Table 4 Global measure of goodness of fit

Repeated approach	Two step approach	Mixed two step approach	PLS-R approach
0.419	0.567	0.588	0.704

7 Conclusions and Future Research

The debate over the reflective or formative approach is still open. In the literature it is possible to find considerations on this theme, sometimes in conflict with each other. The goal of this work has been to present the approaches for the PLS-PM parameter estimation in the presence of HOCs. The performances of the Higher-Order approaches have been compared in different modeling situations through simulations, in which we have tested all types of HOCs in our conceptual framework. The performances of these approaches have been analyzed through a simulation study. The PLS-Regression and the Mixed Two Step approaches, compared with the Repeated Indicators Approach and the Two Step Approach

have almost always shown more stable estimates, both when the sample size changes and when the relationships between the LVs and the MVs change. In all the cases considered, the communalities are always higher when using the PLS-Regression Approach. The PLS-Regression Approach has shown a higher variability than the mixed approach for small samples. As the sample size increases, PLS-Regression improves in terms of estimates, becoming more stable than the Mixed Two Step Approach. In terms of relative error, the results vary with respect to the type of relationship that binds them. The Mixed Two Step and PLS Regression approaches are always the best choices, in terms of bias and MSE of the estimates. They slightly outperform, in terms of prediction accuracy, the Repeated Indicators Approach and the Two Step Approach. In general, if we work on large samples these two methods have the same performance. To be more accurate, the PLS Regression Approach would seem to be the best, because, as shown in the simulations, it works better than the Mixed Two Step Approach. In the empirical example about Sustainability we proposed, we have verified that with small samples if we want to study the formative relationships of the structural model, with constructs measured by their indicators in a formative way, the Mixed Two Step Approach and PLS Component Regression Approach are the most powerful methods, in terms of quality of the model and path significance. It is interesting to note that it is not possible to define strict rules to choose between a reflective or a formative model. The researchers to decide the mode must consider the latent construct and the indicators (observable or observed) at their disposal. The simulation study we have performed considers a simple situation with only one Second-order LV and four First-order LV. In the future we want to study a more complex model considering also Third or Fourth order LV, using more and more MV and LV. In this way will see how the different approaches analyzed perform when the complexity increase. We intend to test, also, the performance of high order PLS path models with qualitative external informations to take into account characteristics about the units or about the variables (Ciavolino et al. 2015). In our application for example we could have considered the qualitative external information: name of the UE Country in order to to compare the performances of the different Nations.

Appendix

See Tables 5, 6, 7, 8, 9, 10, 11 and 12.

Table 5 Path coefficients, relative bias and standard deviation for each approach with respect to the reflective-formative framework

Approach	Size sample		β_1	β_2	β_3	β_4
Repeated indicators	100	Path	0.2942	0.3705	0.3373	0.2717
		RB	− 0.0009	− 0.0041	0.0003	-0.0044
		StD	0.0215	0.0249	0.0244	0.0304
	250	Path	0.2806	0.2879	0.3082	0.3022
		RB	− 0.0001	− 0.0001	0.0009	− 0.0004
		StD	0.0102	0.0091	0.0095	0.0101
	500	Path	0.3279	0.3267	0.3402	0.3148
		RB	0.0000	0.0001	− 0.0014	− 0.0003
		StD	0.0124	0.0120	0.0129	0.0124
	1000	Path	0.3179	0.3065	0.2758	0.3217
		RB	0.0001	− 0.0001	− 0.0001	0.0001
		StD	0.0062	0.0063	0.0061	0.0061
Two step	100	Path	0.3314	0.3270	0.3307	0.2879
		RB	− 0.0001	− 0.0001	0.0001	0.0002
		StD	0.0155	0.0155	0.0153	0.0176
	250	Path	0.2868	0.2911	0.3037	0.2978
		RB	− 0.0001	0.0001	0.0001	0.0002
		StD	0.0053	0.0057	0.0070	0.0062
	500	Path	0.3282	0.3322	0.3191	0.3301
		RB	− 0.0004	0.0005	− 0.0010	− 0.0003
		StD	0.0077	0.0077	0.0074	0.0075
	1000	Path	0.3072	0.3024	0.3029	0.3106
		RB	0.0000	− 0.0001	− 0.0001	0.0003
		StD	0.0039	0.0038	0.0038	0.0038
Mixed two step	100	Path	0.3298	0.3252	0.3256	0.2968
		RB	− 0.0009	− 0.0015	− 0.0009	− 0.0049
		StD	0.0131	0.0156	0.0122	0.0159
	250	path	0.2871	0.2960	0.3013	0.2949
		RB	− 0.0003	− 0.0013	0.0013	0.0007
		StD	0.0051	0.0054	0.0064	0.0056
	500	Path	0.3266	0.3358	0.3201	0.3265
		RB	− 0.0005	− 0.0003	− 0.0007	0.0001
		StD	0.0072	0.0076	0.0070	0.0072
	1000	Path	0.3083	0.3024	0.3019	0.3101
		RB	− 0.0002	− 0.0001	0.0002	0.0001
		StD	0.0036	0.0036	0.0036	0.0038

