



Quantile composite-based path modeling: algorithms, properties and applications

Pasquale Dolce¹ · Cristina Davino² · Domenico Vistocco³

Received: 6 September 2020 / Revised: 5 September 2021 / Accepted: 22 September 2021 /
Published online: 2 November 2021
© The Author(s) 2021

Abstract

Composite-based path modeling aims to study the relationships among a set of constructs, that is a representation of theoretical concepts. Such constructs are operationalized as composites (i.e. linear combinations of observed or manifest variables). The traditional partial least squares approach to composite-based path modeling focuses on the conditional means of the response distributions, being based on ordinary least squares regressions. Several are the cases where limiting to the mean could not reveal interesting effects at other locations of the outcome variables. Among these: when response variables are highly skewed, distributions have heavy tails and the analysis is concerned also about the tail part, heteroscedastic variances of the errors is present, distributions are characterized by outliers and other extreme data. In such cases, the quantile approach to path modeling is a valuable tool to complement the traditional approach, analyzing the entire distribution of outcome variables. Previous research has already shown the benefits of Quantile Composite-based Path Modeling but the methodological properties of the method have never been investigated. This paper offers a complete description of Quantile Composite-based Path Modeling, illustrating in details the method, the algorithms, the partial optimization criteria along with the machinery for validating and assessing the models. The asymptotic properties of the method are investigated through a simulation study. Moreover, an application on chronic kidney disease in diabetic patients is used to provide guidelines

✉ Cristina Davino
cristina.davino@unina.it

Pasquale Dolce
pasquale.dolce@unina.it

Domenico Vistocco
domenico.vistocco@unina.it

¹ Department of Public Health, University of Naples Federico II, Naples, Italy

² Department of Economics and Statistics, University of Naples Federico II, Naples, Italy

³ Department of Political Science, University of Naples Federico II, Naples, Italy

for the interpretation of results and to show the potentialities of the method to detect heterogeneity in the variable relationships.

Keywords Composite-based path modeling · OLS regression · Quantile regression

Mathematics Subject Classification 62-07: (Statistics) Data analysis · 62H99: (Statistics) Multivariate analysis, forse · 62G08: (Statistics) Nonparametric regression · 62P10: (Statistics) Applications to biology and medical sciences

1 Introduction

Several are the approaches to study the relationships among different constructs and between each construct and its corresponding observed or manifest variables (MVs). In most common models, each block of MVs measures a construct, and prior knowledge is used to define the theoretical model. Two are the main parameters in this type of model: the path coefficients and the loadings. Path coefficients represent the relationships between constructs while loadings measure the relationship between constructs and the corresponding MVs.

Covariance structure analysis (Jöreskog 1978) and **Partial Least Square Path Modeling (PLS-PM)** (Esposito Vinzi et al. 2010; Hair et al. 2017) are the two main-stream approaches. Even if they are commonly considered as alternative, they belong to two different families of statistical methods. Covariance structure analysis, essentially used in factor-based **Structural Equation Modeling (SEM)**, exploits the covariance matrix of MVs to estimate the model parameters. PLS-PM instead summarizes each block of MVs in a component, or composite, namely an exact linear combination of the MVs, focusing on the explained variance of MVs (Wold 1982, 1985). Each composite is a proxy of the construct associated to the correspondent block. For the aforesaid reasons, PLS-PM is commonly referred to as a component-based, composite-based or variance-based approach. Herman Wold, who proposed the PLS-PM, referred to this approach as “*Soft Modeling*”. The name indicates that the method requires “*soft*” distributional assumptions, in contrast to the estimation method for factor-based models, which requires strong assumptions on the error distributions (thus the name “*hard modeling*”) (Wold 1975, 1982; Tenenhaus et al. 2005; Chin 1998).

PLS-PM exploits least square regression to estimate the model coefficients and therefore focuses on the conditional mean of the response variables. Many are the cases where the analysis of the average alone could produce an incomplete view of the complex structure of the relationships among variables. When heteroscedastic variances of least square regression residuals occurs, and/or when response variables are highly skewed, the study of the conditional distribution at locations different from the mean can complement the classical approach and provide a richer picture of the investigated phenomenon. Following this idea, **Quantile Compositied-based Path Modeling (QC-PM)**, proposed by Davino and Esposito Vinzi (2016), exploits quantile regression (Koenker and Basset 1978) to look beyond the average. It is a valuable tool to study the relationships among variables, and to model location, scale and shape of the responses. QC-PM can be used to complement PLS-PM to investigate whether the

effects of explanatory constructs change over the entire distribution of the response constructs. It is worth emphasizing that PLS-PM and QC-PM are not competing methods, and therefore their comparison in terms of performance is of little interest. The two methods have different objectives: PLS-PM focuses on the conditional means of the dependent variables providing an instant summary, while QC-PM explores relationships among variables outside of conditional mean. Finding that estimated coefficients vary across the conditional quantiles does not imply that the PLS-PM results are invalid.

Previous research (Davino et al. 2016, 2017, 2018, 2020; Davino and Esposito Vinzi 2016) was oriented to show the relative advantages of QC-PM when the interest is in the effects of explanatory constructs on the entire distribution of the response constructs, and to assist in the interpretation of results and in their use combined with PLS-PM results.

The aim of this paper is to provide a complete and organic description of QC-PM, since its methodological properties have never been investigated. This goal is pursued through several innovative contributions introduced in the paper: a clear and detailed explanation of the method introducing also the case of one block and two blocks, an improvement of the method allowing to handle the measurement invariance issue, an application with artificial data that allows to highlight the potential of the method. More specifically, in order to clarify the characteristics of the method, a step-by-step description of the algorithms, the partial optimization criteria, and the formalization of the models are provided. A relevant part of the present work is devoted to studying the properties of QC-PM, which have never been investigated before.

An analytical discussion of the properties of the involved estimators is a daunting task, due to the complexity of composite-based path modeling. This is a fertile ground for the use of simulation studies, which provide information on the performance of the method in terms of bias, efficiency and robustness of the estimates (Paxton et al. 2001). In particular, we exploit a Monte Carlo simulation design generating data from a composite-based population and considering a set of different scenarios. This allows us to assess the effects of several drivers: correlation within the blocks of MVs, correlation among constructs, effect of heavy-tails and skewness of MV distributions, effect of sample size.

A discussion on the asymptotic properties of the method is offered, along with empirical evidence on the behaviour of the estimators in terms of bias and efficiency. An innovative contribution to the estimation of the outer model is also provided to fulfil the measurement invariance required to compare path coefficients estimates over quantiles.

Moreover, an application on Chronic Kidney Disease (CKD) in diabetic patients is provided to show benefit of using QC-PM as a supplement to PLS-PM. In particular, we apply QC-PM on data already used in a research that proposed a quantile approach to factor based SEM (Wang et al. 2016). Because the original data were not available, they have been artificially derived mimicking the model, the relations among variables and the estimates obtained in the original study (Wang et al. 2016). As the artificial data are generated from a scenario where relationships among variables change with quantiles, the application highlights the potentialities of the method in detecting the

heterogeneity in the variable relationships and stressing its complementary role with the traditional methods for composite-based path modeling.

The paper is organized as follows: Sect. 2 describes QC-PM in detail, formalizing the estimation process starting from the simplest case of one block of MVs and moving until the general path model for multi-block data. Section 3 illustrates the assessment measures of QC-PM in terms of goodness of fit and statistical significance of the estimated coefficients. Section 4 shows the simulation design and the main results, while the applicative potentialities of QC-PM, along with guidelines for the interpretation of results, are provided through the study on the artificial data set concerning CKD in diabetic patients in Sect. 5. Finally, a summary of the proposal and an outline for future developments are given in Sect. 6.

2 QC-PM: quantile composite-based path modeling

QC-PM is strongly related to PLS-PM. Therefore the modeling and estimation procedures have much in common, the same holds for their properties. The theoretical foundations are framed in the iterative algorithm proposed by Wold (1966a, b), the Nonlinear estimation by Iterative Partial Least Squares (NIPALS) algorithm, an alternative algorithm for implementing principal component analysis. More broadly, Partial Least Squares (PLS) refers to a set of iterative alternating Ordinary Least Squares (OLS) algorithms, extending the NIPALS algorithm to implement a large number of multivariate statistical techniques (Esposito Vinzi and Russolillo 2013), depending on the involved MVs. For example, in case of one block of MVs, PLS provides principal component analysis. In case of two blocks of MVs, multivariate regularized regression can be obtained (PLS regression). In case of multi-block data, PLS algorithm produces PLS-PM (Lohmöller 1989).

The numerical solutions of all these methods are obtained through an iterative algorithm, which is the first stage of the procedure. The basic idea of the PLS iterative algorithm, which computes the weights used to define the composites, is to partition the set of model parameters to be estimated in subsets. At each step of the algorithm, one subset of parameters is considered known and held fixed, while the other subset is estimated (Lohmöller 1989). A least squares criterion is adopted to estimate the parameters in each step. The name PLS comes from the use of OLS to face with the least square criterion at each step. For example, in case of multi-block data, namely PLS-PM, the procedure comes down to a set of simple and multiple OLS regressions, and Pearson correlation computation.

The QC-PM algorithm follow exactly the same steps of the PLS-PM algorithm, but replacing simple and multiple OLS regression with their Quantile Regression (QR) counterparts (Koenker 2005). The same for classical Pearson correlation, which is replaced with quantile correlation (Li et al. 2014). As well as PLS-PM, QC-PM is based on a two-stage procedure. The first stage aims at computing the outer weights by an iterative procedure (these weights are then used for computing the composites). In the second stage, model parameters (loadings and path coefficients) are estimated through regression analysis using the composites. As for PLS-PM, partial criteria are optimized at each step.

