**ORIGINAL PAPER**

# Partial possibilistic regression path modeling: handling uncertainty in path modeling

**Rosaria Romano[1]** · **Francesco Palumbo[2]**

## Abstract

The paper presents a new insight of a recently proposed method named partial possibilistic regression path modeling. This method combines the principles of *path modeling* with those of possibilistic regression to model the net of relations among blocks of variables, where a weighted composite summarizes each block. It assumes that randomness can refer back as the measurement error, which is the error in modeling the relations between the observed variables and the corresponding composite, and the vagueness to the structural error, which is the uncertainty in modeling the relations among the composites behind each block of variables. The comparison of the proposed method with a classical composite-based path model is based on a simulation study. A case study on the use of Wikipedia in higher education illustrates a fruitful usability context of the proposed method.

**Keywords** Interval data · Randomness–vagueness · Structural equation modeling · Least absolute values

## 1 Introduction

Socio-economic and natural systems can be defined as having complex relationships between sets (or blocks) of variables. *Regression analysis* is likely the statistical method most widely used to study the dependencies between two sets of variables. However, when the phenomenon increases in complexity, a single equation model becomes inadequate for analyzing and describing the data dependence structures. Complex as the mathematical model may be, it is approximate and can account for a

✉ Rosaria Romano
rosaroma@unina.it

Francesco Palumbo
fpalumbo@unind.it

[1] Department of Economics and Statistics, University of Naples Federico II, Naples, Italy

[2] Department of Political Sciences, University of Naples Federico II, Naples, Italy

simplified abstraction of the reality. Therefore, classical statistical models are developed under the following paradigm (Judd et al. 2011):

$$DATA = MODEL + ERROR.$$

Regression analysis can model only direct relationships between independent and dependent variable(s). It strongly limits the variables in that they cannot have any indirect effects on each other. The path analysis approach (Tukey 1964; Alwin and Hauser 1975) offers a way to overcome such a limitation by allowing to model a set of relationships between observed variables. In other words, path analysis is constructed through a system of simultaneous simple or multiple regressions. Several approaches exist to path models; the most known is structural equation modeling (SEM) (Bollen 1989; Kaplan 2008). SEM combines the ideas behind path analysis with the basic principles of confirmatory factor analysis (Thurstone 1931), which presumes that fewer factors than the number of observed variables are responsible for the shared variance-covariance matrix. Through this analysis, SEM carries the idea that different subsets or blocks of variables are expressions of different concepts. Then latent variables (LVs) are those that cannot be directly observable but are measurable through a set of manifest variables (MVs). The relationships between the LVs define the structural or inner model, whereas the relationships between each LV and its block of MVs define the measurement or outer model. Partial least squares (PLS) represents an alternative approach to path models named PLS path modeling (PLSPM). Like the SEM, PLSPM aims to study the relationships among blocks of MVs, thought PLS estimates the net of linear relations among the blocks through a system of independent equations based on simple and multiple regressions. A peculiarity of PLSPM is that a weighted composite summarizes each block of MVs (Bollen and Bauldry 2011), and the relations among the composites define the structural model. The underlying assumption is that the relevant information regarding the relationships within and between the blocks of MVs are carried through the composites (Dolce et al. 2018). Although PLSPM has been increasing in popularity, it has also been strongly criticized as an alternative *estimation* method of a SEM model (Rönkkö and Evermann 2013; Rönkkö 2014). On the other hand, a recently proposed modified PLSPM allows overcoming some of these weaknesses (Dijkstra and Henseler 2015b, a), and most recently, a new model has been proposed (Dijkstra 2017).

Specifically, in PLSPM, there are two different sources of errors that correspond to the outer model and the inner model, respectively. Residuals represent the gap between the precise mathematical model and reality and are the estimates of the error terms. According to conventional rules, a model fits the data (and it is then useful for describing the reality) if the total residuals are less than a priori defined thresholds. Dealing with two different types of residuals, this choice of the threshold may become arbitrary, and the results interpretation meaningless. The present work aims to consider the innovative approach within the path models framework that is named partial possibilistic regression path modeling (PPRPM) (Romano and Palumbo 2016, 2017), which separately but consistently considers the *measurement* and *structural model errors*. More properly, the following two types of errors are defined: (i) a *measurement model error* that refers to a variability not explained by the composite that the corresponding

MV is supposed to measure (Romano and Palumbo 2017) and (ii) a *structural model error* that refers to a disturbance in the prediction of the dependent composites by the respective predictors. Therefore, measurement model residuals refer to the MVs and can be interpreted using the usual reading key, but structural model residuals cannot; structural model residuals refer to the composites that cannot be directly observed or that are assumed to be determinations of known random variables: these residuals represent the model's inadequacy in describing relationships between composites. In other words, the structural model residuals account for the model's inadequacy to represent real-world complexity.

In statistical reasoning, one cannot disregard the uncertainty that comes with reasoning under partial knowledge using mathematical models (Coppi 2008). Partial knowledge may be due to different causes: data sampling errors, mathematical model inadequacies, and/or fragmented knowledge of causal relationships. Bezdek (1981) describes the uncertainty in mathematical models as belonging to three categories: (i) inaccurate measurements, (ii) random occurrences, and (iii) vague descriptions. In the present study, the researchers assigned the measurement error to the first two categories and the structural error to the third category. Therefore, the solutions of the inner model parameters can be assumed to belong to epistemic sets. According to Couso and Dubois (2014), an epistemic set roughly captures information about a population via observations. It is assumed that the true parameters can be estimated by a random variable that takes values in a given set, but the estimator probability distribution is unknown. Readers interested in the debate between *epistemic* and *ontic* sets may refer to Couso and Dubois (2014) and Dubois (2014). Typically an epistemic model delivers an imprecise output, given the available incomplete information. The present work proposes an epistemic modeling in the context of interval-based representations. Actually, interval valued data can be interpreted as a special case of an $LR_2$ fuzzy number $\tilde{X}$, with membership function $\mu_{\tilde{X}}(z)$ (for $0 \leq z \leq 1$), and defined by four parameters, namely the left center ($c_1$), the right center ($c_2$), the left spread ($l > 0$) and the right spread ($r > 0$):

$$
\mu_{\tilde{X}}(z) = \begin{cases} L\left(\frac{c_1 - z}{l}\right) & z \leq c_1, \\ 1 & c_1 \leq z \leq c_2, \\ R\left(\frac{z - c_2}{r}\right) & z \geq c_2. \end{cases}
$$

Note that $\{L(z), R(z)\} : \mathbb{R} \to [0, 1]$, then if $c_1 = c_2$ and $l = r = 0$ we get an interval (Ferraro and Giordani 2017).

In the proposed approach, measurement model residuals refer to the random occurrence in the data, whereas interval-valued structural model estimations account for the model's vague description of the real world in the epistemic view.

The present work considers an alternative approach within the path models framework: partial possibilistic regression path modeling (PPRPM) (Romano and Palumbo 2016, 2017). The most innovative aspect of PPRPM is the different method it offers for considering uncertainty in the measurement and structural models. On the one hand, this method gives rise to possibilistic regression (PR) (Tanaka and Guo 1999), which accounts for vagueness in the parameters that govern the system structure by yielding

interval path coefficients. On the other hand, this method allows for the least absolute deviation (LAD) regression (Bloomfield and Steiger 1983), which accounts for inaccurate measurements and random occurrences. The work aims to deepen the study of the potential of the PPRPM also through a simulation study that allows evaluating the effect of various measurements and structural errors.

This paper is organized as follows: In Sect. 2, the methodological framework is introduced; in Sect. 3, the proposed method is explained; a simulation study and a practical application of the proposed method are discussed in Sects. 4 and 5, respectively; and finally, in Sect. 6 concluding remarks are made.

## 2 Background

PPRPM depends on the philosophy of the composite-based approach. Each composite is calculated as a weighted aggregate of its corresponding MVs, and its partial result is then used to define the net of relationships within and between the blocks of MVs. Among the composite-based models, PLSPM (Tenenhaus et al. 2005; Wold 1975) is the closest to the formalization presented in this paper. However, notwithstanding that PLSPM is based on the least square criterion, this does not prevent us from adopting (almost) the same notation as in (Vinzi et al. 2010) in the rest of the paper. Differently from the classical composite-based methods, PPRPM introduces two novelties: (i) the relationships in the measurement model are defined through least absolute deviation regression (Bloomfield and Steiger 1983) to model the randomness; (ii) the relationships in the structural model are described through possibilistic regression (Tanaka and Guo 1999) to model the vagueness. The remainder of this section consists of three subsections: Sect. 2.1 PLSPM algorithm, which is the related method used as a comparison; Sect. 2.2 a brief description of the LAD regression and the; Sect. 2.3 PR formalization.