Table 5 (continued)

Approach	Size sample		β_1	β_2	β_3	β_4
PLS-regression	100	Path	0.2927	0.3652	0.3355	0.2820
		RB	0.0003	− 0.0047	0.0001	− 0.0043
		StD	0.0244	0.0234	0.0243	0.0265
	250	Path	0.2811	0.2880	0.3075	0.3024
		RB	0.0018	0.0003	− 0.0011	− 0.0006
		StD	0.0109	0.0098	0.0178	0.0129
	500	Path	0.3277	0.3249	0.3403	0.3168
		RB	0.0006	0.0000	− 0.0017	− 0.0001
		StD	0.0109	0.0106	0.0109	0.0109
	1000	Path	0.3169	0.3072	0.2775	0.3204
		RB	− 0.0001	0.0000	0.0000	0.0001
		StD	0.0058	0.0059	0.0057	0.0057

Table 6 Path coefficients, relative bias and standard deviation for each approach with respect to formative-reflective framework

Approach	Sample size		β_1	β_2	β_3	β_4
Repeated indicators	100	Path	0.8181	0.8365	0.8254	0.7765
		RB	0.0108	0.0109	0.0123	0.0206
		StD	0.0278	0.0299	0.0252	0.0325
	250	Path	0.7781	0.8072	0.8386	0.8012
		RB	0.0077	0.0056	0.0037	0.0059
		StD	0.0256	0.0235	0.0177	0.0200
	500	Path	0.8388	0.8478	0.7939	0.8666
		RB	0.0022	0.0027	0.0037	0.0014
		StD	0.0127	0.0110	0.0148	0.0102
	1000	Path	0.8262	0.8386	0.8195	0.8391
		RB	0.0013	0.0003	0.0012	0.0008
		StD	0.0095	0.0093	0.0099	0.0089
Two step	100	Path	0.8309	0.8185	0.8034	0.7441
		RB	0.0009	0.0100	0.0104	0.0061
		StD	0.0283	0.0362	0.0282	0.0417
	250	Path	0.7719	0.8090	0.8162	0.8061
		RB	0.0035	0.0021	0.0042	0.0024
		StD	0.0268	0.0217	0.0202	0.0190
	500	Path	0.8369	0.8369	0.7931	0.8560
		RB	0.0013	0.0023	0.0020	0.0014
		StD	0.0131	0.0114	0.0153	0.0112
	1000	Path	0.8228	0.8061	0.8096	0.8123
		RB	0.0007	− 0.0001	0.0006	0.0010
		StD	0.0102	0.0111	0.0106	0.0103
Mixed two step	100	Path	0.8181	0.8364	0.8254	0.7765
		RB	0.0107	0.0110	0.0123	0.0206
		StD	0.0278	0.0299	0.0252	0.0325
	250	Path	0.7781	0.8072	0.8386	0.8011
		RB	0.0077	0.0055	0.0038	0.0060
		StD	0.0256	0.0236	0.0177	0.0200
	500	Path	0.8388	0.8478	0.7938	0.8666
		RB	0.0022	0.0027	0.0038	0.0014
		StD	0.0127	0.0110	0.0148	0.0102
	1000	Path	0.8262	0.8385	0.8194	0.8390
		RB	0.0013	0.0003	0.0012	0.0008
		StD	0.0095	0.0093	0.0099	0.0089