The use of quantile based tools in all phases of the algorithm shifts the focus to the entire conditional distributions of the involved response variables, allowing to estimate partial conditional quantiles. Through the use of different conditional quantiles, the whole response distribution can be inspected. Therefore QC-PM is a valuable complement to PLS-PM, as much as quantiles are a complement to the average. For the sake of illustration, next subsections present QC-PM for the simplest model (one block of MVs), for two blocks of MVs, and for multi-block data (the general model), respectively.

The presentation of the QC-PM algorithm will follow the same steps and the same approach generally used to present the PLS-PM algorithm (Lohmöller 1989; Tenenhaus et al. 2005; Esposito Vinzi and Russolillo 2013). A basic knowledge of QR is assumed. Appendix provides a basic introduction of QR idea and goals. For a more detailed description of QR, please refer to Koenker and Basset (1978), Davino et al. (2013) and Furno and Vistocco (2018).

2.1 Quantile path modeling for one block of manifest variables

The simplest model involves one block of MVs and a construct. The relationships between them are depicted in Fig. 1. MVs are denoted with $\mathbf{X} = \{x_{ip}\}$ and represented through rectangles. It is worth to recall that $i = 1, \dots, n$ refers to the observations, n denoting their number, and $p = 1, \dots, P$ refers to the MVs, P denoting the number of MVs in each block. The corresponding construct is labeled with $\xi = \{\xi_i\}$ and is placed in oval or circle. This diagram is called path diagram.

The relationships among MVs and construct in the path diagram can be translated into a system of simultaneous equations. In particular, for each considered quantile $\theta \in (0, 1)$, the link between each MV x_p and ξ is defined through the following equation:

$$x_{ip} = \alpha_p(\theta) + \lambda_p(\theta) \xi_i + \epsilon_{ip}(\theta), \tag{1}$$

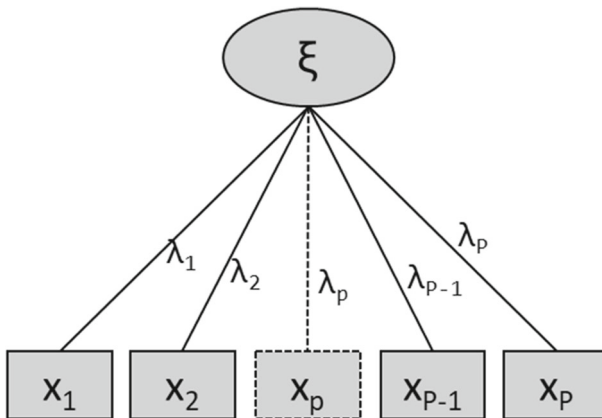


Fig. 1 A path diagram for an hypothetical one-block model

where α_p is a location parameter, λ_p is the loading coefficient, capturing the effect of ξ on \mathbf{x}_p , and $\epsilon_p = \{\epsilon_{ip}\}$ is the error term vector. The only assumption is that the generic θ th conditional quantile of \mathbf{x}_p can be expressed as:

$$Q_\theta(\mathbf{x}_p|\xi) = \alpha_p(\theta) + \lambda_p(\theta)\xi \tag{2}$$

which implies that the θ th quantile of the error term $\epsilon_p(\theta)$ is equal to zero and $\epsilon_p(\theta)$ is independent of ξ . No assumptions on the error distribution are required.

The model captures the common variation among the MVs in \mathbf{X} that depends on the construct ξ . The use of QR allows to encompass the effect of ξ on the whole conditional distribution of each \mathbf{x}_p , effect that might not be equal at different locations, i.e. quantiles. In fact, the construct could be a weak factor at some location of the MV distributions, exerting a stronger effect at other conditional quantiles. In other cases the effect could be almost uniform along the entire distribution. The quantile model offers the opportunity to investigate the possible different situations.

Parameters are estimated through the classical PLS algorithm for one block of variables, replacing OLS regression with QR at each step of the procedure. This corresponds to iteratively optimize a quantile partial criterion. The algorithm consists of three main steps. In the initialization step, arbitrary values are set for the outer weights $\hat{\mathbf{w}}_p$, namely the coefficients used to compute the composite $\hat{\xi}$, which expresses the construct as a linear combination of the MVs. Starting from such initialization values, a first approximation of the composite is computed as a linear combination of the MVs in \mathbf{X} , $\hat{\xi}^{(0)} = \sum_{p=1}^P \hat{w}_p^{(0)} \mathbf{x}_p$. Then, a loop starts and at each s th iteration ($s = 0, 1, 2, \dots$), each MV \mathbf{x}_p is regressed on the composite, minimizing the following quantile loss function:

$$\hat{w}_p^{(s)}(\theta) = \underset{w_p(\theta)}{\operatorname{argmin}} \sum_{i=1}^n \rho_\theta\left(x_{ip} - \alpha_p(\theta) - \xi_i^{(s-1)} w_p(\theta)\right), \tag{3}$$

where $\rho_\theta(\cdot)$ is the check function, which asymmetrically weights positive and negative residuals, namely:

$$\rho_\theta(r) = \begin{cases} \theta r & \text{if } r > 0 \\ (\theta - 1)r & \text{if } r \leq 0. \end{cases} \tag{4}$$

The outer weights are then iteratively computed until convergence through simple QR models, where each MV is the response variable and the composite is the regressor. In each step the estimated weights are normalized and used to update the $\hat{\xi}$, obtained as a linear combination of the response MVs \mathbf{x}_p .

By weighting the MVs considering their quantile covariation with the construct, for each quantile θ the proposed algorithm chooses a linear combination of MVs that is a consistent quantile composite. Once convergence is reached or the maximum number of iterations is achieved, loadings are estimated by means of simple QRs over the corresponding scores. The pseudo-code of the QC-PM iterative algorithm for the case of one block of MVs is provided in Algorithm 1.

Algorithm 1 The QC-PM algorithm for the case of one block of MVs

```

1: for each quantile  $\theta$  do
   STEP 1: Initialization
2:    $s \leftarrow 0$  ▷ iteration counter
3:   Choose arbitrary outer weights  $\hat{w}_p^{(s)}(\theta)$  ( $p = 1, \dots, P$ )
4:   Compute  $\hat{\xi}^{(s)}(\theta) = \sum_{p=1}^P \hat{w}_p^{(s)}(\theta) \mathbf{x}_p$ 
   STEP 2: Iteration
5:   repeat
6:      $s \leftarrow s + 1$  ▷ increment the iteration counter
7:     for all manifest variables  $\mathbf{x}_p$  do
8:       Compute  $\hat{w}_p^{(s)}(\theta)$  solving the quantile regressions:
           
$$x_{ip} = \alpha_p^{(s)}(\theta) + w_p^{(s)}(\theta) \hat{\xi}_i^{(s-1)} + \epsilon_{ip}(\theta) \quad (i = 1, \dots, n; p = 1, \dots, P)$$

9:     end for
10:    Normalize the weights:
           
$$\hat{w}_p^{(s)}(\theta) = \frac{\hat{w}_p^{(s)}(\theta)}{\|\mathbf{X}\hat{\mathbf{w}}^{(s)}(\theta)\|} \quad \text{where } \hat{\mathbf{w}}^{(s)}(\theta) = \{\hat{w}_p^{(s)}(\theta)\}$$

11:    Compute  $\hat{\xi}^{(s)}(\theta)$ :
           
$$\hat{\xi}^{(s)}(\theta) = \sum_{p=1}^P \hat{w}_p^{(s)}(\theta) \mathbf{x}_p$$

12:    until  $\hat{\mathbf{w}}^{(s)}(\theta) \approx \hat{\mathbf{w}}^{(s-1)}(\theta)$ 
   STEP 3: Estimation
13:   for all manifest variables  $\mathbf{x}_p$  do
14:     Estimate  $\hat{\lambda}_p(\theta)$  solving the quantile regressions:
           
$$x_{ip} = \alpha_p(\theta) + \lambda_p(\theta) \hat{\xi}_i^{(s)}(\theta) + \epsilon_{ip}(\theta) \quad (i = 1, \dots, n; p = 1, \dots, P)$$

15:   end for
16: end for

```

2.2 Quantile path modeling for two blocks of manifest variables

Figure 2 depicts the path diagram for an hypothetical two-block model. In such a case, let $\mathbf{X} = \{x_{ip}\}$ and $\mathbf{Y} = \{y_{ij}\}$ denote the two blocks of MVs. In particular, the former consists of the explanatory MVs and the latter of the response MVs. We denote with P the number of explanatory MVs, as in the previous subsection, and with J the number of response MVs, y_j being the generic response MV. Moreover, $\xi = \{\xi_i\}$ is the construct representing the explanatory block, and $\eta = \{\eta_i\}$ the construct representing the dependent block. The general model consists of two sub-models: the inner model and the outer model. The inner model refers to the relationships between the constructs, the outer model between each construct and its block of MVs.

By referring to the outer model, for each quantile $\theta \in (0, 1)$, the MVs \mathbf{x}_p in the explanatory block, and the MVs \mathbf{y}_j in the dependent block, are related to their correspondent constructs through the following system of bilinear equations:

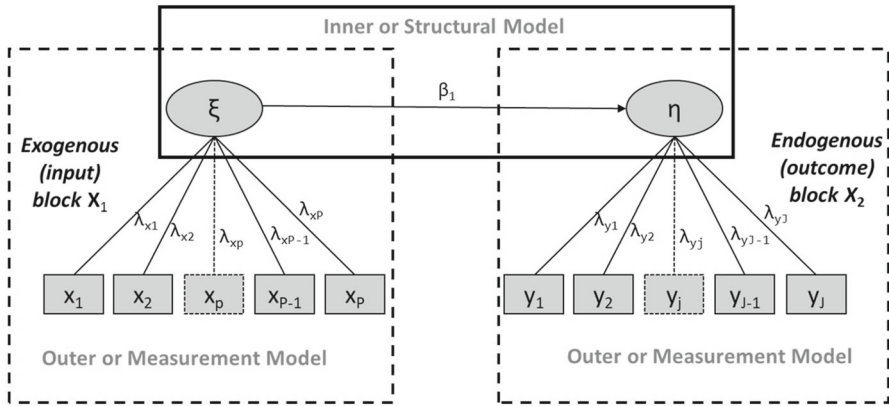


Fig. 2 A path diagram for an hypothetical two-block model

$$x_{ip} = \alpha_{xp}(\theta) + \lambda_{xp}(\theta) \xi_i + \epsilon_{ip}(\theta), \tag{5}$$

$$y_{ij} = \alpha_{yj}(\theta) + \lambda_{yj}(\theta) \eta_i + \omega_{ij}(\theta). \tag{6}$$

where α_{xp} and α_{yj} are location parameters, λ_{xp} is the loading coefficient capturing the effect of ξ on x_p , λ_{yj} is the loading coefficient capturing the effect of η on y_j , while $\epsilon = \{\epsilon_{ip}\}$ and $\omega = \{\omega_{ij}\}$ are the error terms. The usual assumptions on the error terms already mentioned above are required.