### 2.1 Partial least squares path modeling

According to the notation used in Vinzi et al. (2010), the typical data structure in a PLSPM is composed by $H$ ($h = 1, 2 \ldots, H$) blocks of manifest variables recorded on a set of $N$ statistical observations. Each block consists of $P_h$ variables, where the generic variable of the generic block is referred to as $\mathbf{x}_{p_h}$, $p_h = 1, 2, \ldots, P_h$, for any $h \in H$, so that $\sum_h P_h = P$. The $H$ blocks of variables are arranged into a partitioned data matrix

$$\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_h, \ldots, \mathbf{X}_H],$$

where $\mathbf{X}_h$ is the generic block. According to the net of relations among the composites, each block $\mathbf{X}_h$ can be referred to as *endogenous* or *exogenous*. In the structural model, the endogenous composites $\eta$ are dependent variables, while the exogenous composites $\xi$ are independent variables. In the following equations, it is assumed without loss of generality that the composites and the MVs are scaled to zero mean and unit variance

so that the location parameters can be eliminated. The basic structural equation model (Bollen 1989) can be described by the following equations:

$$\mathbf{x}_{p_h} = w_{p_h} \eta_h + \epsilon_{p_h}, \tag{1a}$$

$$\mathbf{x}_{p_h} = w_{p_h} \xi_h + \delta_{p_h}, \tag{1b}$$

$$\eta_h = B\eta_{h'} + \Gamma\xi_{h''} + \zeta_h, \tag{1c}$$

where $h \neq h' \neq h''$ may refer either to endogenous or exogenous composites. The equations defined in (1a) and (1b), which are also referred to as outer relations, form the measurement model, and the generic equation defined in (1c) formalizes the structural model. This equation is referred to as inner relation. Here the $w_{p_h}$ is the loading associated to the $p_h$ MV in the $h$ block, $B$ and $\Gamma$ correspond to the path coefficients linking the corresponding $\eta$ and $\xi$ composites. Although in the classic SEM notation the variables associated with the endogenous composite are commonly referred to as $\mathbf{y}_{p_h}$, here it is preferred to keep the $\mathbf{x}_{p_h}$ notation consistently with the previous notation in which the various $\mathbf{X_h}$ blocks were defined as constitutive elements of the partitioned matrix $\mathbf{X}$.

Regarding the error terms, $\zeta_h$ represents errors in the inner relations (i.e., disturbances in the prediction of endogenous composites), whereas $\epsilon_h$ and $\delta_h$ represent imprecision in the measurement process.

In PLSPM, an iterative procedure permits estimations of the composite scores and loadings, while structural coefficients are obtained from ordinary least square regressions of the estimated composites. As PLSPM notation makes no distinction between endogenous and exogenous composites or between blocks of MVs (Vinzi et al. 2010), the following equations refer to any block of MVs as $\mathbf{X}_h$ and each composite as $\xi_h$, where $h = 1, \ldots, H$.

The algorithm computes the composite scores by alternating the outer and inner estimations until convergence. The procedure starts on the centered (or standardized) MVs by choosing arbitrary weights $w_{p_h}$. In the external estimation, the $h$-th composite is estimated as a linear combination of the corresponding MVs:

$$\mathbf{v}_h \propto \sum_{p_h=1}^{P_h} w_{p_h} \mathbf{x}_{p_h} = \mathbf{X}_h \mathbf{w}_h, \tag{2}$$

where $\mathbf{v}_h$ is the standardized outer estimate of the composite $\boldsymbol{\xi}_h$, and the symbol $\propto$ means that the left side of the equation corresponds to the standardized right side. In the internal estimation, the composite is estimated by considering its links with the other $h'$ adjacent[1] composites:

$$\boldsymbol{\vartheta}_h \propto \sum_{h'} e_{hh'} \mathbf{v}_{h'}, \tag{3}$$

where $\boldsymbol{\vartheta}_h$ is the standardized inner estimate of the latent variable $\boldsymbol{\xi}_h$ and the inner weights $e_{hh'}$, according to the so-called *centroid scheme* (Tenenhaus et al. 2005), are

---

[1] In the PLSPM notation, two variables are said to be adjacent if they are related to each other in the same structural equation (Chin et al. 1998).

equal to the sign of the correlation between $\mathbf{v}_h$ and $\mathbf{v}_{h'}$ (where $h, h' = 1, \ldots, H$). Alternative weighting schemes are provided in the original algorithm (Lohmöller 1989). In PLSPM, these first two steps update the outer weights $w_{p_h}$, which are the regression coefficients in the simple regressions of the $p$-th manifest variable of the $h$-th block $\mathbf{x}_{p_h}$ on the inner estimate of the $h$-th latent variable $\boldsymbol{\vartheta}_h$. The outer weights also correspond to the covariances as $\boldsymbol{\vartheta}_h$ becomes standardized:

$$w_{p_h} = cov(\mathbf{x}_{p_h} \boldsymbol{\vartheta}_h). \tag{4}$$

Even for the external weights, the original algorithm provides an alternate scheme (Lohmöller 1989). The algorithm iterates until convergence, which can formally be demonstrated only for one- and two-block models (Lyttkens et al. 1975) in the general case; moreover in the model with more than two blocks, recent works showed that the PLSPM algorithm optimizes different criteria according to the mode chosen for the computation of the outer weights (Hanafi 2007; Tenenhaus and Tenenhaus 2011). After convergence, the structural or path coefficients are estimated through single and multiple linear regressions of the estimated composites:

$$\boldsymbol{\xi}_h = \beta_{h0} + \sum_{h':\boldsymbol{\xi}_{h'} \to \boldsymbol{\xi}_h} \beta_{hh'} \boldsymbol{\xi}_{h'}, + \zeta_h, \tag{5}$$

where $\boldsymbol{\xi}_h$ is the generic dependent composite, and $\beta_{hh'}$ is the generic path coefficient interrelating the $h'$-th independent composites to the $h$-th dependent composite (where $h \neq h'$). The notation $\to$ indicates that $\xi_{h'}$ and $\xi_h$ are adjacent composites.

## 2.2 Least absolute deviation regression

LAD regression aims to study the relationship between a dependent variable $Y$ and a set of predictors $X_1, \ldots, X_m, \ldots, X_M$ through the linear function

$$Y = \lambda_1 X_1 + \cdots + \lambda_m X_m + \cdots + \lambda_M X_M + \varepsilon, \tag{6}$$

where $\lambda_m$ indicates the generic regression coefficient, $X_1$ is a unitary vector, and $\varepsilon$ is the error term. Parameters $\lambda_m$ are estimated by solving the following linear programming problem:

$$\{\lambda_1, \ldots, \lambda_m, \ldots, \lambda_M\} = \arg \min_{\lambda_m} = |Y - (\lambda_1 X_1 + \cdots + \lambda_m X_m + \cdots + \lambda_M X_M)| \tag{7}$$

LAD regression does not have an analytical solving method; thus, an iterative approach is required. However, LAD regression is resistant to effects caused by outliers in the data since it places equal emphasis on all observations (Koenker 2009). Those interested in learning more about LAD are referred to Bloomfield and Steiger (1983) work.

### 2.3 Possibilistic regression

The purpose of PR is to explain a dependent variable as an interval output in terms of the variation in the explanatory variables. Generally speaking, PR defines the relation between one dependent variable $Y$ and a set of $M$ predictors $X_1, \ldots, X_m, \ldots, X_M$ through a linear function containing interval-valued coefficients:

$$Y = \tilde{\omega}_1 X_1 + \cdots + \tilde{\omega}_m X_m + \cdots + \tilde{\omega}_M X_M, \tag{8}$$

where $\tilde{\omega}_m$ denotes the generic interval-valued coefficient, and $\overline{\omega}_m$ and $\underline{\omega}_m$ are the upper and lower bounds, respectively. Interval-valued coefficients, referred to as interval coefficients throughout the rest of this paper, are also defined in terms of the midpoint and the spread (also called the range), $\tilde{\omega}_m = \{c_m, a_m\}$:

$$c_m = \frac{1}{2}(\underline{\omega}_m + \overline{\omega}_m) \quad a_m = \frac{1}{2}(\overline{\omega}_m - \underline{\omega}_m).$$

There are no restrictive assumptions on the model. Any deviations between the data and the linear models were assumed to be caused by the vagueness of the parameters and not by measurement errors, unlike classical statistical regression. This means that there is no external error component in PR. All uncertainties are embedded in the spread of the coefficients, such that PR minimizes the total spread of the interval coefficients

$$\min_{a_m} \sum_{m=1}^{M} \left( \sum_{n=1}^{N} a_m |x_{nm}| \right), \quad \forall\, m = 1, \ldots, M, \ \forall\, n = 1, \ldots, N, \tag{9}$$

under the following linear constraints

$$\sum_{m=1}^{M} c_m x_{nm} + \alpha \sum_{m=1}^{M} a_m |x_{nm}| \geq y_n,$$
$$\sum_{m=1}^{M} c_m x_{nm} - \alpha \sum_{m=1}^{M} a_m |x_{nm}| \leq y_n, \quad \forall n = 1, \ldots, N, \tag{10}$$

satisfying the following conditions: (i) $a_m \geq 0$; (ii) $c_m \in R$; (iii) $x_{n1} = 1$.