Table 6 (continued)

Approach	Sample size		β_1	β_2	β_3	β_4
PLS-regression	100	Path	0.8266	0.8337	0.8045	0.7462
		RB	0.0009	0.0099	0.0157	0.0117
		StD	0.0292	0.0376	0.0303	0.0411
	250	Path	0.7758	0.8072	0.8194	0.8061
		RB	0.0040	0.0026	0.0048	0.0029
		StD	0.0269	0.0227	0.0207	0.0192
	500	Path	0.8414	0.8359	0.7930	0.8570
		RB	0.0012	0.0025	0.0026	0.0016
		StD	0.0134	0.0116	0.0153	0.0113
	1000	Path	0.8308	0.8091	0.8119	0.8165
		RB	0.0005	0.0002	0.0006	0.0014
		StD	0.0103	0.0110	0.0107	0.0102

Table 7 Path coefficients, relative bias and standard deviation for each approach with respect to reflective-reflective framework

Approach	Sample size		β_1	β_2	β_3	β_4
Repeated indicators	100	Path	0.8501	0.7815	0.7888	0.7089
		RB	0.0013	0.0015	0.0020	0.0053
		StD	0.0278	0.0402	0.0357	0.0644
	250	Path	0.7448	0.8161	0.8005	0.8080
		RB	− 0.0027	0.0009	0.0004	0.0004
		StD	0.0346	0.0216	0.0238	0.0218
	500	Path	0.8397	0.8323	0.7692	0.8453
		RB	0.0008	0.0006	0.0014	0.0006
		StD	0.0140	0.0136	0.0192	0.0118
	1000	Path	0.7896	0.7970	0.7999	0.7858
		RB	0.0003	0.0007	0.0002	0.0004
		StD	0.0111	0.0118	0.0118	0.0120
Two step	100	Path	0.8329	0.7788	0.7956	0.7222
		RB	0.0028	− 0.0002	0.0038	0.0011
		StD	0.0285	0.0406	0.0341	0.0604
	250	Path	0.7475	0.8127	0.8064	0.8019
		RB	− 0.0027	0.0012	− 0.0002	0.0004
		StD	0.0322	0.0212	0.0225	0.0211
	500	Path	0.8293	0.8292	0.7802	0.8479
		RB	0.0008	0.0005	0.0014	0.0005
		StD	0.0143	0.0132	0.0179	0.0112
	1000	Path	0.7936	0.7837	0.7946	0.7962
		RB	0.0004	0.0007	0.0001	0.0002
		StD	0.0107	0.0120	0.0113	0.0113
Mixed two step	100	Path	0.8308	0.7839	0.7945	0.7235
		RB	0.0032	0.0007	0.0041	0.0024
		StD	0.0296	0.0399	0.0338	0.0574
	250	Path	0.7495	0.8116	0.8069	0.8018
		RB	− 0.0026	0.0009	0.0003	0.0008
		StD	0.0319	0.0215	0.0222	0.0210
	500	Path	0.8299	0.8291	0.7801	0.8484
		RB	0.0013	0.0005	0.0016	0.0002
		StD	0.0142	0.0131	0.0179	0.0111
	1000	Path	0.7966	0.7853	0.7938	0.7974
		RB	0.0004	0.0008	0.0002	0.0004
		StD	0.0105	0.0118	0.0112	0.0110

Table 7 (continued)