The inner model specifies the dependence relationships between the two constructs. The dependent construct η is linked to the explanatory construct ξ by the following model:

$$\eta_i(\theta) = \beta_0(\theta) + \beta_1(\theta) \xi_i + \zeta_i(\theta), \tag{7}$$

where β_1 is the so-called path coefficient capturing the effects of ξ on the dependent construct η , and $\zeta = \{\zeta_i\}$ is the inner error variable.

The procedure for the estimation of the model parameters requires a multi-step algorithm and follows the same structure of the PLS algorithm for two blocks of MVs, defined for example in Esposito Vinzi and Russolillo (2013), where partial criteria are optimized iteratively. The pseudo code of the QC-PM algorithm for the case of two blocks of MVs is detailed in Algorithm 2.

Algorithm 2 The QC–PM algorithm for the case of two blocks of MVs

1: **for** each quantile θ **do**
STEP 1: Initialization
 2: $s \leftarrow 0$ ▷ iteration counter
 3: Choose arbitrary outer weights $\hat{w}_{xp}^{(s)}(\theta)$ ($p = 1, \dots, P$)
 4: Compute $\hat{\xi}^{(s)}(\theta) = \sum_{p=1}^P \hat{w}_{xp}^{(s)}(\theta) \mathbf{x}_p$
STEP 2: Iteration
 5: **repeat**
 6: $s \leftarrow s + 1$ ▷ increment the iteration counter
 7: ▷ dependent block
 8: **for all** manifest variables \mathbf{y}_j of the dependent block **do**
 9: Compute $\hat{w}_j^{(s)}(\theta)$ solving the quantile regressions:

$$y_{ij} = \alpha_j^{(s)}(\theta) + w_j^{(s)}(\theta) \hat{\xi}_i^{(s-1)} + \epsilon_{ij}(\theta) \quad (i = 1, \dots, n; j = 1, \dots, J)$$

 10: **end for**
 11: Normalize the weights:

$$\hat{w}_{yj}^{(s)}(\theta) = \frac{\hat{w}_{yj}^{(s)}(\theta)}{\|\mathbf{Y}\hat{\mathbf{w}}_y^{(s)}(\theta)\|} \quad \text{where } \hat{\mathbf{w}}_y^{(s)}(\theta) = \{\hat{w}_{yj}^{(s)}(\theta)\}$$

 12: Compute $\hat{\boldsymbol{\eta}}^{(s)}(\theta)$:

$$\hat{\boldsymbol{\eta}}^{(s)}(\theta) = \sum_{j=1}^J \hat{w}_{yj}^{(s)}(\theta) \mathbf{y}_j$$

 13: ▷ explanatory block
 14: **for all** manifest variables \mathbf{x}_p of the explanatory block **do**
 15: Compute $\hat{w}_p^{(s)}(\theta)$ solving the quantile regressions:

$$x_{ip} = \alpha_p^{(s)}(\theta) + w_p^{(s)}(\theta) \hat{\eta}_i^{(s)}(\theta) + \epsilon_{ip}(\theta) \quad (i = 1, \dots, n; p = 1, \dots, P)$$

 16: **end for**
 17: Normalize the weights:

$$\hat{w}_{xp}^{(s)}(\theta) = \frac{\hat{w}_{xp}^{(s)}(\theta)}{\|\mathbf{X}\hat{\mathbf{w}}_x^{(s)}(\theta)\|} \quad \text{where } \hat{\mathbf{w}}_x^{(s)}(\theta) = \{\hat{w}_{xp}^{(s)}(\theta)\}$$

 18: Compute $\hat{\xi}^{(s)}(\theta)$:

$$\hat{\xi}^{(s)}(\theta) = \sum_{p=1}^P \hat{w}_{xp}^{(s)}(\theta) \mathbf{x}_p$$

 19: **until** $[\hat{\mathbf{w}}_x^{(s)}(\theta), \hat{\mathbf{w}}_y^{(s)}(\theta)] \approx [\hat{\mathbf{w}}_x^{(s-1)}(\theta), \hat{\mathbf{w}}_y^{(s-1)}(\theta)]$
STEP 3: Estimation
 20: **for all** manifest variables \mathbf{x}_p **do** ▷ explanatory block
 21: Estimate $\hat{\lambda}_{xp}(\theta)$ solving the quantile regressions:

$$x_{ip} = \alpha_p(\theta) + \lambda_p(\theta) \hat{\xi}_i^{(s)}(\theta) + \epsilon_{ip}(\theta) \quad (i = 1, \dots, n; p = 1, \dots, P)$$

 22: **end for**
 23: **for all** manifest variables \mathbf{y}_j **do** ▷ dependent block
 24: Estimate $\hat{\lambda}_{yj}(\theta)$ solving the quantile regressions:

$$y_{ij} = \alpha_j(\theta) + \lambda_j(\theta) \hat{\eta}_i^{(s)}(\theta) + \epsilon_{ij}(\theta) \quad (i = 1, \dots, n; j = 1, \dots, J)$$

 25: **end for**
 26: Estimate the path coefficient $\beta_1(\theta)$ solving the quantile regressions: ▷ path coefficient

$$\eta_i = \beta_0(\theta) + \beta_1(\theta) \hat{\xi}_i(\theta) + \zeta_i(\theta)$$

 27: **end for**

In the initialization step of the algorithm, arbitrary values are set for the outer weights $\hat{\mathbf{w}}_{xp}$ to compute a first approximation of the composite as a linear combination of the MVs in \mathbf{X} , $\hat{\boldsymbol{\xi}}^{(0)} = \sum_{p=1}^P \hat{w}_{xp}^{(0)} \mathbf{x}_p$. Then, the iterative algorithm step proceeds over two phases. At each s th iteration ($s = 0, 1, 2, \dots$), the response MVs \mathbf{y}_j are regressed on the approximation of the composite $\hat{\boldsymbol{\xi}}^{(s-1)}$, minimizing the following quantile loss function:

$$\hat{w}_{yj}^{(s)}(\theta) = \operatorname{argmin}_{w_{yj}(\theta)} \sum_{i=1}^n \rho_{\theta} \left(y_{ij} - \alpha_p(\theta) - \hat{\xi}_i^{(s-1)} w_{yj}(\theta) \right), \tag{8}$$

where $\rho_{\theta}(\cdot)$ is the check function defined as above.

In the second phase, the estimated $\hat{w}_{yj}^{(s)}(\theta)$, for $j = 1, \dots, J$, are used to compute the composite $\hat{\boldsymbol{\eta}}^{(s)}$ through a linear combination of the response MVs \mathbf{y}_j , and then the explanatory MVs \mathbf{x}_{xp} , are regressed on the obtained linear combination, minimizing the following quantile loss function:

$$\hat{w}_{xp}^{(s)}(\theta) = \operatorname{argmin}_{w_{xp}(\theta)} \sum_{i=1}^n \rho_{\theta} \left(x_{ip} - \alpha_p(\theta) - \hat{\eta}_i^{(s)} w_{xp}(\theta) \right). \tag{9}$$

Finally, an updated approximation of the composite $\hat{\boldsymbol{\xi}}^{(s)}$ is obtained as a linear combination of the explanatory MVs \mathbf{x}_p , using the weights $\hat{w}_{xp}^{(s)}(\theta)$.

These two phases are iteratively repeated until convergence of the outer vectors, $\mathbf{w}_x(\theta) = \{w_{xp}(\theta)\}$ and $\mathbf{w}_y(\theta) = \{w_{yj}(\theta)\}$, is achieved. Then loadings and path coefficients are estimated through quantile regression.

QC-PM algorithm returns, for each quantile θ , a linear combination of the explanatory MVs \mathbf{x}_{xp} by weighting the corresponding MVs on the basis of their quantile covariation with the linear combination of the response MVs \mathbf{y}_j .

2.3 Quantile path modeling for multi-block data

Figure 3 depicts a path model for multi-block data using the case of three blocks. The general model for K blocks follows the same logic. Let us assume that P variables are collected in a table \mathbf{X} of data partitioned in K blocks: $\mathbf{X} = [\mathbf{X}_1, \mathbf{X}_2, \dots, \mathbf{X}_K]$. Let $\mathbf{X}_k = \{x_{ipk}\}$ be a generic block of MVs, where $i = 1, \dots, n$, with n denoting the number of observations, $p_k = 1, \dots, P_k$, with P_k being the number of MVs in the k th block. We denote by $\boldsymbol{\xi}_k = \{\xi_{ik}\}$ and $\mathbf{x}_{pk} = \{x_{ipk}\}$ the LV and a generic MV of the k th block, respectively.

A construct that never appears as a dependent variable in the model is called exogenous, while the endogenous constructs play only the role of dependent variables or of both dependent and explanatory variables. In Fig. 3, for example, $\boldsymbol{\xi}_1$ is an exogenous construct and $\boldsymbol{\xi}_2$ and $\boldsymbol{\xi}_3$ are endogenous constructs.