The constraints in Eq. (10) guarantee the inclusion of the whole given data set within the estimated boundaries, where $x_{nm}$ represents the generic value of $X_m$ and $n = (1, \ldots, N)$. The degree of possibility $\alpha$ varying in ]0, 1] is a subjective measure that depends on the context; decreasing the $\alpha$ coefficient expands the estimated intervals. In the rest of the paper the $\alpha$ coefficient is set to 1 since this corresponds to the minimum $a_m$. Those interested in learning more about PR, and in the choice of $\alpha$ specifically, are referred to Tanaka and Guo (1999).

Wang and Tsaur (2000) provided a suitable interpretation of the regression interval by proposing an index of confidence (IC), which is similar to the traditional $R^2$ in statistics. The index is defined as the ratio between the *SSR* (*regression sum of squares*) and the *SST* (*total sum of squares*), where the former represents the variation in the interval midpoints between the lower and upper bounds, and the latter measures the total variation in the observed dependent variables between the lower and upper

bounds. Then $IC = SSR/SST$, with $0 \leq IC \leq 1$, and it gives a measure of the variation in $Y$ between $\underline{Y}$ and $\overline{Y}$. A higher IC means that a well-estimated PR is modeled and can support a better prediction. The literature offers several alternatives to fuzzy or interval regression (Diamond 1990; Kim et al. 1996; Marino and Palumbo 2002; Ferraro et al. 2010; Petit-Renaud and Denœux 2004), yet Tanaka and Guo (1999) possibilistic approach remains the only one that estimates interval-valued regression coefficients capable of embedding the entire error component. One of the strongest criticism of PR concerns its sensitivity to anomalous observations. Some procedures for reducing the effect of outliers have been presented (Wang et al. 2015; Nadimi et al. 2013).

## 3 Handling uncertainty in path models: Partial possibilistic regression path modeling

In composite-based SEM, the three residual terms $\epsilon$, $\delta$, and $\zeta$ (see Eqs. (1a)–(1c)) play a crucial role in the modeling process. PLSPM aims to minimize the sum of the residual variances of all the dependent variables in the model, both latent and observed (Vinzi et al. 2010). Without loss of generality, in the following, the two residual terms $\epsilon$ and $\delta$ are attributed to the measurement error. Therefore, $\zeta$ represents the error in the structural (inner) relations (i.e., disturbances in the prediction of endogenous composites), whereas $\epsilon$ and $\delta$ represent the impreciseness in the measurement (outer) process. The randomness is relegated to the measurement model, and in the inner model, the only source of uncertainty is due to the relations among the composites. However, in the model, through the alternate inner and outer estimation of the composite scores, the two error components (structural and measurement) interact and coexist. Then, the error components are never modeled simultaneously in the same equation, and it follows that an analytical formulation of their propagation is not a priori possible (Baudrit et al. 2007). According to the epistemic approach to the partial knowledge, PPRPM treats the vagueness in the prediction of the composites differently from the imprecision in the measurement of MVs. Thus, PPRPM differs from composite-based SEM in that elements in the coefficient matrices, such as $B$ and $\Gamma$ in Eq. 1c, are interval-valued, but vector residual $\zeta$ is no longer included in the model. Therefore, PPRPM gives rise to PR that accounts for the imprecise nature or vagueness in our understanding of the phenomena, by including interval-valued path coefficients in the structural model. Consistently with the structural model estimation that minimizes the sum of the spreads of the interval parameters, the estimation process of the measurement model is based on the least absolute values (LAD). LAD regression minimizes the sum of the absolute values of the residuals. Measurement model residuals estimate the same type of error as in the PLSPM outer model, that is the imprecision in the measurement process. It is worth noting that, the presence of outliers affects the measurement model; therefore, their impact on the structural model is then mitigated by the LAD outer estimation because of its well-known robustness (Dodge 1997; Koenker 2009) that helps significantly to protect the model against the effect of anomalous observations. The thrust of the paper does not concern the sensitivity of the model

but the determination of a precise model that describes the unobserved constructs that take the form of intervals (Dubois 2014).

## 3.1 The algorithm

The PPRPM estimation process is an $L^1$ norm problem that independently minimizes the *sum of the absolute values of the residuals* in the measurement model and the *sum of all the ranges of the interval-valued coefficients* in the structural model. PPRPM follows the same iterative procedure as PLSPM, alternating the inner and outer estimations of the composite scores; however, the algorithm computes the outer weights and the path coefficients in a different way.

The outer weight $w_{ph}$ is the regression coefficient in the LAD regression of the $p$-th MV of the $h$-th block $\mathbf{x}_{ph}$ on the inner estimate of the $h$-th composite $\vartheta_h$:

$$\mathbf{x}_{ph} = w_{ph}\vartheta_h + \epsilon_{ph}. \tag{11}$$

The structural (or path) coefficient is the regression coefficient in the PR among the estimated composites:

$$\tilde{\xi}_j = \tilde{\beta}_{0j} + \sum_{h:\xi_h \to \xi_j} \tilde{\beta}_{hj}\xi_h, \tag{12}$$

where $\xi_j$ ($j = 1, \ldots, J$ and $J < H$) is the generic endogenous (dependent) latent variable and $\tilde{\beta}_{hj}$ is the generic interval path coefficient in terms of the midpoint and the range $\tilde{\beta}_{hj} = \{c_{hj}; a_{hj}\}$, or equivalently $[\underline{\beta}_{hj}, \overline{\beta}_{hj}] = [c_{hj} \pm a_{hj}]$, which interrelates the $h$-th exogenous (independent) variable to the $j$-th endogenous variables (where $h \neq j$). The higher the midpoint coefficient, the higher the contribution to the prediction of the endogenous composite is. The higher the spread of the coefficient, the higher the vagueness in the relation among the composites is.

In PPRPM, the model can be validated using the same criteria defined in the PLSPM framework. In particular, this criterion applies to the assessment of the measurement model, which can be validated through the *communality index* (Tenenhaus et al. 2005). However, the same reasoning cannot be extended to the validation of the structural model, and even less so to that of the global model. In PPRPM, each structural equation is modeled with PR, which includes the error term in its parameters; thus, there is no residual term. The quality of the model can be measured here with the IC index presented in Sect. 2.3.

About the whole computational complexity of the algorithm, the outer and inner estimations are obtained by solving a linear programming (LP) problem through the simplex algorithm. Specifically, in the outer model estimation, the algorithm independently computes $P$ LAD regressions, and each solution has a computational complexity equal to $O(\ln N)$ (where $N$ refers to the total number of observations). The inner model estimation algorithm solves as many LP problems as the number of structural equations with the simplex algorithm. Each problem involves a simplex algorithm with the computational complexity depending on $N$ and on the number of
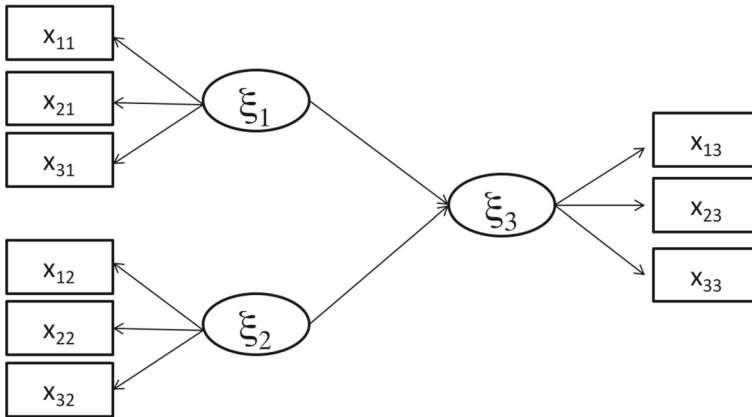
**Fig. 1** Path diagram of the structural and measurement models in the simulation study

independent variables according to $O(\ln mN)$, where $m$ refers to the number of predictors. The whole algorithm (the outer and the inner step) iterates until convergence is satisfied. According to the evidence from the simulation studies (see the following section), the whole algorithm requires *negligible* computational time, as the complexity is of $\ln N$ order, and the convergence is empirically achieved.