Approach	Sample size		β_1	β_2	β_3	β_4
PLS-regression	100	Path	0.8479	0.7792	0.7888	0.7141
		RB	0.0019	0.0070	− 0.0044	0.0033
		StD	0.0294	0.0465	0.0387	0.0613
	250	Path	0.7470	0.8153	0.8002	0.8069
		RB	− 0.0023	0.0009	0.0004	0.0002
		StD	0.0327	0.0213	0.0233	0.0216
	500	Path	0.8389	0.8318	0.7711	0.8449
		RB	0.0008	0.0007	0.0015	0.0004
		StD	0.0139	0.0135	0.0184	0.0118
	1000	Path	0.7893	0.7963	0.7998	0.7868
		RB	0.0005	0.0006	0.0001	0.0004
		StD	0.0109	0.0116	0.0115	0.0116

Table 8 Path coefficients, relative bias and standard deviation for each approach with respect to formative-formative framework

Approach	Sample size		β_1	β_2	β_3	β_4
Repeated indicators	100	Path	0.3007	0.3008	0.3008	0.3008
		RB	− 0.0039	− 0.0039	− 0.0040	− 0.0040
		StD	0.0068	0.0068	0.0068	0.0068
	250	Path	0.2891	0.2891	0.2892	0.2892
		RB	− 0.0012	− 0.0012	− 0.0012	− 0.0012
		StD	0.0038	0.0038	0.0038	0.0038
	500	Path	0.3197	0.3196	0.3196	0.3197
		RB	− 0.0015	− 0.0015	− 0.0015	− 0.0015
		StD	0.0041	0.0041	0.0041	0.0041
	1000	Path	0.2967	0.2967	0.2967	0.2967
		RB	− 0.0004	− 0.0004	− 0.0004	− 0.0004
		StD	0.0023	0.0023	0.0023	0.0023
Two step	100	Path	0.2912	0.2884	0.2656	0.3102
		RB	− 0.0028	− 0.0021	0.0012	− 0.0111
		StD	0.0103	0.0081	0.0126	0.0249
	250	Path	0.2772	0.3092	0.2756	0.2924
		RB	− 0.0028	− 0.0021	0.0012	− 0.0017
		StD	0.0115	0.0081	0.0146	0.0116
	500	Path	0.3229	0.3349	0.3071	0.3099
		RB	− 0.0018	− 0.0011	− 0.0019	− 0.0009
		StD	0.0078	0.0073	0.0116	0.0114
	1000	Path	0.3115	0.2815	0.2922	0.2984
		RB	− 0.0004	0.0003	− 0.0009	− 0.0005
		StD	0.0059	0.0071	0.0070	0.0061
Mixed two step	100	Path	0.3008	0.3007	0.3008	0.3008
		RB	− 0.0039	− 0.0039	− 0.0040	− 0.0040
		StD	0.0068	0.0068	0.0068	0.0068
	250	Path	0.2891	0.2891	0.2892	0.2891
		RB	− 0.0012	− 0.0012	− 0.0013	− 0.0012
		StD	0.0038	0.0038	0.0038	0.0038
	500	Path	0.3196	0.3197	0.3196	0.3196
		RB	− 0.0015	− 0.0015	− 0.0015	− 0.0015
		StD	0.0041	0.0041	0.0041	0.0041
	1000	Path	0.2967	0.2967	0.2967	0.2968
		RB	− 0.0004	− 0.0004	− 0.0003	− 0.0004
		StD	0.0023	0.0023	0.0023	0.0023

Table 8 (continued)

Approach	Sample size		β_1	β_2	β_3	β_4
PLS-regression	100	Path	0.3048	0.2762	0.3428	0.2728
		RB	− 0.0133	0.0079	− 0.0100	0.0003
		StD	0.0291	0.0315	0.0253	0.0339
	250	Path	0.2777	0.3146	0.2766	0.2859
		RB	− 0.0038	− 0.0030	0.0019	− 0.0005
		StD	0.132	0.0106	0.0165	0.0135
	500	Path	0.3212	0.3393	0.3093	0.3058
		RB	− 0.0015	− 0.0017	− 0.0017	− 0.0009
		StD	0.0087	0.0096	0.0131	0.0124
	1000	Path	0.3154	0.2802	0.2915	0.2969
		RB	− 0.0003	0.0012	− 0.0017	− 0.0007
		StD	0.0071	0.0081	0.0081	0.0074

