
Bradley Wilson, RMIT University, School of Applied Communication. Jörg Henseler, Radboud University, Nijmegen School of Management, The Netherlands.

Abstract

Many studies in the social sciences are increasingly modeling higher-order constructs. PLS can be used to investigate models at a higher level of abstraction (Lohmöller, 1989). It is often chosen due to its’ ability to estimate complex models (Chin, 1998). The primary goal of this paper is to demonstrate the relative robustness of various item and construct characteristics on the reproduction of parameter true scores when utilising the two-stage, hierarchical components indicators (repeated indicators) and a newer hybrid technique (where indicators are not repeated). Our Monte Carlo study mirrors a simple substantive branding example. We vary pertinent dimensions such as: sample size, differing item reliabilities and inner weighting schemes. Our contribution is twofold. Firstly, we provide an overview of the approaches to model reflective second-order constructs with PLS. Secondly, based on our simulation, we provide suggestions when to use each approach.

Three PLS-based Approaches to Estimating Path Models with Higher-Order Constructs

In the extant literature, two approaches have been suggested, the Two-Stage Approach, and the Hierarchical Components Approach. Additionally, we propose a newer Hybrid Approach. We acknowledge that is some definitional concern regarding the use of the terminology component, factor and latent variables that affect the area of PLS, Principal Component Analysis, Covariance-Based Structural Equation Modeling (CBSEM) and Exploratory Factor Analysis domains (du Toit, du Toit, Joreskog and Sorbom 1999; Pedhazar and Pedhazur, 1991). We understand these issues but for the purposes of this paper we refer to constructs as latent constructs or variables interchangeably. Others have also followed this convention in PLS writing and reporting (Henseler, and Fassott, 2007; O'Cass, 2002). Chin, Marcolin and Newsted, 2003, p. 197) refer to PLS as a “components-based structural equation modeling technique.” CBSEM and PLS methods are seen as complementary methods (Barclay, Higgins and Thompson, 1995). A description of the three approaches investigated is presented next.

The Two-Step Approach

The two-stage approach is when latent variable scores are initially estimated without the second-order construct present, but with all of the first-order constructs only within the model (Agarwal and Karahanna, 2000, Henseler, Wilson, Götz and Hautvast, 2007). The latent variable scores are subsequently used as indicants in a separate higher-order structural model analysis. Hence, a two-stage approach. This is typical of how analysts previously used factor scores prior to running further regression analyses. It may offer advantages when estimating higher-order models with formative indicants (Diamantopoulos and Winklhofer, 2001; Reinartz, Krafft, and Hoyer, 2004). A clear disadvantage of any two-stage approach is that any construct that is investigated in stage two is not taken into account when estimating the latent variable scores at stage one. This could encourage “interpretation confounding” (Burt, 1973). Similar arguments have followed the use of the two-step modeling approach advocated
by Anderson and Gerbing (1988) in the CBSEM literatures. The implementation is not one simultaneous PLS run. Falk and Miller (1992) talk of the advantage of PLS estimation in that it takes into account ‘its nearest neighbor’ during iteration. To follow such an approach may not fully capitalise on the “consistency at large” assumption that PLS is based around. PLS can overcome some of the problems of “indeterminancy” experienced when using CBSEM techniques (Falk and Miller, 1992). We assume that the reader is familiar with PLS estimation procedures.

The Hierarchical Components Approach

The hierarchical components model was suggested originally by Wold (1982) (see also Lohmöller, 1989, p. 130-133; Chin et al. 2003). Also known as the Repeated Indicators Approach (Lohmöller, 1989; Wold, 1982) or Superblock Approach (Tenenhaus, Esposito Vinzi, Chatelin and Lauro, 2005) is the most popular approach when estimating higher order constructs with PLS (Venaik, 1999; Wilson, 2007; Zhang, Li, and Sun, 2006). “A second order factor is directly measured by observed variables for all the first order factors. While this approach repeats the number of manifest variables used, the model can be estimated by the standard PLS algorithm (Reinartz, Kraft and Hoyer, 2003, p. 19).” The manifest indicators are repeated to also represent the higher order construct.

![Figure. 1. Conceptual Representation of Hierarchical Components Model.](image)

Latent scores are saved during analysis to be used within future analyses. At this point in time, this approach appears to be the one most favoured by analysts when using PLS to model higher constructs. We believe this is because it has been presented most clearly by key prominent PLS methodologists (e.g., Wold and Lohmöller). A disadvantage of this approach is that there is a perceived effect of possibly biasing the estimates by relating variables of the same type together via PLS estimation. That is, the exogeneous variables in effect becomes the endogenous variables. We are yet to see any Monte Carlo study that compares the performance of various modeling approaches.