## 4 A simulation study

The simulations focus on the effect of various measurement and structural errors, combined with different degrees of skewness and diverse sample size. They were mainly based on four previous studies (Cassel et al. 1999, 2000; Westlund et al. 2001; Vilares et al. 2010).

The sensitivity of the results was investigated with respect to:

- Skewness (symmetric, highly skewed).
- Sample size (50, 500).
- Level of noise in the structural model ($\zeta$: 10, 30%).
- Level of noise in the measurement model ($\epsilon$: 10, 30%).

The PLSPM and PPRPM estimations of the structural and measurement models were compared in terms of bias and precision (mean squared error (MSE)). The distribution of the estimated scores of the composites is also of special interest.

### 4.1 Data-generating process

The data were generated according to a structural model (see Fig. 1) consisting of two exogenous composites ($\xi_1$ and $\xi_2$) and one endogenous composite ($\xi_3$).

The inner model was defined as:

$$\xi_3 = \beta_1\xi_1 + \beta_2\xi_2 + \zeta,$$
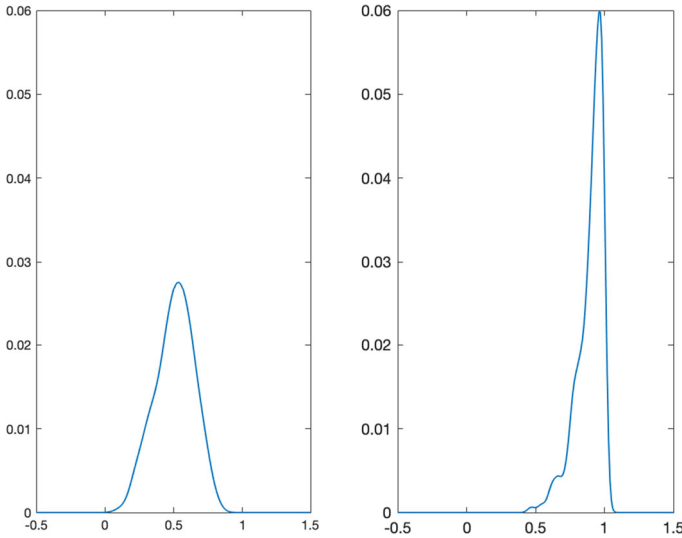
**Fig. 2** Examples of distributions of exogenous composite scores generated in the simulation study: symmetric (left plot) and highly skewed (right plot)

**Table 1** Simulation design

| | | $\beta(6, 6)$ | | $\beta(9, 1)$ | |
|---|---|---|---|---|---|
| $\varepsilon$ (%) | $\zeta$ (%) | $n = 50$ | $n = 500$ | $n = 50$ | $n = 500$ |
| 10 | 10 | A | B | I | L |
| 30 | 10 | C | D | M | N |
| 10 | 30 | E | F | O | P |
| 30 | 30 | G | H | Q | R |

where $\beta_1$ and $\beta_2$ are the path coefficients, and $\zeta$ is the random disturbance effect. The following values were assumed: $\beta_1 = 0.9$; $\beta_2 = 0.3$. The measurement model equations for the generic latent variables $\xi_h$, with $h = 1, \ldots, 3$, were:

$$x_{1h} = \lambda_{1h}\xi_h + \epsilon_{1h},$$
$$x_{2h} = \lambda_{2h}\xi_h + \epsilon_{2h},$$
$$x_{3h} = \lambda_{3h}\xi_h + \epsilon_{3h},$$

where $\lambda_{1h}$, $\lambda_{2h}$, and $\lambda_{3h}$ are the loadings, and $\epsilon_{1h}$, $\epsilon_{2h}$, and $\epsilon_{3h}$ are the random noise effects. The following values were assumed: $\lambda_{ph} = 0.75; 0.80; 0.85$ for $p = 1, \ldots, 3$. The exogenous composites $\xi_h$ were generated from the beta distribution $\beta_{u,v}$: B(6,6) symmetric case, and B(9,1) highly skewed. The noises $\zeta$ and $\epsilon$ were realizations of the continuous uniform distribution $U(-a, a)$, with an expectation of zero. Variance accounted for two levels of the corresponding dependent variable variance: 10% (low noise) and 30% (medium noise). The simulation of the disturbance components is a ticklish choice in such a scheme. As a matter of fact, it is not feasible to simultaneously

**Table 2** Comparison of simulation settings A, B, C, D, from 500 replicates

| | A | | B | | C | | D | |
|---|---|---|---|---|---|---|---|---|
| | PLSPM | PPRPM | PLSPM | PPRPM | PLSPM | PPRPM | PLSPM | PPRPM |
| Bias | | | | | | | | |
| $\beta_1$ | − 0.038 | − 0.026 | − 0.044 | − 0.035 | − 0.116 | − 0.104 | − 0.126 | − 0.114 |
| $\beta_2$ | − 0.012 | − 0.019 | − 0.014 | − 0.017 | − 0.036 | − 0.042 | − 0.041 | − 0.039 |
| Average | − 0.025 | − 0.023 | − 0.029 | − 0.026 | − 0.076 | − 0.073 | − 0.083 | − 0.076 |
| Mean square error | | | | | | | | |
| $\beta_1$ | 0.003 | 0.013 | 0.002 | 0.007 | 0.016 | 0.035 | 0.016 | 0.028 |
| $\beta_2$ | 0.005 | 0.021 | 0.001 | 0.007 | 0.009 | 0.043 | 0.002 | 0.015 |
| Average | 0.004 | 0.017 | 0.001 | 0.007 | 0.013 | 0.039 | 0.009 | 0.022 |
| Skewness | | | | | | | | |
| $\xi_1$ | 0.004 | 0.004 | 0.003 | 0.003 | 0.022 | 0.022 | − 0.002 | − 0.002 |
| $\xi_2$ | − 0.001 | − 0.007 | − 0.005 | − 0.006 | 0.002 | 0.004 | − 0.004 | − 0.005 |
| $\xi_3$ | 0.011 | 0.011 | 0.002 | 0.002 | 0.020 | 0.021 | − 0.001 | − 0.001 |

**Table 3** Comparison of simulation settings E, F, G, H, from 500 replicates

| | E | | F | | G | | H | |
|---|---|---|---|---|---|---|---|---|
| | PLSPM | PPRPM | PLSPM | PPRPM | PLSPM | PPRPM | PLSPM | PPRPM |
| Bias | | | | | | | | |
| $\beta_1$ | − 0.142 | − 0.144 | − 0.146 | − 0.142 | − 0.211 | − 0.214 | − 0.218 | − 0.214 |
| $\beta_2$ | − 0.034 | − 0.050 | − 0.051 | − 0.052 | − 0.063 | − 0.083 | − 0.067 | − 0.060 |
| Average | − 0.088 | − 0.097 | − 0.099 | − 0.097 | − 0.137 | − 0.148 | − 0.142 | − 0.137 |
| Mean square error | | | | | | | | |
| $\beta_1$ | 0.023 | 0.037 | 0.022 | 0.027 | 0.049 | 0.078 | 0.048 | 0.061 |
| $\beta_2$ | 0.009 | 0.028 | 0.003 | 0.010 | 0.015 | 0.049 | 0.005 | 0.018 |
| Average | 0.016 | 0.032 | 0.013 | 0.018 | 0.032 | 0.064 | 0.027 | 0.040 |
| Skewness | | | | | | | | |
| $\xi_1$ | 0.010 | 0.009 | 0.004 | 0.004 | − 0.026 | − 0.025 | − 0.002 | − 0.002 |
| $\xi_2$ | 0.006 | 0.001 | − 0.001 | − 0.002 | − 0.009 | − 0.008 | 0.002 | 0.002 |
| $\xi_3$ | − 0.006 | − 0.007 | 0.007 | 0.007 | 0.000 | 0.000 | − 0.003 | − 0.003 |