As for the case of two blocks of MVs, the general model consists of the inner model and the outer model. For each quantile $\theta \in (0, 1)$, in the outer model it is assumed

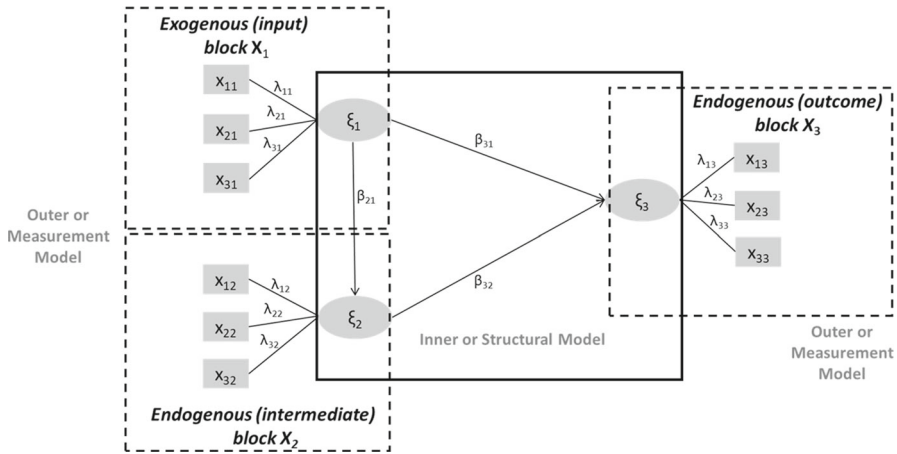


Fig. 3 A path model with three blocks of MVs

that each MV x_{pk} is related to its own construct through the following equations:

$$x_{ipk} = \alpha_{pk}(\theta) + \lambda_{pk}(\theta) \xi_{ik} + \epsilon_{ipk}(\theta), \tag{10}$$

where α_{pk} is the location parameter, $\xi_k = \{\xi_{ik}\}$ is the construct representing the k th block of MVs, λ_{pk} is the loading coefficient, capturing the effect of ξ_k on x_{pk} and $\epsilon_k = \{\epsilon_{ipk}\}$ is the error term vector, using the usual above mentioned assumption on the errors.

The inner model captures and specifies the dependence relationships among constructs. A generic endogenous construct, $\xi_{k'}$, is linked to the related explanatory constructs, ξ_k , $k \in J_{k'}$, where $J_{k'} = \{k : \xi_{k'} \text{ is predicted by } \xi_k\}$, by:

$$\xi_{ik'}(\theta) = \beta_{k'0}(\theta) + \sum_{k \in J_{k'}} \beta_{k'k}(\theta) \xi_{ik} + \zeta_{ik'}(\theta), \tag{11}$$

where $\beta_{k'k}$ is the path coefficient capturing the effects of ξ_k on the dependent construct $\xi_{k'}$, and $\zeta_{k'} = \{\zeta_{ik'}\}$ is the inner error variable vector, with the usual assumption on the errors.

A description of the general QC-PM algorithm is provided in Algorithm 3.

Algorithm 3 The general QC-PM algorithm for the case of multi-block data

- 1: **for** each quantile θ **do**
STEP 1: Initialization
- 2: $s \leftarrow 0$ ▷ iteration counter
- 3: Choose arbitrary outer weights $\hat{w}_{p_k}^{(s)}(\theta)$ ($p_k = 1, \dots, P_k; k = 1, \dots, K$)
- 4: Compute $\hat{\xi}_k^{(s)}(\theta) = \sum_{p_k=1}^{P_k} \hat{w}_{p_k}^{(s)}(\theta) \mathbf{x}_{p_k}$
STEP 2: Iteration
- 5: **repeat**
- 6: $s \leftarrow s + 1$ ▷ increment the iteration counter
- 7: ▷ inner approximation phase ($k \in J_{k'}$)
- 8: Compute quantile correlation:

$$\tau(\theta)_{k'k}^{(s)} = qcor(\theta)(\hat{\xi}_k^{(s-1)}(\theta), \hat{\xi}_{k'}^{(s-1)}(\theta))$$

- 9: Compute inner scores as,

$$\hat{\xi}_{k'}^{(s)}(\theta) = \sum_{k \in J_{k'}} \tau(\theta)_{k'k}^{(s)} \hat{\xi}_k^{(s-1)}(\theta), \quad \text{where } J_{k'} = \{k : \xi_{k'} \text{ is predicted by } \xi_k\}$$

$$\hat{\xi}_k^{(s)}(\theta) = \sum_{k' \in J_k} \tau(\theta)_{k'k}^{(s)} \hat{\xi}_{k'}^{(s-1)}(\theta), \quad \text{where } J_k = \{k' : \xi_k \text{ predicts } \xi_{k'}\}$$
- 10: ▷ outer approximation phase ($k = 1, \dots, K$)
- 11: **for all** manifest variables \mathbf{x}_{p_k} **do**
- 12: Compute $\hat{w}_{p_k}^{(s)}(\theta)$ solving the quantile regressions:

$$x_{ip_k} = \alpha_{p_k}^{(s)}(\theta) + w_{p_k}^{(s)}(\theta) \hat{\xi}_i^{(s)}(\theta) + \epsilon_{ip_k}(\theta) \quad (i = 1, \dots, n; p_k = 1, \dots, P_k)$$

- 13: **end for**
- 14: Normalize the weights:

$$\hat{w}_{p_k}^{(s)}(\theta) = \frac{\hat{w}_{p_k}^{(s)}(\theta)}{\|\mathbf{X}\hat{\mathbf{w}}_k^{(s)}(\theta)\|} \quad \text{where } \hat{\mathbf{w}}_k^{(s)}(\theta) = \{\hat{w}_{p_k}^{(s)}(\theta)\}$$

- 15: Compute $\hat{\xi}_k^{(s)}(\theta) = \sum_{p_k=1}^{P_k} \hat{w}_{p_k}^{(s)}(\theta) \mathbf{x}_{p_k}$
- 16: **until** $\hat{\mathbf{w}}^{(s)} \approx \hat{\mathbf{w}}^{(s-1)}$
STEP 3: Estimation
- 17: **for all** manifest variables \mathbf{x}_{p_k} **do** ▷ estimation of loadings
- 18: Estimate $\hat{\lambda}_{\cdot p}(\theta)$ solving the quantile regressions:

$$x_{ip} = \alpha_p(\theta) + \lambda_p(\theta) \hat{\xi}_i^{(s)}(\theta) + \epsilon_{ip}(\theta) \quad (i = 1, \dots, n; p = 1, \dots, P)$$

- 19: **end for**
- 20: Estimate $\beta_{k'k}(\theta)$, solving the quantile regressions: ▷ estimation of path coefficients

$$\xi_{ik'}(\theta) = \beta_{k'0}(\theta) + \sum_{k' \in J_{k'}} \beta_{k'k}(\theta) \xi_{ik}(\theta) + \zeta_{ik'}(\theta)$$

- 21: **end for**

The weight vector $\mathbf{w}(\theta) = \{w_{p_k}(\theta)\}$, ($p_k = 1, \dots, P_k; k = 1, \dots, K$), used to define the composites, is computed by an iterative algorithm that proceeds over two phases, so-called inner and outer approximation phases, iteratively repeated until convergence of the outer vectors $\mathbf{w}(\theta)$ is achieved (i.e., the change of the outer weights from one iteration to the next is smaller than a predefined tolerance).

In the inner phase, composites are approximated as weighted aggregates of the adjacent composites: two composites are adjacent if there exists a link in the inner model connecting the corresponding constructs, that is, an arrow going from one construct to the other in the path diagram, independently of the direction. The inner weights are defined as the values of the quantile correlation between the composites obtained at the previous step. According to the PLS-PM terminology, this mode to compute inner weights is called *factorial* inner scheme. Another scheme can be also applied, called *centroid* scheme, where the inner weights are computed as the signs of the quantile correlation between the composites (Tenenhaus et al. 2005). These two schemes generally provides very close results, but *factorial* scheme is more advisable when correlation between composites is close to zero. In this case, correlation may oscillate from small negative to small positive values during the iteration cycles, and *factorial* scheme is more advisable because it takes into account the strength of the correlation, instead of just the sign. It is worth to note that unlike Pearson correlation, quantile correlation is not a symmetric measure (Li et al. 2014), hence it is necessary to specify the role played by the involved constructs in each equation (i.e., explanatory or dependent one) for the calculation of the inner weights.

In the outer estimation phase, composites are approximated through a normalized weighted aggregate of the corresponding MVs. Outer weights are computed through simple quantile regressions, where each MV is regressed on the corresponding inner approximation composite. Then, the weights are normalized so that $\text{var}[\mathbf{X}_k \mathbf{w}_k(\theta)] = 1$. According to the PLS-PM terminology (Tenenhaus et al. 2005), this mode to compute outer weights is called *Mode A*. The so-called *Mode B* is also feasible in QC-PM, computing the outer weights as regression coefficients in the quantile multiple regression of the inner approximation composite on its own MVs. Basically, *Mode B* takes account of collinearity among MVs of the some blocks, while *Mode A* ignores this collinearity.

In the QC-PM iterative procedure, at each s th iteration ($s = 0, 1, 2, \dots$), the following partial criterion are then optimized:

$$\hat{\mathbf{w}}_{p_k}(\theta) = \underset{w_{p_k}(\theta)}{\operatorname{argmin}} \sum_{i=1}^n \rho_{\theta} \left(x_{ip_k} - \alpha_p(\theta) - \xi_{ik}^{(s-1)} w_{p_k}(\theta) \right) \tag{12}$$

where $\rho_{\theta}(\cdot)$ is the check function defined as above.