The Hybrid Approach

The Hybrid Approach builds on an idea of Wold (1982) originally meant for modeling nonlinear structural relationships. The hybrid approach was inspired by and adapted from the work of Marsh, Wen, and Hau (2006) within the CBSEM literature when investigating interactions and quadratic effects. They argue when creating product terms that “each of the multiple indicators should only be used once in the formation …(cross-product terms)., to avoid creating artificially correlated residuals when the same variable is used in the construction of more than one product terms. p. 245”. To implement this technique within PLS would randomly split all variables so that half are represented on their respective first order construct side and the other half of indicants are represented on the second order construct side (E.g., items are not repeated). So for instance, if we are to alter Figure 1
slightly, this would mean that item 1 and 3 may still measure the two first order constructs and items 2 and 4 would be for the higher order construct. We believe this approach has not been trialed in PLS and could overcome the criticism that is directed towards the hierarchical components model in that the indicators are repeated and therefore via PLS iteration and estimation the analyst could be in some way relating the same items together. Naturally, the hybrid approach circumvents this criticism. During the runtime of the algorithm, the second-order construct is generated by a proxy which is then assigned to the second-order construct (to derive latent variable scores and path coefficients).

We believe, the need for research is clear given that researchers are using PLS with increasing regularity (Dawes, Lee and Dowling, 1998; Fornell et al., 1996; O’Cass and Fenech, 2003). Numerous approaches exist yet the analyst knows little about performance and robustness properties of technique for these model types. Previous studies provide limited direction. This study also makes a valuable contribution by investigating a newer hybrid approach.

**An Example from Brand Management**

Prior to this Monte Carlo study there has been no formal guidance for social researchers modeling higher order constructs with PLS. The researcher often just selected what procedures other researchers had followed in the past. Here is one such small example. Fournier (1994) after extensive qualitative research developed an item battery to measure the quality of the person-brand bond. She termed this Brand Relationship Quality. “Brand relationship quality (BRQ) is best thought of as a customer-based indicator of the strength and depth of the person-brand relationship (Fournier, 1994, p. 124)”. We illustrate only the repeated indicators approach here to reserve further presentation space. For a description of the sample characteristics readers should consult Wilson (2007). There were two main item batteries (BRQ Scale: 62 items, Future Purchase Intention: 10 items) utilised. Fournier supplied an extended BRQ scale version. The items used a 7-point scale. All measures were treated as being reflective in keeping with the initial mode they were specified. A final sample size of 1290 was obtained (25.8% response rate). The example was analysed with SPAD 6.0. Adequate unidimensionality and appropriate construct and discriminant validity was established (Wilson, 2007). In this example, 27.69% of the variation in intention is explained by BRQ. The path coefficient between BRQ and intention indicates a fairly strong positive impact ($\beta = 0.5254$). Falk and Miller (1992) would indicate that this is a significant finding in social research. Presentation space is reserved for the Monte Carlo design and selected results.

**A Monte Carlo Simulation Study**

In order to complete this project we created our own implementation of the PLS algorithm. We used R 2.3.1 (R Development Core Team, 2006). Firstly, we define an underlying true population model (see figure 2) and determine the factor attributes for the Monte Carlo design. Secondly, we generated random data that emerges from the model parameters. Thirdly, given the random data, we let each PLS approach estimate each model under each factor combination. The simulation model mirrors our BRQ $\rightarrow$ future intentions example that is presented previously. In this study, 1000 Monte Carlo samples per condition were run resulting in 216,000 observations.

Fixed Factor:

1. Approach (Two-Stage; Repeated Indicators; Hybrid)
We used the following random factors (with categories) in the Monte Carlo experiment:
2. effect size of the effect of the second order construct on first-order construct ($f^2$: 0.02, 0.15, 0.35)
3. loadings ($\lambda$) in the model, including loadings between first order constructs and second order construct (Square root of 0.5 (ca. 0.7071); square root of 0.75 (ca. 0.8660))
4. number of observations or sample size (50, 200, 800)
5. number of indicators ($k$: 4, 8)
6. inner weighting scheme (Centroid Scheme, Factor Weighting Scheme)

The three approaches were compared on the basis of their capability to:
- capture the true relationship parameter between second-order $\xi$ and $\eta$
- create reliable latent variable scores for both $\xi$ and $\eta$
- predict accurately the endogenous variable $\eta$

![Population Model](image)

**Figure 2. True Population Simulation Model**

**Results and Discussion**

The PLS estimation outcomes have been measured for each run. Extensive analysis and reporting is currently in progress and full results will be presented at the conference. We are aiming to increase the complexity of our design prior to the conference by adding cells such as expanding the number of indicators, loadings and parameters (homogenous and heterogeneous), great sample size and investigating maybe another model structure to allow for greater contrasts. The results in short at this preliminary stage of analysis include:

- The reliability of the second-order construct is almost completely independent of the chosen modelling approach (significant at the 10%-level only), and depends only on the item loadings ($p<.01$), which was to be expected.
- We see that for small effects, the reliability-corrected hybrid approach and the reliability-corrected two-stage approach deliver the most consistent estimates. Particularly in the case of fewer indicators (4), these approaches highly outperform the repeated indicators approach in this regard. The same results pattern holds true for medium effects. The repeated indicators approach performs best when the number of indicants is larger and at determining larger effects sizes.
- In terms of consistency, we see that the centroid scheme is best chosen for small effects and that the factor weighting scheme is best utilised for medium effects. For large effects, the choice of the weighting scheme is of limited influence.
- Results reveal that with small sample sizes (50), the repeated indicators approach is not able to deliver consistent estimates for small effects.
- The centroid scheme performs poorly for small sample sizes (50), but is not likely to overestimate effects as does the factor weighting scheme.
• The repeated indicators approach delivers heavily biased estimates under less reliable measurement compared with the other approaches. With higher loadings this approach starts to improve dramatically for medium and large effects and in some respects is superior.
• There has been adequate convergence for all methods using different inner weighting schemes with the PLS R programme.

The above preliminary results should give the reader an idea on what will be fully reported at the conference. Even these preliminary results provide considerable guidance on what approach to follow when faced with different study factors. Structural equation analysts are often trying to find modeling solutions to these problems with literature that provides limited consolidated guidance. This is definitely a first step in what will be a more concerted research agenda with a more complex research design.

This work is not without limitations. A Monte Carlo design can provide guidance for the chosen population model under study. Often social researchers are selecting structural equation models of greater complexity to investigate so the chosen population model may not mirror our one structural path design. Also, Chin and Newsted (1999, p. 325) state that the PLS estimates are "inconsistent" relative to the CBSEM model due to the fact that they are aggregates of the observed variables, which in part includes measurement error. The estimates will approach the "true" latent variable scores as both the number of indicators per block and the sample size increases.” Our preliminary work confirms these findings. The point at which this happens is of interest within our Monte Carlo work to provide pragmatic guidelines for appropriate use of PLS for different model types. Chin and Newsted (1999) go on to emphasize that more indicants and also more cases are needed to improve PLS estimation robustness. We also can confidently state that our results also indicate this is the case. The hybrid approach although operating against these principles compared with the hierarchical components approach showed relatively robust performance. This may offer the modeller a new choice that has not been previously considered. Our design needs to be improved to have smaller sample size graduations (at the lower range), smaller number of indicator graduations, testing different distributional patterns and non-homogeneous outside loading patterns. Such work is now in progress.

As we limited our study to PLS-based approaches, other structural equation modeling techniques CBSEM [summated scales and congeneric models (Joreskog, 1971)] were not considered. Some researchers have used CBSEM in estimating congeneric models before using PLS for the structural modeling stages (Grace and O’Cass, 2005). This approach is also worthy of future exploration within a Monte Carlo design to act as another basis for comparison.

As we studied only reflective measurement models, it remains unclear whether our results will be generalised to PLS path models with formative measures where PLS has significant demonstrated advantages (Ringle, Wilson and Götz, 2007; Vilares, Almeida and Coelho, 2007). All higher order model structure combinations represented within the Jarvis et al. (2003) study need to be investigated. That is, all model types need to have separate PLS Monte Carlo studies completed.

We believe that researchers need to approach tasks with more methodological certainty given the increasing popularity for investigating complex higher order structural models. It is our contention that such research will increasingly be undertaken even if it is in a model generation exploratory capacity. Practically, market researchers are looking for such flexibility. We believe that Monte Carlo work within PLS is some way behind what has been established in the CBSEM domain (Boomsma, 1983; Gerbing and Anderson, 1993; Hu and Bentler, 1999; Tanaka, 1987). Some recent studies in PLS are aiming to alter this dearth in
knowledge to provide researchers with greater guidance (Cassel, Hackl and Westlund, 1999; Henseler, Wilson and Dijkstra, 2007; Tenenhaus et al., 2005). We encourage others to follow a similar research agenda to expand the body of PLS knowledge.
References


For an excellent explanation of the PLS algorithm its' respective estimation in establishing outside and inside parameter estimates see (Chin, 1998; Chin and Newsted, 1999; Fornell and Cha, 1994; Tenenhaus et al., 2005).

We realise that this separation of items could occur in many ways. This is worthy of future investigation. This in analogous to all the approaches that have been tried and tested when using parcelling or testlets within the CBSEM literature (Landis, Beal and Tesluk, 2000).

This criticism cannot be referenced but has come about from discussions with the PLS academic community. Some researchers currently sparingly use the hierarchical components approach as they believe the “jury is out” on its’ overall performance. Others avoid investigating models with higher-order relations altogether at this time with PLS. However, some researchers are starting to use PLS when investigating higher order conceptual models.

The first author would like to acknowledge and thank Professor Fournier for her initial support and inspiration.

We acknowledge it is difficult to extend such research to incorporate LISREL approaches for higher order constructs at the item level due to the inability to use the repeated indicators approach. There are also many complex identification and estimation problems (Chen et al., 2001; Dillon, Mulani and Kumar, 1987; Rindskopf, 1984) that can become more prevalent with the use of CBSEM with higher order constructs.