**Table 4** Comparison of simulation settings I, L, M, N, from 500 replicates

| | I | | L | | M | | N | |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| | PLSPM | PPRPM | PLSPM | PPRPM | PLSPM | PPRPM | PLSPM | PPRPM |
| Bias | | | | | | | | |
| $\beta_1$ | − 0.040 | − 0.048 | − 0.046 | − 0.053 | − 0.122 | − 0.151 | − 0.128 | − 0.126 |
| $\beta_2$ | − 0.006 | − 0.016 | − 0.014 | − 0.013 | − 0.033 | − 0.054 | − 0.039 | − 0.042 |
| Average | − 0.023 | − 0.032 | − 0.030 | − 0.033 | − 0.078 | − 0.103 | − 0.084 | − 0.084 |
| Mean square error | | | | | | | | |
| $\beta_1$ | 0.004 | 0.016 | 0.002 | 0.010 | 0.018 | 0.051 | 0.017 | 0.029 |
| $\beta_2$ | 0.006 | 0.023 | 0.001 | 0.009 | 0.010 | 0.042 | 0.002 | 0.018 |
| Average | 0.005 | 0.019 | 0.002 | 0.010 | 0.014 | 0.047 | 0.010 | 0.023 |
| Skewness | | | | | | | | |
| $\xi_1$ | − 1.205 | − 1.201 | − 1.322 | − 1.321 | − 1.026 | − 1.017 | − 1.142 | − 1.142 |
| $\xi_2$ | − 1.183 | − 1.132 | − 1.314 | − 1.294 | − 0.977 | − 0.845 | − 1.144 | − 1.115 |
| $\xi_3$ | − 0.890 | − 0.888 | − 1.009 | − 1.009 | − 0.759 | − 0.748 | − 0.871 | − 0.870 |

**Table 5** Comparison of simulation settings O, P, Q, R, from 500 replicates

| | O | | P | | Q | | R | |
|---|---|---|---|---|---|---|---|---|
| | PLSPM | PPRPM | PLSPM | PPRPM | PLSPM | PPRPM | PLSPM | PPRPM |
| Bias | | | | | | | | |
| $\beta_1$ | − 0.144 | − 0.159 | − 0.149 | − 0.149 | − 0.209 | − 0.237 | − 0.220 | − 0.224 |
| $\beta_2$ | − 0.041 | − 0.055 | − 0.048 | − 0.039 | − 0.058 | − 0.086 | − 0.072 | − 0.079 |
| Average | − 0.092 | − 0.107 | − 0.098 | − 0.094 | − 0.133 | − 0.161 | − 0.146 | − 0.152 |
| Mean square error | | | | | | | | |
| $\beta_1$ | 0.024 | 0.045 | 0.022 | 0.029 | 0.048 | 0.089 | 0.049 | 0.065 |
| $\beta_2$ | 0.011 | 0.030 | 0.003 | 0.010 | 0.014 | 0.054 | 0.006 | 0.021 |
| Average | 0.018 | 0.037 | 0.013 | 0.019 | 0.031 | 0.071 | 0.027 | 0.043 |
| Skewness | | | | | | | | |
| $\xi_1$ | − 1.213 | − 1.208 | − 1.326 | − 1.326 | − 1.043 | − 1.024 | − 1.147 | − 1.145 |
| $\xi_2$ | − 1.183 | − 1.127 | − 1.319 | − 1.299 | − 0.927 | − 0.805 | − 1.143 | − 1.114 |
| $\xi_3$ | − 0.632 | − 0.630 | − 0.694 | − 0.694 | − 0.533 | − 0.525 | − 0.598 | − 0.597 |
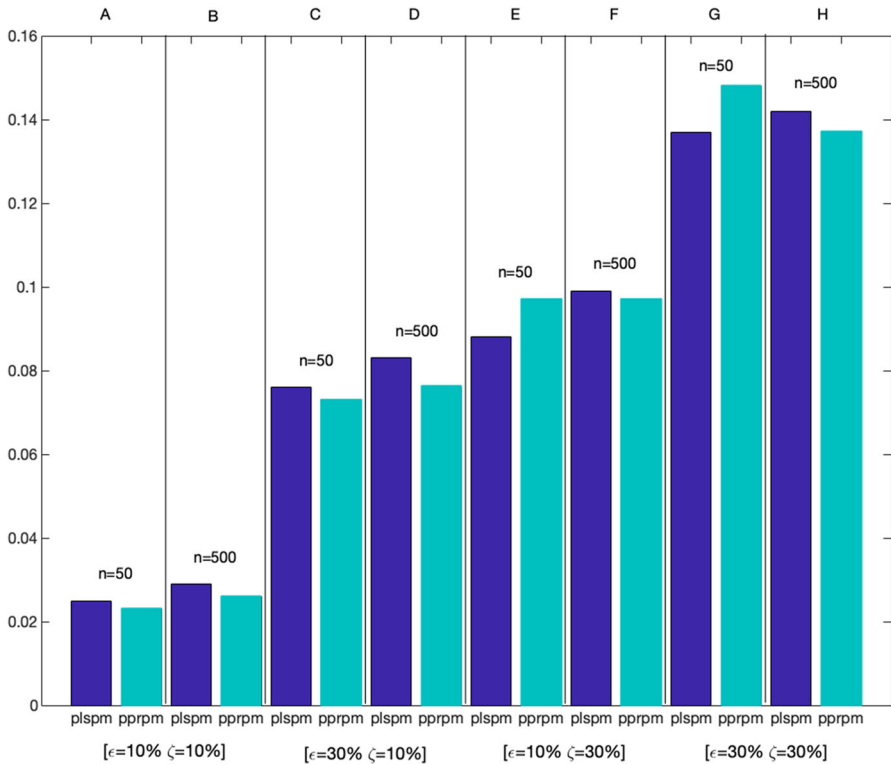
**Fig. 3** Bias of scenarios A, B, C, D, E, F, G, and H with symmetric MVs ($\beta(6, 6)$)

take under control the endogenous composites and the MVs distributions, because they are defined as the sum of the variables. To appreciate the skewness effect on the model parameter estimations (and on the latent variables), according to Cassel et al. (1999, 2000) and Westlund et al. (2001), the two sources of uncertainty were generated from the uniform distribution. According to our formulation, the vagueness in the model inadequacy is fully represented by the whole set of relationships among the composites. To appreciate the vagueness, the simulation study evaluates the effects of the different scenarios defined as the combination of the latent variables and the error term simulations.

By way of example, Fig. 2 shows the corresponding kernel density plot of the distributions of the exogenous composite scores in the symmetric (left plot) and skewed context (right plot). For the two cases, samples of 50, and of 500 were generated, and the data was then re-scaled to the interval [1, 9]. The sampling distributions consist of 500 replicates of the model estimations.

Table 1 shows the selected simulation settings. As can be seen, the complete design includes 16 scenarios, derived from the intersection of the four factors ($\epsilon$, $\zeta$, $n$, $\beta_{u,v}$), each with two levels. Scenarios from A to H allow us to compare the two methods with increasing sample size and as the level of noise in the inner and outer mod-
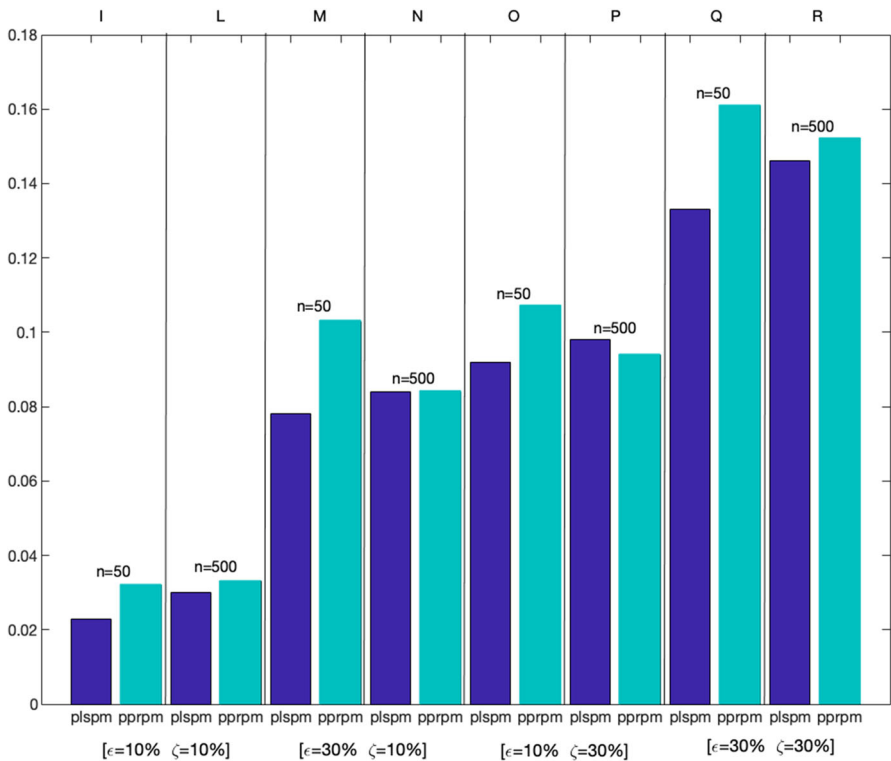
**Fig. 4** Bias of scenarios I, L, M, N, O, P, Q, and R with skewed MVs ($\beta(9, 1)$)

els increases. The remaining scenarios compare the two methods according to the increasing skewness in the MVs.