When *Mode B* is applied, the following criterion is instead minimized:

$$\hat{\mathbf{w}}_k^{(s)}(\theta) = \underset{\mathbf{w}_k(\theta)}{\operatorname{argmin}} \sum_{i=1}^n \rho_{\theta} \left(\xi_{ik}^{(s-1)}(\theta) - \alpha_p(\theta) - \mathbf{X}_{ik} \mathbf{w}_k(\theta) \right) \tag{13}$$

When convergence is achieved, loadings and path coefficients are estimated through QR.

As a matter of fact, QC-PM provides, for each quantile of interest, a set of outer weights, loadings and path coefficients, offering a more complete picture of the relationships among variables both in the outer model and in the inner model.

The algorithm provides quantile-based composites, and it is useful to deal with heterogeneity both in the structural model and in the measurement model. In such a case the interest is in evaluating how weights and composites vary across quantile. However, if the interest is in comparing estimated models over quantiles, the measurement invariance (Henseler et al. 2016) has to be fulfilled. If weights, and consequently composites, change over quantiles, a proper comparison among path coefficients estimated at different quantiles is indeed not reliable, because the same concept may not be measured across quantiles. To this end, a test on the weights defined as a variant of the Wald test described in Koenker and Basset (1982) can be exploited. The null hypothesis of the test states that the weights are identical. In case of significant differences among weights, or in case there is the requirement to keep the weights fixed, a new variant of QC-PM can be implemented simply setting the quantile to the median in the iterative procedure. The use of the median in the iterative procedure (step 2) of the algorithm is in line with the approach proposed in Wang et al. (2016) for factor-based SEM. In such an approach the quantile varies only in step 3, to obtain quantile-dependent path coefficients. The median approach can be generalized to the whole iterative process to provide measurement invariance.

3 Model assessment and validation

Once the algorithm converges and estimates for loadings and path coefficients are obtained, there are many tools for assessing both the inner and outer model. Results, namely loadings and path coefficients, can also be validated from an inferential point of view (Davino et al. 2016).

Goodness of fit measures most commonly used in PLS-PM cannot be directly adapted to QC-PM. Moving from OLS to QR requires indeed amendment. The introduction of an effective goodness of fit approach in QR is still an open issue in the scientific literature (Koenker and Machado 1999; He and Zhu 2003). This does make it odd to directly compare OLS and QR, even considering that the two methods optimize different criteria. Therefore, a direct comparison between PLS-PM and QC-PM is not possible.

Starting with the inner model, the coefficient of determination R^2 of the endogenous constructs (Esposito Vinzi et al. 2010) is the criterion mostly used in PLS-PM. Considering that QR loss function is not based on a least squares criterion but rather on a least absolute deviation criterion in terms of weighted residuals, the use of R^2 in QC-PM goes against the underlying rationale of the method. This issue is particularly relevant since most of the assessment indexes in PLS-PM are based on the multiple linear determination coefficient or squared Pearson correlation coefficient. It is against this background we employ the *pseudo- R^2* proposed by Koenker and Machado (1999), so to have a measure that simulates the role and interpretation of the R^2 for QC-PM

assessment. It is important, however, to bear in mind that the *pseudo-R*² is designed differently.

QC-PM estimates a set of parameters for each conditional quantile θ of interest and, consequently, it requires a set of assessment measures for each estimated model. In particular, for each θ , the *pseudo-R*² compares the residual absolute sum of weighted differences using the selected model (*RASW*) with the total absolute sum of weighted differences using a model with the only intercept (*TASW*). *RASW* corresponds to the residual sum of squares in classical regression, *TASW* to the total sum of squares of the dependent variable. *Pseudo-R*² aims to evaluate if the full model (i.e. the model with the regressors) is better in terms of residuals the “restricted” (the model with the only intercept). More precisely, the *pseudo-R*² is calculated as one minus the ratio between *RASW* and *TASW*. In essence, *pseudo-R*² can be considered as a local measure of goodness of fit for a particular quantile as it measures the contribute of the selected regressors to the explanation of the dependent variable with respect to the trivial model without regressors. With an *R*², *pseudo-R*² values range between 0 and 1: the more it is close to 1, the more the model with regressors can be considered a good model (i.e., the θ th conditional quantile function is significantly altered by the effect of the covariates). If on one hand the *pseudo-R*² will always be smaller than the *R*² and a direct comparison with *R*² in PLS-PM is not feasible, on the other hand *pseudo-R*² is useful to the end of identifying locations in the distribution of outcome variable where model may show a better/worse fit (for example, if the model fits in the tail, there’s not guarantee that it fits well anywhere else) (Kováč and Želinský 2013).

For the sake of generality, we consider below the case of multi-block QC-PM. Once convergence is reached and composites are obtained, several QRs are carried out in the inner part of the model, according to the number of considered quantiles. Such QRs estimate the path coefficients linking endogenous and exogenous constructs. As stated in Sect. 2.3, a generic endogenous construct, $\xi_{k'}$, is linked to the related explanatory constructs, ξ_k , $k \in J_{k'}$, where $J_{k'} = \{k : \xi_{k'} \text{ is predicted by } \xi_k\}$. For the convenience of the reader, we report again Eq. (11) that describes this relationship:

$$\xi_{ik'}(\theta) = \beta_{k'0}(\theta) + \sum_{k \in J_{k'}} \beta_{k'k}(\theta) \xi_{ik}(\theta) + \zeta_{ik'}(\theta). \tag{14}$$

Since $\hat{\xi}_{k'}(\theta)$ represents the residuals of the model explaining the k' th endogenous construct, for each considered quantile θ , *RASW* is the corresponding minimizer:

$$RASW_{k'}(\theta) = \sum_{\hat{\xi}_{k'}(\theta) \geq 0} \theta \left| \hat{\xi}_{k'}(\theta) \right| + \sum_{\hat{\xi}_{k'}(\theta) < 0} (1 - \theta) \left| \hat{\xi}_{k'}(\theta) \right|, \tag{15}$$

where positive and negative residuals are asymmetrically weighted, respectively with weights equal to θ and $(1 - \theta)$. The *TASW* is instead:

$$TASW_{k'}(\theta) = \sum_{\xi_{k'} \geq \theta} \theta \left| \xi_{k'} - \theta \right| + \sum_{\xi_{k'} < \theta} (1 - \theta) \left| \xi_{k'} - \theta \right|. \tag{16}$$

Therefore, the obtained *pseudo*– R^2 can be computed as follows:

$$pseudo-R_{k'}^2(\theta) = 1 - \frac{RASW_{k'}(\theta)}{TASW_{k'}(\theta)}. \tag{17}$$

The *pseudo*– R^2 ranges between 0 and 1, since $RASW(\theta)$ is always less than or equal $TASW(\theta)$. It indicates, for each considered quantile, whether the presence of the covariates influences the correspondent conditional quantile of the response variable. It is worth noticing that the *pseudo*– R^2 is not a symmetric measure, assuming a different value when the role of the variables is reversed. The index, computed for each inner equation, measures the amount of variability of a given endogenous construct explained by its explanatory constructs. The average of all the *pseudo*– R^2 indexes provides a synthesis of the evaluations regarding the inner model.

As regards to the outer model, the assessment is carried out considering the relations between each construct and its own MVs and the estimate of the error term vector $\hat{\epsilon}_k$. The *pseudo*– R^2 can be used to assess convergent validity for each outer model, applying for each block the average of the *pseudo*– R^2 indexes of the related MVs, and can be used for assessing the quality of the whole outer models computing a weighted average of all measures over all the blocks, using the number of MVs for each block as weights. In particular, a *pseudo*– $R_{pk}^2(\theta)$ is computed on the basis of Eq. (10) considering the k th block, for each MV and for each considered quantile θ . This measure, called *Communality* $_{pk}(\theta)$, with $p = 1, \dots, P, k = 1, \dots, K$, indicates how much of the MV' variance can be explained by the corresponding component. The communality of the block k results:

$$Communality_{yk}(\theta) = \frac{1}{P_k} \sum_{pk=1}^{P_k} pseudo-R_{pk}^2(\theta). \tag{18}$$

The quality of the whole outer model is finally obtained through the average of the *Communality* indexes of all the blocks.

It should be noted that, as described in Sect. 2.3, if the quantile in the iterative procedure is set equal to the median to solve the measurement invariance issue, the assessment of the outer model is limited to the quantile $\theta = 0.5$.

Another measure of assessment is provided by the *Redundancy* index, which is defined only for the endogenous block. Please note that low levels of redundancy does not necessarily mean that the structural model is poorly specified. This index only combines the evaluations of both the inner model and the outer model (Lohmöller 1989; Hair et al. 2011), thus can be used as a measure of assessment of the global model, but specific measures for the two sub-models are also needed.

Redundancy can be computed for each endogenous MV or for the whole block, as an average of the redundancies of its MVs. For each MV of the endogenous block $\mathbf{x}_{pk'}$, *Redundancy* is computed multiplying its *Communality* measure by the *pseudo*– R^2 obtained in the corresponding inner model:

$$Redundancy_{pk'}(\theta) = Communality_{pk'}(\theta) \times pseudo-R_{k'}^2(\theta) \tag{19}$$

The overall *Redundancy* of the block k' is obtained averaging the measures associated to the MVs of the endogenous block:

$$Redundancy_{y_{k'}}(\theta) = \sum_{p_{k'}=1}^{P_{k'}} \frac{Redundancy_{p_{k'}}(\theta)}{P_{k'}}. \quad (20)$$

Possible variants could exploit different goodness of fit measures available in the quantile framework as well as the amendment of some assessment indexes proposed in PLS-PM literature (Benitez et al. 2020; Hair et al. 2020, 2017, 2019; Amato et al. 2004).