Table 9 Relative Bias with respect to reflective-formative framework

Approach	Path	100	250	500	1000
Repeated indicators	β_1	− 0.00183	− 0.00054	− 0.00132	− 0.00253
	β_2	0.01428	− 0.00020	− 0.00161	− 0.00057
	β_3	− 0.00460	− 0.00262	0.00078	− 0.00192
	β_4	− 0.00826	− 0.00103	− 0.00068	− 0.00054
Two step	β_1	− 0.00145	− 0.00124	− 0.00112	− 0.00197
	β_2	− 0.00218	− 0.00127	− 0.00143	− 0.00099
	β_3	− 0.00354	− 0.00329	0.00120	− 0.00156
	β_4	− 0.00678	− 0.00230	− 0.00098	− 0.00087
Mixed two step	β_1	− 0.00165	− 0.00326	− 0.00049	− 0.00114
	β_2	0.00793	0.00270	0.00102	− 0.00368
	β_3	− 0.01051	− 0.00045	0.00141	− 0.00111
	β_4	0.00109	− 0.00308	− 0.00431	0.00072
PLS-regression	β_1	0.00156	0.00018	− 0.00361	− 0.00307
	β_2	0.00713	− 0.00062	− 0.00193	0.00091
	β_3	− 0.00055	− 0.00476	0.00198	− 0.00291
	β_4	− 0.01307	0.00169	0.00074	− 0.00029

Table 10 Relative Bias with respect to formative-reflective framework

Approach	Path	100	250	500	1000
Repeated indicators	β_1	0.01363	0.01005	0.00458	0.00116
	β_2	0.01455	0.00464	0.00299	0.00044
	β_3	0.00855	0.00387	0.00443	0.00240
	β_4	0.01462	0.00235	0.00226	0.00195
Two step	β_1	0.01187	0.0067	0.00238	0.00101
	β_2	0.01954	0.00563	0.00192	0.00105
	β_3	0.00642	0.00276	0.00220	0.00188
	β_4	0.01095	0.00208	0.00213	0.00182
Mixed two step	β_1	0.01368	0.01003	0.00446	0.00116
	β_2	0.01469	0.00461	0.00306	0.00042
	β_3	0.00847	0.00392	0.00447	0.00243
	β_4	0.01455	0.00238	0.00230	0.00196
PLS-regression	β_1	0.01637	0.00504	0.00271	0.00102
	β_2	0.01509	0.00500	0.00217	0.00043
	β_3	0.01321	0.00205	0.00160	0.00137
	β_4	0.01221	0.00226	0.00302	0.00171

Table 11 Relative Bias with respect to reflective-reflective framework

Approach	Path	100	250	500	1000
Repeated indicators	β_1	− 0.00050	− 0.00045	0.00124	− 0.00008
	β_2	0.00717	0.00229	0.00018	0.00017
	β_3	0.00502	0.00078	0.00246	0.00130
	β_4	0.00897	0.00102	0.00023	0.00058
Two step	β_1	0.00087	0.00032	0.00099	0.00013
	β_2	− 0.00056	− 0.00194	− 0.00023	− 0.00034
	β_3	0.00439	0.00056	0.00195	0.00112
	β_4	0.00546	0.00079	0.00012	0.00044
Mixed two step	β_1	− 0.00056	0.00088	0.00201	− 0.00024
	β_2	0.00723	0.00164	− 0.00028	0.00026
	β_3	0.00531	0.00114	0.00206	0.00134
	β_4	0.00966	0.00066	0.00056	0.00061
PLS-regression	β_1	− 0.00016	0.00198	0.00146	− 0.00016
	β_2	0.00671	0.00059	0.00031	0.00008
	β_3	0.00367	0.00019	0.00235	0.00152
	β_4	0.00967	0.00116	0.00003	0.00059