## 4.2 Results

The results in Table 2 offer a comparison between the two methods with increasing sample size ($n_A = 50$, $n_B = 500$). The estimations are negatively biased for both methods. According to the literature (Schneeweiss 1993), this result was expected for PLSPM; the values shown in Table 2 confirm these results for PLSPM and show that the same occurs for PPRPM. The bias does not seem to be influenced by the increasing sample size. As expected, the PPRPM parameter MSE is larger than the PLSPM MSE in all the scenarios. In PPRPM there is no error term in the inner model; thus, the error component reflects on the parameter estimations. This error component represents the extent of vagueness that in PLSPM is discarded in the model residuals. Moving, respectively, from scenarios A to C and B to D, the bias increases according to the increase in the measurement error term (from 10 to 30%).

The results in Table 3 allow comparing the bias and the precision when the level of noise in the inner and outer models increases. More specifically, from scenarios E to G ($n = 50$), and from F to H ($n = 500$) only the measurement model error varies
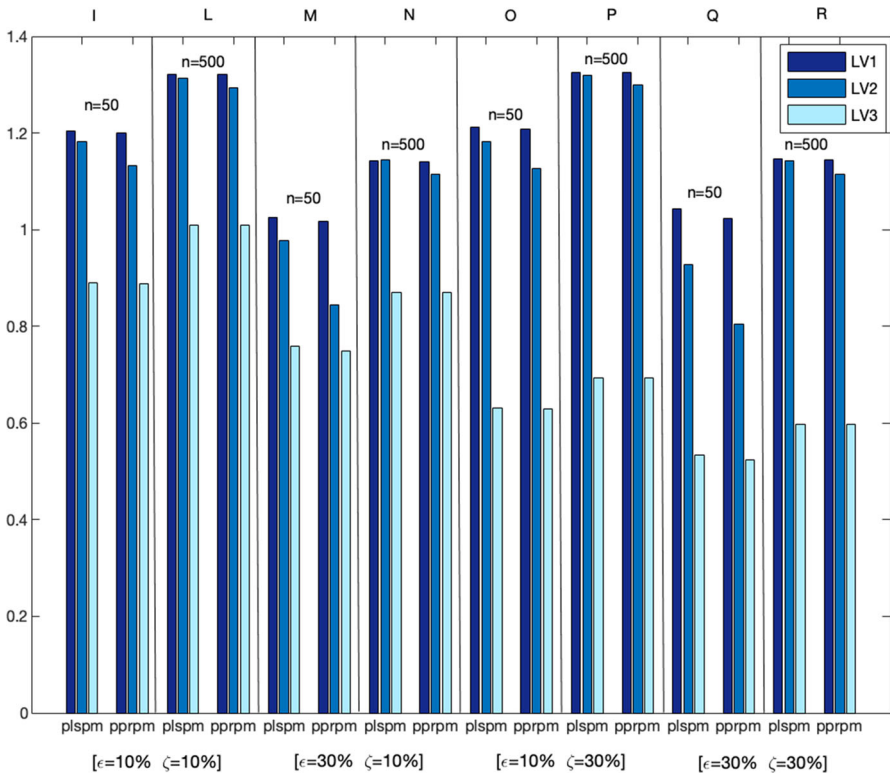
**Fig. 5** Absolute values of skewness of MVs in scenarios I, L, M, N, O, P, Q, and R

(10–30%). The Table 3 shows that the noise level increase in the outer models has a proportional and direct effect on the estimation of the structural model.

Tables 4 and 5 aim to show the effect of the skewness on the model parameter estimations. These tables replicate the same structure of the two previous tables, but the exogenous composites are generated from the skewed Beta distributions. In this case, what is important to note is that PLSPM is slightly affected by the skewness. This effect is well-known among the PLSPM community and is generally considered an advantage in terms of robustness (Cassel et al. 2000; Vilares et al. 2010). However, skewed distributions mean that in one or more composites there is a tendency toward the higher or lower scores in the scale. This is important evidence for the researcher and should be taken properly into account. In PLSPM, such skewed behavior affects the model residuals. However, it is taken into account by the inner model parameters in PPRPM. Tables 4 and 5 highlight that the PPRPM estimations have higher bias, generally. This is due to the larger spreads of the interval-valued parameters toward the direction of the asymmetry. Figures 3 and 4 summarize the results of the simulation study. The average distortion (absolute values) in the scenarios are compared. Figure 3 shows that in all scenarios except E and G the PPRPM has a smaller distortion than that of PLSPM. That is, with increasing structural error (from 10 to 30%), PPRPM has greater distortion, but it becomes smaller when the sample size increases (from

**Table 6** Indicators used to measure the constructs

| Label | Indicators |
| --- | --- |
| SA1 | It is important to share academic content in open platforms |
| SA2 | It is important to publish research results in other media than academic journals or books |
| SA3 | It is important that students become familiar with online collaborative environments |
| PU1 | The use of Wikipedia makes it easier for students to develop new skills |
| PU2 | The use of Wikipedia improves students' learning |
| PU3 | Wikipedia is useful for teaching |
| IM1 | The use of Wikipedia is well considered among colleagues |
| IM2 | In academia, sharing open educational resources is appreciated |
| IM3 | My colleagues use Wikipedia |
| BI1 | In the future I will recommend the use of Wikipedia to my colleagues and students |
| BI2 | In the future I will use Wikipedia in my teaching activity |
| UB1 | I use Wikipedia to develop my teaching materials |
| UB2 | I use Wikipedia as a platform to develop educational activities with students |
| UB3 | I recommend my students to use Wikipedia |
| UB4 | I recommend my colleagues to use Wikipedia |
| UB5 | I agree my students use Wikipedia in my courses |

$n = 50$ to $n = 500$). Figure 4 shows that in the presence of skewed variables, PPRPM always has higher distortion than PLSPM, but the difference between the two methods is decreased as the sample size increases. Skewness has been measured by the standardized third moment coefficient. Consistently with the simulation settings (scenarios from I to R), all composite's distributions are negatively skewed (Tables 4, 5); for sake of legibility Fig. 5 visualizes the absolute values of the skewness. Results show the common tendency of the two methods to present lower asymmetry for the endogenous latent variable, that is, the variable obtained as a linear combination of the two exogenous variables that are simulated as skewed variables. In addition, in all scenarios, the PPRPM produces slightly lower values, especially for the second composite. This result confirms the ability of the LAD regression implemented in the PPRPM measurement model to be more robust than the least squares regression implemented in the PLSPM measurement model.

## 5 Empirical evidence: The use of Wikipedia in higher education

In this section, the results of a practical application are discussed. The proposed data set is available in the UCI Machine Learning Repository (Lichman 2013). The aim of the study was to investigate university faculty members' perceptions and practices in using Wikipedia as a teaching resource (Meseguer-Artola et al. 2016). The survey was conducted among students at two Spanish universities, but only data gathered from the Universitat Oberta de Catalunya (UOC) was used in this application.
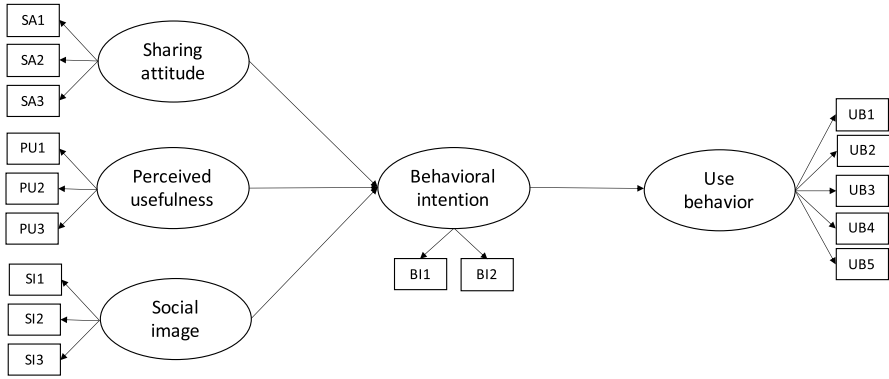
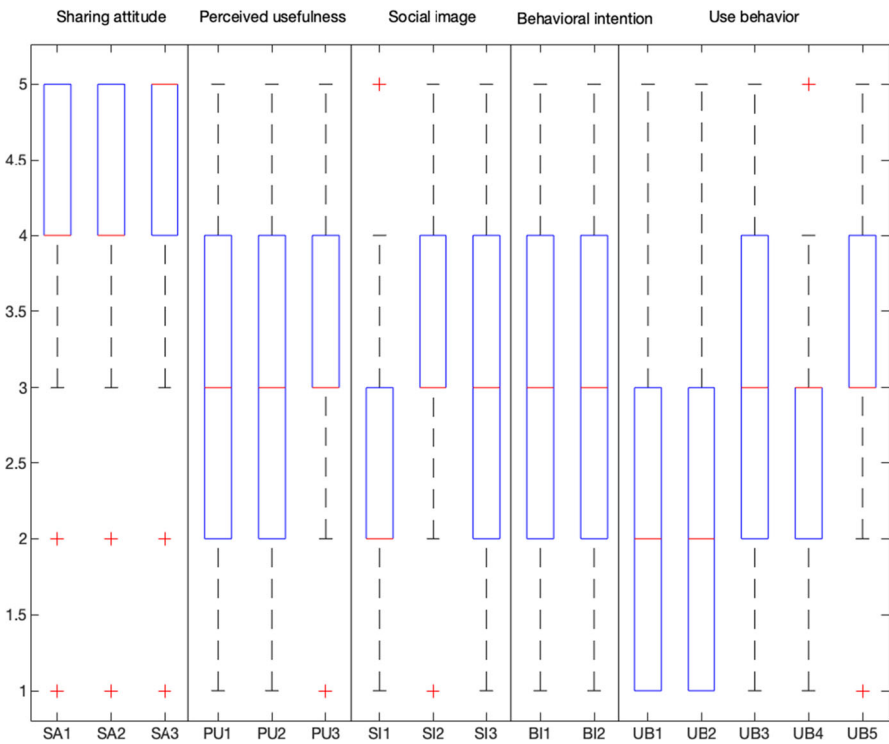**Fig. 6** Model graph and hypotheses