Two main approaches can be used to evaluate the statistical significance of the coefficients related to the different quantiles. The first approach exploits the asymptotic normal distribution of QR estimators (Koenker and Basset 1978). Such estimators are indeed asymptotically normal, with variance–covariance matrix depending on the model assumptions. Independent and identically distributed errors, independent and not identically distributed errors, and dependent errors determine obvious differences in the variance–covariance matrix. The alternative resorts to bootstrap theory (Efron and Tibshirani 1993), commonly used both in PLS-PM and QR. Bootstrap permits to estimate the standard errors of the coefficients using a distribution free approach. QR literature counts several bootstrap procedures, the xy -pair method (Parzen et al. 1994) being the simplest and widespread solution. It is also known as design matrix bootstrap. Bootstrap standard errors are exploited to compute confidence intervals and to perform hypothesis tests. Resampling methods are also useful in case of small samples. For example, a jackknife approach could be used to estimate the standard errors of the coefficients. Statistical tests could be also easily introduced in QC-PM to test if coefficients at different quantiles can be considered statistically different (Gould 1997). We will not expand on the details here. Readers who are interested can consult Davino et al. (2013) and Furno and Vistocco (2018) for a thorough explanation and all bibliographic indications on inference in QR.

4 A simulation study

A Monte Carlo simulation study has been designed to investigate the performance of QC-PM considering different scenarios. As already stated above, QC-PM and PLS-PM are complementary rather than alternative approaches. Therefore, it is advisable to use both methods in real data application: in many cases, indeed, the focus on conditional mean is not sufficient and a more comprehensive look at the entire conditional distribution is necessary.

Nevertheless, since there is not a real competing method to QC-PM, namely a composite-based approach focusing on conditional quantile, we chose to investigate properties of the QC-PM estimators comparing them with PLS-PM results. It is well known that PLS-PM produces consistent estimates for composite-based model parameters and performs well in the considered scenarios.

4.1 Simulation design and data generation

We operate in the context proposed by Schlittgen et al. (2020), generating data from composite-based populations using the *cbsem* R package (Schlittgen 2019). Determination of the covariance matrix in the procedure proposed by Schlittgen et al. (2020) can be derived considering three scenarios, named formative–formative (ff), formative–reflective (fr) and reflective–reflective (rr). We used the scenario rr, where outer weights are not required. The procedure requires path coefficients, loadings and variances and covariances of exogenous constructs. The parameters must be chosen such that sets of weights can be found to fulfill the equation defining the covariance matrix (see the Vignette from the *cbsem* R package for further details) (Schlittgen 2019).

We set the relationships in the model assuming the theoretical path model represented in Fig. 3 and then we simulated data for the given values of the parameters. The postulated inner model is:

$$\begin{aligned}\xi_2 &= \beta_{20} + \beta_{21}\xi_1 + \zeta_2 \\ \xi_3 &= \beta_{30} + \beta_{31}\xi_1 + \beta_{32}\xi_2 + \zeta_3.\end{aligned}$$

The outer model can be instead written as:

$$\begin{aligned}x_{11} &= \alpha_{11} + \lambda_{11}\xi_1 + \epsilon_{11} \\ x_{21} &= \alpha_{21} + \lambda_{21}\xi_1 + \epsilon_{21} \\ x_{31} &= \alpha_{31} + \lambda_{31}\xi_1 + \epsilon_{31} \\ x_{12} &= \alpha_{12} + \lambda_{12}\xi_2 + \epsilon_{12} \\ x_{22} &= \alpha_{22} + \lambda_{22}\xi_2 + \epsilon_{22} \\ x_{32} &= \alpha_{32} + \lambda_{32}\xi_2 + \epsilon_{32} \\ x_{13} &= \alpha_{13} + \lambda_{13}\xi_3 + \epsilon_{13} \\ x_{23} &= \alpha_{23} + \lambda_{23}\xi_3 + \epsilon_{23} \\ x_{33} &= \alpha_{33} + \lambda_{33}\xi_3 + \epsilon_{33}.\end{aligned}$$

The simulation study considered different scenarios both in the outer and in the inner part of the model. Moreover, the effect of sample size and non-normality distributions were also considered. The cases of homogeneous blocks (no differences among loadings) and heterogeneous blocks (large differences among loadings) were used to assess the outer model. For the inner model, the effect of different correlation levels between constructs was investigated.

That, in short, are the design-factors we considered for the simulation study: sample size, homogeneity of blocks, size effect and variable distributions. In particular, we used the following levels for each design-factor.

Sample sizes. We set $n \in \{50, 100, 200, 300, 400, 500, 1000\}$. The value $n=50$ allows us to investigate the performance of the method in case of application with small sample size. The other values used for n are instead typically encountered

Table 1 Levels of the design-factors considered in the simulation design

Sample size	n = 50, 100, 200, 300, 400, 500, 1000	
Loadings	Homogeneous Blocks For each block k: $\lambda_1 = 0.9$ $\lambda_2 = 0.9$ $\lambda_3 = 0.9$	
	Heterogeneous Blocks For each block k: $\lambda_1 = 0.9$ $\lambda_2 = 0.6$ $\lambda_3 = 0.3$	
Path Coefficients	$\beta_{21} = \beta_{31} = \beta_{32} = 0.2$	
	$\beta_{21} = \beta_{31} = \beta_{32} = 0.3$	
	$\beta_{21} = \beta_{31} = \beta_{32} = 0.4$	
	$\beta_{21} = \beta_{31} = \beta_{32} = 0.5$	
Manifest variables distribution	Data generated from normal distribution (skew=0, kurt=0)	
	Data generated from exponential distribution (skew=0, kurt=6)	

in research applications, the largest values being useful to study the asymptotic properties of QC-PM estimators.

Loadings. We set $\lambda_{pk} = 0.9$ ($p = 1, 2, 3; k = 1, 2, 3$) for homogeneous blocks and loadings $\lambda_{1k} = 0.9, \lambda_{2k} = 0.6, \lambda_{3k} = 0.3$ ($k = 1, 2, 3$) for heterogeneous blocks, these last values to reflect very large differences among loadings.

Path coefficients. We set, for all inner relationships, $\beta_{k'k} \in \{0.2, 0.3, 0.4, 0.5\}$ to take into account different levels of correlations among constructs.

Skewness and Kurtosis. We set both equal to 0 for normal distribution, while skewness = 2 and kurtosis = 6 were used for mimicking exponential distributions.

The total number of scenarios obtained from the combination of the above described levels of the design-factors is equal to 112 (7 sample sizes \times 2 loadings \times 4 path coefficients \times 2 skewness and kurtosis). For each considered scenario, we generated 500 replications.

A synthesis schema of the simulation design is offered in Table 1.

The R software environment (R Core Team 2020) for statistical computing were used to generate and analyze data.

Data with non-normal distribution were generated using the technique described in Vale and Maurelli (1983), who extended the method proposed by Fleishman (1978).

The performance of QC-PM was assessed considering the Relative Bias (*RBias*) and the Root Mean Square Error (*RMSE*) of the estimates on the basis of the 500 replications. *RBias* was computed as:

$$RBias = \frac{1}{S} \sum_{s=1}^S \frac{(\hat{\theta}_s - \theta)}{\theta} \quad s = 1, 2, \dots, 500$$

where S represents the number of replications in the simulation, $\hat{\theta}_s$ is the estimate for the generic replication, and θ is the corresponding population parameter. Instead, $RMSE$ was computed as:

$$\sqrt{\frac{1}{S} \sum_{s=1}^S (\hat{\theta}_s - \theta)^2} \quad s = 1, 2, \dots, 500$$

Clearly, because $MSE = Var(\hat{\theta}) + bias(\hat{\theta})^2$, $RMSE$ entails information on both bias and variability of the estimates.

4.2 Simulation results

In presenting simulation results, we choose to focus on the effect of sample size on the bias and efficiency, and, consequently, on the consistency of the estimates. Since the number of considered scenarios (112) is too large, in the following we focus only on the more interesting and enlightening scenarios. In particular, we present results for all the considered levels of loadings in the three blocks of variables for PLS-PM and for QC-PM at quantile $\theta = 0.5$ (measurement model was indeed restricted to the median regression model).

Additionally, since there are interesting differences between homogeneous blocks and heterogeneous blocks, we reported results for both the cases.

For the inner model, we present results for all the three path coefficient estimates obtained through PLS-PM and QC-PM at quantiles $\theta \in \{0.25, 0.5, 0.75\}$, but only for $\beta_{k'k} = 0.3$, a small/moderate effect. Indeed, as known in literature (Tenenhaus 2008), results are similar as correlations between composites increases.

The following subsections details results according to the three considered factors, that is sample size, level of heterogeneity within blocks and degree of skewness/kurtosis of the distribution. The resulting scenarios are organized in three groups:

- *Group 1*, focusing at the effect of sample size,
- *Group 2*, focusing at the effect of the level of heterogeneity within blocks,
- *Group 3*, focusing at the effect of the degree of skewness/kurtosis of the distribution.

The effect of sample size

The first set of considered scenarios (from hereinafter *Group 1*) allows us to focus only on the effect of sample size, neutralising, as far as possible the effect of the other factors: outer blocks are considered homogeneous with high correlations among the MVs (all λ values set equal to 0.9), data generated from normal distributions and path coefficients set equal to 0.3. These settings result in seven scenarios for Group 1.