Table 12 Relative Bias with respect to formative-formative framework

Approach	Path	100	250	500	1000
Repeated indicators	β_1	− 0.00949	− 0.00422	− 0.00222	− 0.00210
	β_2	− 0.00947	− 0.00427	− 0.00221	− 0.00210
	β_3	− 0.00940	− 0.00431	− 0.00220	− 0.00209
	β_4	− 0.00936	− 0.00426	− 0.00225	− 0.00209
Two Step	β_1	− 0.00795	− 0.00299	− 0.00185	− 0.00201
	β_2	− 0.00567	− 0.00444	− 0.00211	− 0.00188
	β_3	− 0.00880	− 0.00322	− 0.00190	− 0.00199
	β_4	− 0.00679	− 0.00400	− 0.00202	− 0.00166
Mixed two step	β_1	− 0.00921	− 0.00433	− 0.00221	− 0.00212
	β_2	− 0.00949	− 0.00414	− 0.00225	− 0.00211
	β_3	− 0.00954	− 0.00443	− 0.00220	− 0.00212
	β_4	− 0.00945	− 0.00414	− 0.00221	− 0.00199
PLS-regression	β_1	0.00261	− 0.00788	− 0.00386	− 0.00650
	β_2	− 0.02790	− 0.00143	− 0.00302	− 0.00101
	β_3	0.00811	0.00118	− 0.00152	0.00263
	β_4	− 0.02214	− 0.00901	− 0.00114	− 0.00403

References

Andreev, P., Heart, T., Maoz, H., & Pliskin, N. (2009) . Validating formative partial least squares (PLS) models: Methodological review and empirical illustration. ICIS 2009 proceedings, 193.

Aria, M., Capaldo, G., Iorio, C., Orefice, C. I., Riccardi, M., & Siciliano, R. (2018). PLS Path Modeling for causal detection of project management skills: A research field in National Research Council in Italy. *Electronic Journal of Applied Statistical Analysis*, *11*(2), 516–545.

Bagozzi, R. P. (2007). *On the meaning of formative measurement and how it differs from reflective measurement: Comment on Howell, Breivik, and Wilcox (2007)*. Washingdon. DC: American Psychological Association.

Blalock, H. M. (1982). Conceptualization and measurement in the social sciences.

Becker, J.-M., Klein, K., & Wetzels, M. (2012). Hierarchical latent variable models in PLS-SEM: Guidelines for using reflective-formative type models. *Long Range Planning*, *45*(5–6), 359–394.

Bollen, K. A. (1989). *Structural equations with latent variables*. New York: Wiley.

Bollen, K., & Lennox, R. (1991). Conventional wisdom on measurement: A structural equation perspective Psychological bulletin. *American Psychological Association*, *110*(2), 305.

Bollen, K. A., & Ting, K. F. (2000). A tetrad test for causal indicators. *Psychological methods*, *5*(1), 3.

Cataldo, R., Grassia, M. G., Lauro, N. C., & Marino, M. (2017). Developments in Higher-Order PLS-PM for the building of a system of Composite Indicators Quality & Quantity. *Springer*, *51*(2), 657–674.

Cataldo, R., Crocetta, C, Grassia, M. G, Lauro, N. C., Marino, M., & Voytsekhovska, V. (2020). Methodological PLS-PM framework for SDGs system Social Indicators Research, Springer, 1–23.

Cheah, J. H., Ting, H., Ramayah, T., Memon, M. A., Cham, T. H., & Ciavolino, E. (2019). A comparison of five reflective-formative estimation approaches: Reconsideration and recommendations for tourism research. *Quality & Quantity*, *53*(3), 1421–1458.

Chin, W. W. (1998a). Issues and opinion on structural equation modeling. *MIS Quarterly*, *22*(1), vii–xvi.

Chin, W. W. (1998b). The partial least squares approach to structural equation modeling. In G. A. Marcoulides (Ed.), *Modern business research methods* (pp. 295–336). Mahwah: Lawrence Erlbaum Associates.

Chin, W. W. (2010). How to write up and report PLS analyses. In V. V. Esposito, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares (PLS): Concepts, methods and applications* (pp. 655–690). Berlin: Springer.

Chin, W. W., Marcolin, B., & Newsted, P. (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. *Information systems research, INFORMS*, *14*(2), 189–217.