**Fig. 7** Boxplots of the indicators

Five constructs were considered for the current research: *Sharing attitude* (SA), *perceived usefulness* (PU), *social image* (SI), *behavioral intention* (BI), and *use behavior* (UB). Additional details on the specific indicators used to measure these constructs and shown in Table 6 can be found in Meseguer-Artola et al. (2016).

The hypotheses considered in this paper are shown in Fig. 6. The model assumes that teachers' *behavioral intention* to use Wikipedia is directly influenced by their
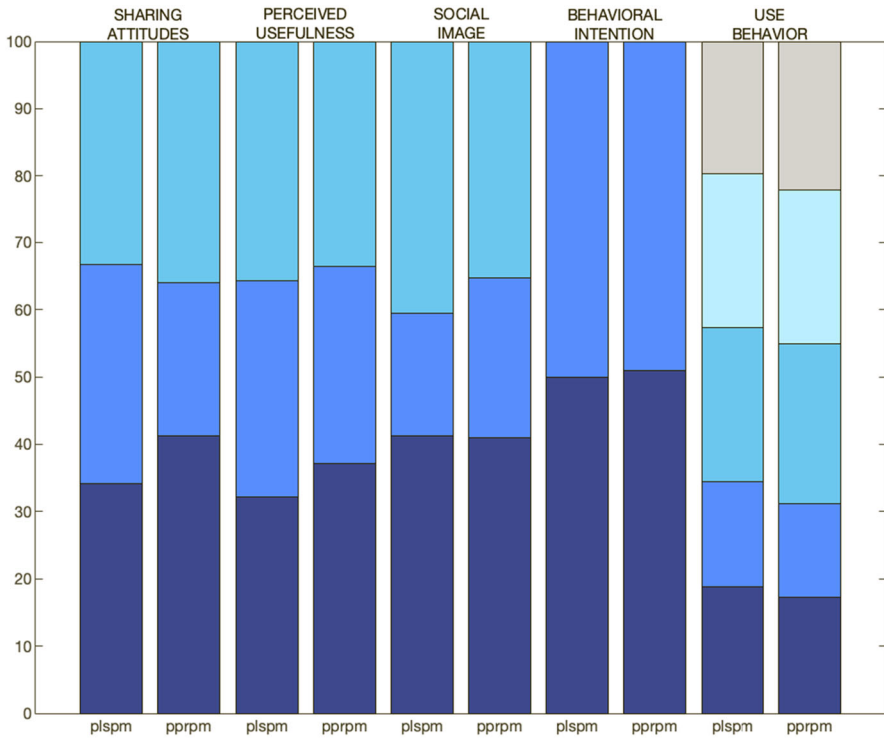
**Fig. 8** Measurement model results: weights

*sharing attitude*, their perceptions of the *social image* of Wikipedia, and the *perceived usefulness* of Wikipedia. Furthermore, it is assumed that the teachers' *use behavior* is determined by their *behavioral intention*.

An exploratory analysis (see Fig. 7) of the indicators shows that some of the distributions are highly skewed, which is common for questionnaires operated on an ordinal scale. Thus, the decision to adopt LAD regression for the measurement model seems appropriate for this type of data. Figures 8 and 9 show the measurement model results, that is, the outer weights and loadings of PLSPM and PPRPM. As shown in Fig. 8, the main differences are found in the *sharing attitude* construct, which presented skewed indicators.

The results for the structural model are reported in Table 7. The path coefficients and the fit indices for PLSPM and PPRPM are shown. The results highlight the role of PU as the most important predictor of BI. In PLSPM, BI (0.52) is higher than SA and SI, in PPRPM, BI has a coefficient with the highest midpoint (0.20) compared to the other constructs. PPRPM also provides component-wise information on the uncertainty of the relationships. This applies to the relationship between BI and SA, whose path coefficient has a range equal to 0.21, and the relationship between BI and SI, whose path coefficient has a range equal to 0.18. Although both approaches show the highest coefficient for the relationship between UB and BI, PPRPM also highlights that this relationship is characterized by greater uncertainty, with a range equal to 0.35.
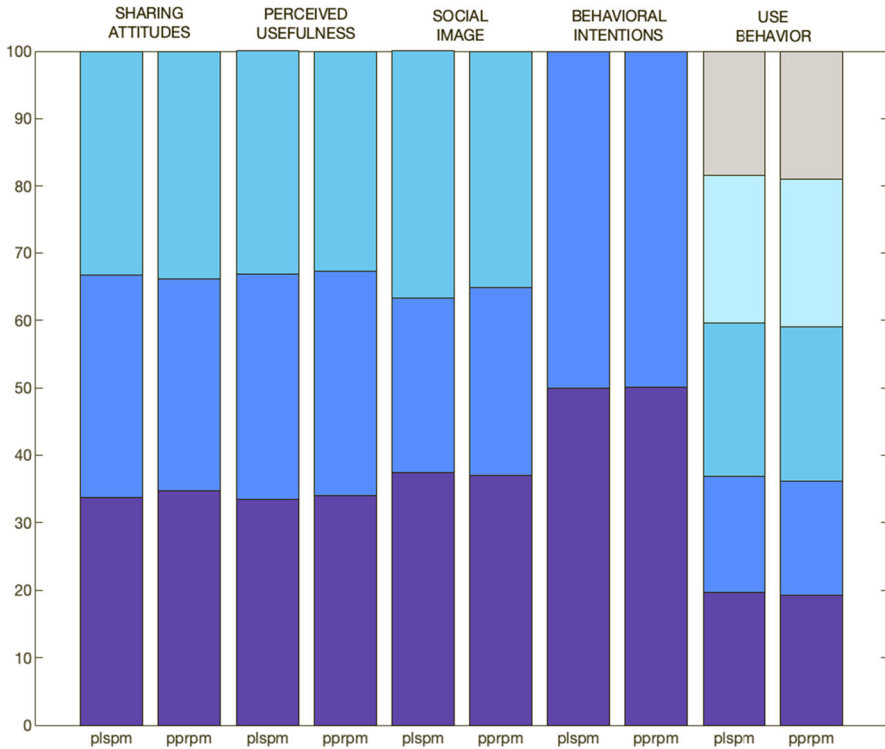
**Fig. 9** Measurement model results: loadings

**Table 7** Structural model results

| Relations | PLSPM | | | PPRPM | |
| | Path coefficient | $R^2$ | | Path coefficient | IC |
|---|---|---|---|---|---|
| BI < SA | 0.14 | 0.50 | | {0.17; 0.21} | 0.85 |
| BI < PU | 0.52 | | | {0.20; 0.00} | |
| BI < SI | 0.22 | | | {0.05; 0.18} | |
| UB < BI | 0.79 | 0.63 | | {0.44; 0.35} | 0.86 |

## 6 Conclusion and perspectives

The partial least squares approach to structural equation model estimation has experienced significant growth over the last decade, thanks to a large number of applications in areas where the theory had been less developed. Researchers consider using PLSPM when the primary objective is to predict and explain target constructs (Hair Jr et al. 2016, p. 14). As demonstrated in previous studies and the current simulation study, PLSPM can integrate data under very limited assumptions and works efficiently with small sample sizes. However, despite its extreme flexibility, PLSPM does not allow the user to properly appreciate global goodness of fit. This limitation also depends on the presence of two different kinds of residuals: the measurement and the struc-

tural model ones. By introducing the possibilistic regression approach to structural model relationships, the current proposed method offers an alternative approach for considering different kinds of residuals.