Figure 4 shows the distribution of loadings for such scenarios for the outcome block (Block 3). The results for the other two blocks are not shown since they do not differ much from those of Block 3. The two columns of Fig. 4 depict the results for PLS-PM and QC-PM (recall that we set $\theta=0.5$ for the outer model). The seven scenarios of the group, corresponding to the different values of the sample size n , are represented on

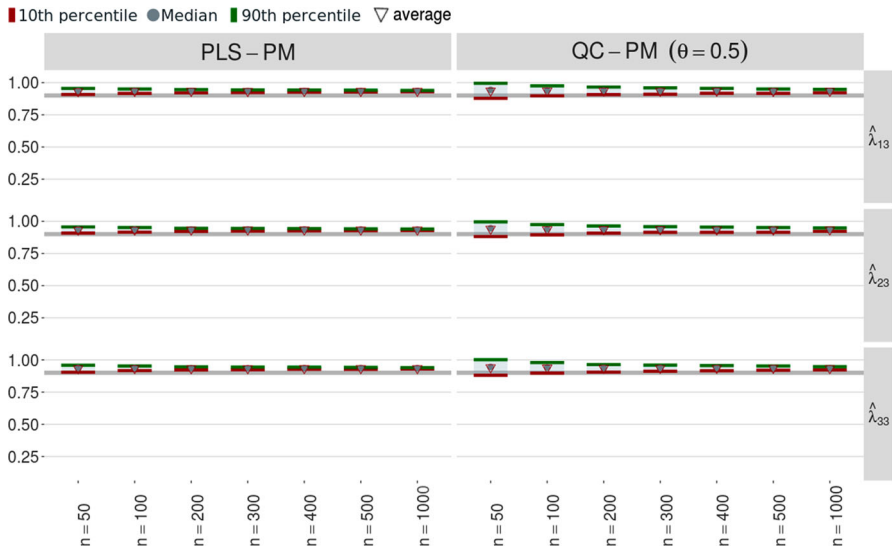


Fig. 4 Diminutive distribution charts of the loadings of scenarios belonging to *Group 1* with $n \in \{50, 100, 200, 300, 400, 500, 1000\}$, homogeneous blocks, path coefficients equal to 0.3 and Normal distributed data

the horizontal axis. Finally, the rows refer to the coefficients $\lambda_{13}, \lambda_{23}, \lambda_{33}$ associated with the MVs of the block.

Data are represented through diminutive distribution charts (Rudis 2019), a variant of boxplots aimed to visualize distribution characteristics: each box ranges from the 10th percentile to the 90th percentile, the triangle indicates the mean value of the distribution and the circle the median. Both PLS-PM and QC-PM show distributions that converge to the true parameter value (horizontal line at the value 0.9) as n increases. Moreover, the variability of the estimates is rather small, although for QC-PM slightly higher. Regarding bias, it is interesting to note, that for the quantile model, the 10th percentile has a smaller distance from the true parameter than the PLS-PM, whatever the value of n .

Table 2 reports the values of *RBias* and *RMSE* for each loading distribution. In particular, the table shows the average values of *RBias* and *RMSE* for each block (columns) and each scenario (rows). The table is row-partitioned according to the model. The values of *RBias* and *RMSE* for the loading estimates do not change substantially between QC-PM and PLS-PM when blocks are homogeneous and variables are normally distributed, with large sample sizes (at least 300 observations). At the lowest considered sample size ($n = 50$), QC-PM always shows higher bias compared to PLS-PM and such behavior is also confirmed combining bias with variability of the estimates (see *RMSE* columns).

The distribution of the path coefficients (Fig. 5) is also affected by sample size and shows less marked differences between the two methods (the comparison between Figs. 4 and 5 must be done with caution because the two figures have different vertical scales). As in Fig. 4, each diminutive distribution chart refers to a scenario (horizontal axis) but here the three rows refer to the path coefficient ($\beta_{21}, \beta_{31}, \beta_{32}$). Each vertical

Table 2 Values of *RBias* and *RMSE* for loadings of the scenarios belonging to *Group 1* with $n \in \{50, 100, 200, 300, 400, 500, 1000\}$, homogeneous blocks, path coefficients equal to 0.3 and Normal distributed data (levels of the varying factor, sample size, are in the first column)

	<i>RBias</i>			<i>RMSE</i>		
	Block X1	Block X2	Block X3	Block X1	Block X2	Block X3
<i>PLS-PM</i>						
$n = 50$	0.035	0.036	0.037	0.040	0.042	0.039
$n = 100$	0.039	0.039	0.038	0.038	0.038	0.037
$n = 200$	0.038	0.038	0.038	0.036	0.036	0.035
$n = 300$	0.038	0.038	0.038	0.035	0.035	0.035
$n = 400$	0.038	0.038	0.038	0.035	0.035	0.035
$n = 500$	0.039	0.038	0.038	0.035	0.035	0.035
$n = 1000$	0.038	0.038	0.038	0.035	0.035	0.035
<i>QC-PM</i>						
$n = 50$	0.043	0.039	0.044	0.066	0.082	0.066
$n = 100$	0.040	0.043	0.042	0.056	0.052	0.050
$n = 200$	0.041	0.039	0.041	0.044	0.042	0.043
$n = 300$	0.039	0.039	0.040	0.040	0.040	0.040
$n = 400$	0.039	0.039	0.039	0.039	0.039	0.039
$n = 500$	0.040	0.039	0.039	0.039	0.038	0.038
$n = 1000$	0.039	0.039	0.039	0.036	0.036	0.036

Results for PLS-PM and QC-PM estimated at $\theta = 0.5$ are reported each scenario

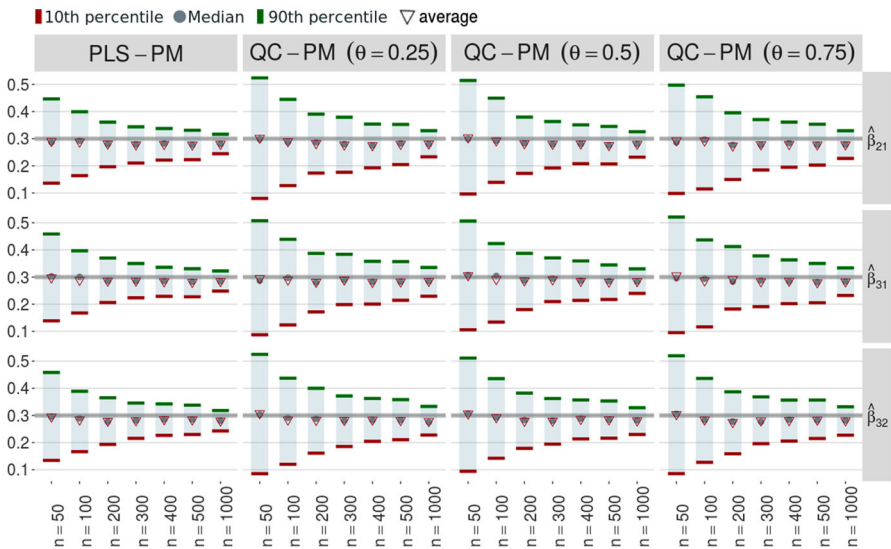


Fig. 5 Diminutive distribution charts of the path coefficients of scenarios belonging to *Group 1* with $n \in \{50, 100, 200, 300, 400, 500, 1000\}$, homogeneous blocks, path coefficients equal to 0.3 and Normal distributed data

Table 3 Values of *RBias* and *RMSE* for the path coefficients of scenarios belonging to *Group 1* with $n \in \{50, 100, 200, 300, 400, 500, 1000\}$, homogeneous blocks, path coefficients equal to 0.3 and Normal distributed data (the levels of the varying factor, the sample size, are in the first column)

	<i>RBias</i>			<i>RMSE</i>		
	β_{21}	β_{31}	β_{32}	β_{21}	β_{31}	β_{32}
<i>PLS-PM</i>						
$n = 50$	-0.031	-0.007	-0.016	0.120	0.121	0.122
$n = 100$	-0.039	-0.037	-0.053	0.092	0.091	0.088
$n = 200$	-0.062	-0.047	-0.068	0.067	0.066	0.069
$n = 300$	-0.074	-0.046	-0.062	0.058	0.051	0.055
$n = 400$	-0.068	-0.055	-0.049	0.050	0.047	0.048
$n = 500$	-0.075	-0.062	-0.052	0.047	0.044	0.044
$n = 1000$	-0.066	-0.053	-0.065	0.035	0.033	0.035
<i>QC-PM $\theta = 0.25$</i>						
$n = 50$	0.008	-0.015	0.026	0.163	0.160	0.163
$n = 100$	-0.033	-0.033	-0.053	0.123	0.122	0.123
$n = 200$	-0.057	-0.061	-0.057	0.090	0.089	0.094
$n = 300$	-0.074	-0.033	-0.058	0.079	0.072	0.074
$n = 400$	-0.083	-0.064	-0.054	0.071	0.065	0.062
$n = 500$	-0.066	-0.058	-0.059	0.063	0.058	0.060
$n = 1000$	-0.063	-0.050	-0.072	0.043	0.043	0.048
<i>QC-PM $\theta = 0.5$</i>						
$n = 50$	0.014	0.020	0.021	0.160	0.150	0.154
$n = 100$	-0.024	-0.028	-0.025	0.118	0.111	0.115
$n = 200$	-0.059	-0.043	-0.069	0.083	0.079	0.084
$n = 300$	-0.066	-0.030	-0.066	0.071	0.064	0.070
$n = 400$	-0.062	-0.050	-0.045	0.060	0.059	0.060
$n = 500$	-0.082	-0.059	-0.051	0.060	0.053	0.055
$n = 1000$	-0.066	-0.052	-0.063	0.043	0.039	0.042
<i>QC-PM $\theta = 0.75$</i>						
$n = 50$	-0.023	0.018	0.013	0.152	0.161	0.161
$n = 100$	-0.025	-0.044	-0.053	0.128	0.123	0.119
$n = 200$	-0.087	-0.027	-0.084	0.096	0.088	0.093
$n = 300$	-0.072	-0.054	-0.063	0.077	0.074	0.071
$n = 400$	-0.069	-0.048	-0.060	0.068	0.063	0.065
$n = 500$	-0.074	-0.066	-0.055	0.063	0.061	0.058
$n = 1000$	-0.073	-0.057	-0.063	0.047	0.043	0.045

block refers to a model: PLS-PM and QC-PM for $\theta \in \{0.25, 0.5, 0.75\}$. It is interesting, in the case of the path coefficients, the correspondence of the mean and median values of the estimates with the true value of the parameter (horizontal line at 0.3).