Ciavolino, E. (2012). General distress as second order latent variable estimated through PLS-PM approach. *Electronic Journal of Applied Statistical Analysis*, *5*(3), 458–464.

Ciavolino, E., Carpita, M., & Nitti, M. (2015). High-order pls path model with qualitative external information. *Quality & Quantity*, *49*(4), 1609–1620.

Ciavolino, E., & Nitti, M. (2013). Simulation study for PLS path modelling with high-order construct: A job satisfaction model evidence. In *Advanced dynamic modeling of economic and social systems* (pp. 185–207). Springer, Berlin, Heidelberg.

Coltman, T., Devinney, T. M., Midgley, D. F., & Venaik, S. (2008). Formative versus reflective measurement models: Two applications of formative measurement. *Journal of Business Research, Elsevier*, *61*(12), 1250–1262.

DeVellis, R. F. (2016). Scale development: Theory and applications (Vol. 26). Sage publications.

Diamantopoulos, A., Riefler, P., & Roth, K. P. (2008). Advancing formative measurement models. *Journal of Business Research*, *61*(12), 1203–1218.

Diamantopoulos, A., & Siguaw, J. A. (2006). Formative versus reflective indicators in organizational measure development: A comparison and empirical illustration. *British Journal of Management*, *17*(4), 263–282.

Diamantopoulos, A., & Winklhofer, H. (2001). Index construction with formative indicators: An alternative to scale development. *Journal of Marketing Research*, *38*(2), 269–277.

Eboli, L., Forciniti, C., & Mazzulla, G. (2018). Formative and reflective measurement models for analysing transit service quality. *Public Transport*, *10*(1), 107–127.

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, American Psychological Association*, *5*(2), 155.

Esposito, V. V., Trinchera, L., & Amato, S. (2010). PLS path modeling: From Foundations to Recent developments and open issues for model assessment and improvement. In V. V. Esposito, W. W. Chin, J. Henseler, & H. Wang (Eds.), *Handbook of partial least squares (PLS): Concepts, methods and applications*. Berlin: Springer.

Hair, J. F, Jr., Sarstedt, M., Ringle, C. M., & Gudergan, S. P. (2017). Advanced issues in partial least squares structural equation modeling. Sage Publications.

Howell, R. D., Breivik, E., & Wilcox, J. B. (2007). Reconsidering formative measurement. *Psychological Methods*, *12*(2), 205.

Jarvis, D., MacKenzie, S., & Podsakoff, P. (2003). A critical review of construct indicators and measurement model misspecification in marketing and consumer research. *Journal of Consumer Research*, *30*(3), 199–218.

Johnson, R., Rosen, C., Djurdjevic, E., & Taing, M. (2012). Recommendations for improving the construct clarity of higher-order multidimensional constructs. *Human Resource Management Review*, *22*(2), 66–72.

Joseph, F., Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2014). *A primer on partial least squares structural equation modeling (PLS-SEM)*. Thousand Oaks: SAGE Publications Inc.

Kock, N. (2013) .WarpPLS 4.0 User Manual, ScriptWarp Systems, Laredo Texas.

Kock, N. (2015). How Likely is Simpson's Paradox in Path Models? *International Journal of e-Collaboration (IJeC)*, *11*(1), 1–7.

Law, K., Wong, C., & Mobley, W. (1998). Toward a taxonomy of multidimensional constructs. *Academy of Management Review*, *23*(4), 741–755.

Lohmöller, J. B. (2013). *Latent variable path modeling with partial least squares*. Berlin: Springer.

Lauro, N. C., Grassia, M. G., & Cataldo, R. (2018). Model based composite indicators: New developments in partial least squares-path modeling for the building of different types of composite indicators. *Social Indicators Research*, *135*(2), 421–455.

MacCallum, R. C., & Browne, M. W. (1993). The use of causal indicators in covariance structure models: Some practical issues. *Psychological Bulletin*, *114*(3), 533.

MacKenzie, S., Podsakoff, P., & 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.