Structural models can be defined by a more or less complex net of relations, and simulation studies provide empirical proof of the stability and consistency of PPRPM estimates, compared with traditional PLSPM. Such evidence encourages the use of PPRPM as interval-valued parameters method that represent a richer information. More empirical evidence and additional simulation studies are necessary to assess the capabilities of PPRPM. Future research should focus on procedures for protecting against unwanted effects caused by outliers.

# References

Alwin DF, Hauser RM (1975) The decomposition of effects in path analysis. Am Sociol Rev 40:37–47

Baudrit C, Couso I, Dubois D (2007) Joint propagation of probability and possibility in risk analysis: towards a formal framework. Int J Approx Reason 45(1):82–105

Bezdek JC (1981) Models for pattern recognition. Pattern recognition with fuzzy objective function algorithms. Springer, Boston, MA, pp 1–13

Bloomfield P, Steiger WL (1983) LAD in linear regression. In: Least absolute deviations. Progress in probability and statistics, vol 6. Springer, Birkhäuser Boston, pp. 37–76

Bollen KA (1989) Structural equations with latent variables. Wiley, New York

Bollen KA, Bauldry S (2011) Three Cs in measurement models: causal indicators, composite indicators, and covariates. Psychol Methods 16(3):265–284

Cassel C, Hackl P, Westlund AH (1999) Robustness of partial least-squares method for estimating latent variable quality structures. J Appl Stat 26(4):435–446

Cassel CM, Hackl P, Westlund AH (2000) On measurement of intangible assets: a study of robustness of partial least squares. Total Qual Manag 11(7):897–907

Chin WW et al (1998) The partial least squares approach to structural equation modeling. Mod Methods Bus Res 295(2):295–336

Coppi R (2008) Management of uncertainty in statistical reasoning: the case of regression analysis. Int J Approx Reason 47(3):284–305

Couso I, Dubois D (2014) Statistical reasoning with set-valued information: ontic vs. epistemic views. Int J Approx Reason 55(7):1502–1518

Diamond P (1990) Least squares fitting of compact set-valued data. J Math Anal Appl 147(2):351–362

Dijkstra TK (2017) A perfect match between a model and a mode. In: Latan H, Noonan R (eds) Partial least squares path modeling. Springer, Cham, pp 55–80

Dijkstra TK, Henseler J (2015a) Consistent and asymptotically normal PLS estimators for linear structural equations. Comput Stat Data Anal 81:10–23

Dijkstra TK, Henseler J (2015b) Consistent partial least squares path modeling. MIS Q 39(2):297–316

Dodge Y (1997) Lad regression for detecting outliers in response and explanatory variables. J Multivar Anal 61(1):144–158

Dolce P, Vinzi VE, Lauro NC (2018) Non-symmetrical composite-based path modeling. Adv Data Anal Classif 12(3):759–784

Dubois D (2014) On various ways of tackling incomplete information in statistics. Int J Approx Reason 55(7):1570–1574

Ferraro MB, Coppi R, Rodríguez GG, Colubi A (2010) A linear regression model for imprecise response. Int J Approx Reason 51(7):759–770

Ferraro MB, Giordani P (2017) Possibilistic and fuzzy clustering methods for robust analysis of non-precise data. Int J Approx Reason 88:23–38

Hair JF Jr, Hult GTM, Ringle C, Sarstedt M (2016) A primer on partial least squares structural equation modeling (PLS-SEM). Sage, Thousand Oaks

Hanafi M (2007) PLS path modelling: computation of latent variables with the estimation mode b. Comput Stat 22(2):275–292

Judd CM, McClelland GH, Ryan CS (2011) Data analysis: a model comparison approach. Routledge, New York

Kaplan D (2008) Structural equation modeling: foundations and extensions, vol 10. Sage, Newbury Park

Kim KJ, Moskowitz H, Koksalan M (1996) Fuzzy versus statistical linear regression. Eur J Oper Res 92(2):417–434

Koenker R (2009) The median is the message: Wilson and Hilfertys experiments on the law of errors. Am Stat 63(1):20–25

Lichman M (2013) UCI machine learning repository. School of Information and Computer Science, University of California, Irvine, CA

Lohmöller J (1989) Latent variable path modeling with partial least squares. Physica-Verlag, Heildelberg

Lyttkens E, Areskoug B, Wold, H (1975) The convergence of NIPALS estimation procedures for six path models with one or two latent variables. Technical report, University of Göteborg

Marino M, Palumbo F (2002) Interval arithmetic for the evaluation of imprecise data effects in least squares linear regression. Ital J Appl Stat 14(3):277–291

Meseguer-Artola A, Aibar E, Lladós J, Minguillón J, Lerga M (2016) Factors that influence the teaching use of Wikipedia in higher education. J Assoc for Inf Sci Technol 67(5):1224–1232

Nadimi R et al (2013) Outlier detection in fuzzy linear regression with crisp input-output by linguistic variable view. Appl Soft Comput 13(1):734–742

Petit-Renaud S, Denœux T (2004) Nonparametric regression analysis of uncertain and imprecise data using belief functions. Int J Approx Reason 35(1):1–28

Romano R, Palumbo F (2016) Partial possibilistic regression path modeling. In: Abdi H, Esposito Vinzi V, Russolillo G, Saporta G, Trinchera L (eds) The multiple facets of partial least squares and related methods, vol 173. Springer, Cham, pp 155–168

Romano R, Palumbo F (2017) Handling uncertainty in structural equation modeling. In: Ferraro MB, Giordani P, Vantaggi B, Gagolewski M, Ángeles Gil M, Grzegorzewski P, Hryniewicz O (eds) Soft methods for data science. Springer, Cham, pp 431–438

Rönkkö M (2014) The effects of chance correlations on partial least squares path modeling. Organ Res Methods 17(2):164–181

Rönkkö M, Evermann J (2013) A critical examination of common beliefs about partial least squares path modeling. Organ Res Methods 16(3):425–448

Schneeweiss H (1993) Consistency at large in models with latent variables. In: Haagen K, Bartholomew DJ, Deistler M (eds.) Statistical modelling and latent variables. Elsevier, Amsterdam

Tanaka H, Guo P (1999) Possibilistic data analysis for operations research. Physica Verlag, Heidelberg, Germany

Tenenhaus A, Tenenhaus M (2011) Regularized generalized canonical correlation analysis. Psychometrika 76(2):257–284

Tenenhaus M, Vinzi VE, Chatelin Y-M, Lauro C (2005) PLS path modeling. Comput Stat Data Anal 48(1):159–205

Thurstone L (1931) The theory of multiple factors. Edwards Brothers, Ann Arbor

Tukey JW (1954) Causation, regression, and path analysis. In: Kempthorne O, Bancroft TA, Gowen JW, Lush JH (eds) Statistics and mathematics in biology. Iowa State College Press, Ames, pp 35–66

Vilares MJ, Almeida MH, Coelho PS (2010) Comparison of likelihood and pls estimators for structural equation modeling: a simulation with customer satisfaction data. In: Esposito Vinzi V, Chin WW, Henseler J, Wang H (eds) Handbook of partial least squares: concepts, methods and applications. Springer, Berlin, Heidelberg, pp 289–305

Vinzi VE, Trinchera L, Amato S (2010) PLS path modeling: from foundations to recent developments and open issues for model assessment and improvement. In: Esposito Vinzi V, Chin WW, Henseler J, Wang H (eds) Handbook of partial least squares. Springer, Berlin, Heidelberg, pp 47–82

Wang C, Li J, Guo P (2015) The normalized interval regression model with outlier detection and its real-world application to house pricing problems. Fuzzy Sets Syst 274:109–123

Wang H-F, Tsaur R-C (2000) Insight of a fuzzy regression model. Fuzzy Sets Syst 112(3):355–369

Westlund AH, Cassel CM, Eklöf J, Hackl P (2001) Structural analysis and measurement of customer perceptions, assuming measurement and specifications errors. Total Qual Manag 12(7–8):873–881

Wold H (1975) Soft modelling by latent variables: the non-linear iterative partial least squares (NIPALS) approach. J Appl Probab 12(S1):117–142