Mazziotta, M., & Pareto, A. (2013). Methods for constructing composite indices: one for all or all for one? *Rivista Italiana di Economia Demografia e Statistica*, LXVIII, n.2.

Mazziotta, M., & Pareto, A. (2019). Use and misuse of PCA for measuring well-being. *Social Indicators Research*, *142*(2), 451–476.

Paxton, P., Curran, P. J., Bollen, K. A., Kirby, J., & Chen, F. (2001). Monte carlo experiments: Design and implementation. *Structural Equation Modeling*, *8*(2), 287–312.

Petter, S., Straub, D., & Rai, A. (2007). *Specifying formative constructs in information systems research* (pp. 623–656). JSTOR: MIS quarterly.

Rajala, R., & Westerlund, M. (2010). Antecedents to consumers' acceptance of mobile advertisements: A hierarchical construct PLS structural equation model. In *XLIIIth Hawaii International Conference on Systems Sciences (HICSS)*.

Reinartz, W., Echambadi, R., & Chin, W. W. (2002). Generating non-normal data for simulation of structural equation models using Mattson's method. *Multivariate Behavioural Research*, *37*(2), 227–244.

Rindskopf, D., & Rose, T. (1988). Second order factor analysis: Some theory and applications. *Multivariate Behavioral Research*, *23*(1), 51–67.

Roni, S. M., Djajadikerta, H., & Ahmad, M. A. N. (2015). PLS-SEM approach to second-order factor of deviant behaviour: Constructing perceived behavioural control. *Procedia Economics and Finance*, *28*, 249–253.

Roy, S., Tarafdar, M., Ragu-Nathan, T. S., & Erica, M. (2012). The effect of misspecification of reflective and formative constructs in operations and manufacturing management research. *Electronic Journal of Business Research Methods*, *10*(1).

Sanchez, G. (2013). *PLS Path Modeling with R*. Berkeley: Trowchez Editions.

Simonetto, A. (2012). Formative and reflective models: State of the art. *Electronic Journal of Applied Statistical Analysis*, *5*(3), 452–457.

Tenenhaus, M., Amato, S., Esposito, & Vinzi, V. (2004). A global goodness-of-fit index for PLS structural equation modelling. In *Proceedings of the $XLII^{th}$ SIS Scientific Meeting*, 1, 739–742.

Tenenhaus, M., Esposito, V. V., Chatelin, Y. M., & Lauro, N. C. (2005). PLS path modeling. *Computational Statistics and Data Analysis*, *48*(1), 159–205.

Wagner, C. H. (1982). Simpson's paradox in real life. *The American Statistician*, *36*(1), 46–48.

Wetzels, M., Odekerken-Schröder, G., & Van Oppen, C. (2009). *Using PLS path modeling for assessing hierarchical construct models: Guidelines and empirical illustration* (pp. 177–195). JSTOR: CMIS quarterly.

Wilson, B., & Henseler, J. (2007). *Modeling reflective higher-order constructs using three approaches with PLS path modeling: a Monte Carlo comparison*. Otago: Department of Marketing, School of Business, University of Otago.

Wold, H. (1982). Soft modeling: The basic design and some extensions. *Systems under Indirect Observation*, *2*, 343.

Affiliations

Corrado Crocetta[1] · Laura Antonucci[1] · Rosanna Cataldo[2] · Roberto Galasso[2] · Maria Gabriella Grassia[2] · Carlo Natale Lauro[3] · Marina Marino[2]

Laura Antonucci
laura.antonucci@unifg.it

Rosanna Cataldo
rosanna.cataldo2@unina.it

Roberto Galasso
robe.galasso@gmail.com

Maria Gabriella Grassia
mgrassia@unina.it

Carlo Natale Lauro
clauro@unina.it

Marina Marino
mari@unina.it

[1] Department of Economics, University of Foggia, Largo Papa Giovanni Paolo II, 71121 Foggia, Italy

[2] Department of Social Science, University "Federico II", Vico Monte della Pietá, 80138 Naples, Italy

[3] Department of Economics and Statistics, University "Federico II", Via Cintia, Monte Sant'Angelo, 80126 Naples, Italy