Table 3 shows, for each horizontal block, the results obtained from the corresponding model in terms of *RBias* and *RMSE* values for the path coefficient estimates, in the

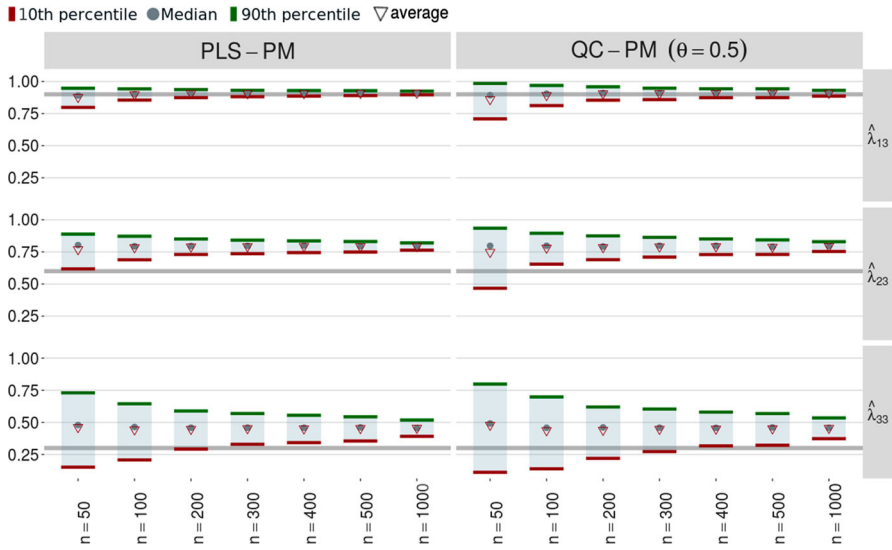


Fig. 6 Diminutive distribution charts of the loadings of scenarios belonging to *Group 2* with $n \in \{50, 100, 200, 300, 400, 500, 1000\}$, heterogeneous blocks, path coefficients equal to 0.3 and Normal distributed data

seven scenarios. PLS-PM and QC-PM perform very similar both in terms of bias and efficiency. In general, as sample size increases, *RMSE* decreases for all path coefficient estimates. The *RMSE* is slightly higher in QC-PM at quantile $\theta = 0.25$ and $\theta = 0.75$.

The effect of the level of heterogeneity

The convergence of estimates as n increases is also confirmed in the case of heterogeneous blocks. Figure 6 shows the diminutive distribution charts of the loadings of the scenarios belonging to the second group of scenarios, *Group 2*, still encompassing normal distributions and with path coefficients equal to 0.3 but with MVs differently correlated to each construct (0.9, 0.6 and 0.3).

In this group of scenarios, unlike *Group 1*, the distribution of loadings gets closer to the true population parameter especially for λ_{13} . However, the heterogeneity of the blocks has a distortive effect on the estimates of the path coefficients (Fig. 7) both in case of PLS-PM and QC-PM. Tables 4 and 5 confirm decreasing bias and variability as n increases. In this case but both *RBias* and *RMSE* are always higher than the homogeneous case.

The effect of the degree of skewness/kurtosis

The third group of scenarios, *Group 3*, worth to be mentioned aims to show the effect of an asymmetric distribution in the data generation process. Also this group includes seven scenarios (varying the sample size), with homogeneous loadings (equal to 0.9) and path coefficients equal to 0.3. Data are here generated by an Exponential distribution. Convergence is confirmed for this group as well, but there are differences between

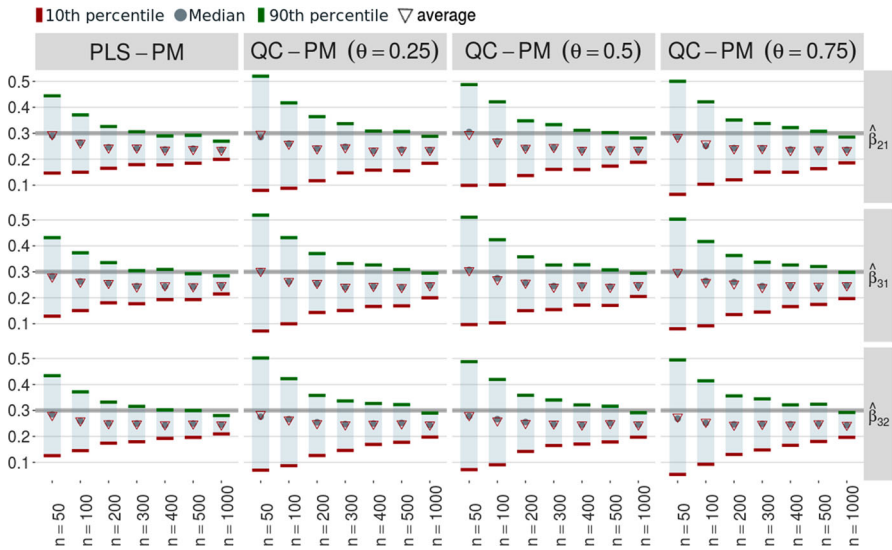


Fig. 7 Diminutive distribution charts of the path coefficients of scenarios belonging to *Group 2* with $n \in \{50, 100, 200, 300, 400, 500, 1000\}$, heterogeneous blocks, path coefficients equal to 0.3 and Normal distributed data

Table 4 Values for the *RBias* and *RMSE* for the loadings (averages for each block) of the scenarios belonging to *Group 2* with $n \in \{50, 100, 200, 300, 400, 500, 1000\}$, heterogeneous blocks, path coefficients equal to 0.3 and Normal distributed data (the levels of the varying factor, the sample size, are in the first column)

	<i>RBias</i>			<i>RMSE</i>		
	Block X1	Block X2	Block X3	Block X1	Block X2	Block X3
<i>PLS-PM</i>						
$n = 50$	0.307	0.295	0.282	0.208	0.205	0.186
$n = 100$	0.265	0.282	0.262	0.160	0.164	0.153
$n = 200$	0.274	0.252	0.272	0.141	0.138	0.138
$n = 300$	0.274	0.274	0.279	0.134	0.133	0.133
$n = 400$	0.272	0.273	0.279	0.130	0.129	0.129
$n = 500$	0.282	0.275	0.281	0.129	0.129	0.128
$n = 1000$	0.280	0.283	0.284	0.124	0.125	0.124
<i>QC-PM</i>						
$n = 50$	0.304	0.297	0.296	0.248	0.247	0.230
$n = 100$	0.258	0.277	0.256	0.188	0.193	0.178
$n = 200$	0.270	0.244	0.261	0.155	0.153	0.151
$n = 300$	0.266	0.271	0.272	0.146	0.145	0.144
$n = 400$	0.267	0.268	0.275	0.138	0.138	0.137
$n = 500$	0.276	0.277	0.276	0.136	0.136	0.134
$n = 1000$	0.277	0.282	0.283	0.128	0.129	0.128

For each scenario, results for PLS-PM and QC-PM estimated at $\theta = 0.5$ are shown

Table 5 Path coefficients *RBias* and *RMSE* values of scenarios belonging to *Group 2* with $n = [50, 100, 200, 300, 400, 500, 1000]$, heterogeneous blocks, path coefficients equal to 0.3 and Normal distributed data (the levels of the varying factor, the sample size, are in the first column)

	<i>RBias</i>			<i>RMSE</i>		
	β_{21}	β_{31}	β_{32}	β_{21}	β_{31}	β_{32}
<i>PLS-PM</i>						
$n = 50$	-0.015	-0.064	-0.061	0.117	0.117	0.120
$n = 100$	-0.124	-0.130	-0.131	0.095	0.095	0.100
$n = 200$	-0.186	-0.145	-0.167	0.084	0.075	0.080
$n = 300$	-0.186	-0.191	-0.170	0.075	0.076	0.073
$n = 400$	-0.213	-0.176	-0.180	0.078	0.070	0.070
$n = 500$	-0.207	-0.188	-0.170	0.074	0.069	0.065
$n = 1000$	-0.217	-0.172	-0.183	0.070	0.059	0.061
<i>QC-PM $\theta = 0.25$</i>						
$n = 50$	-0.009	0.010	-0.047	0.167	0.167	0.165
$n = 100$	-0.137	-0.120	-0.119	0.129	0.128	0.133
$n = 200$	-0.197	-0.145	-0.166	0.111	0.098	0.104
$n = 300$	-0.183	-0.194	-0.182	0.092	0.092	0.091
$n = 400$	-0.226	-0.183	-0.174	0.092	0.084	0.081
$n = 500$	-0.220	-0.198	-0.169	0.087	0.080	0.076
$n = 1000$	-0.218	-0.175	-0.183	0.076	0.065	0.067
<i>QC-PM $\theta = 0.5$</i>						
$n = 50$	-0.012	0.019	-0.063	0.148	0.158	0.156
$n = 100$	-0.108	-0.097	-0.131	0.122	0.126	0.128
$n = 200$	-0.190	-0.142	-0.164	0.100	0.091	0.096
$n = 300$	-0.179	-0.194	-0.171	0.085	0.088	0.085
$n = 400$	-0.216	-0.178	-0.183	0.087	0.081	0.080
$n = 500$	-0.210	-0.196	-0.166	0.082	0.080	0.073
$n = 1000$	-0.213	-0.173	-0.182	0.074	0.063	0.065
<i>QC-PM $\theta = 0.75$</i>						
$n = 50$	-0.049	-0.004	-0.082	0.161	0.162	0.167
$n = 100$	-0.133	-0.133	-0.149	0.128	0.131	0.133
$n = 200$	-0.196	-0.152	-0.186	0.106	0.100	0.103
$n = 3000$	-0.194	-0.198	-0.175	0.095	0.095	0.092
$n = 4000$	-0.219	-0.174	-0.185	0.092	0.083	0.083
$n = 500$	-0.211	-0.182	-0.168	0.085	0.078	0.076
$n = 1000$	-0.219	-0.175	-0.187	0.077	0.067	0.068

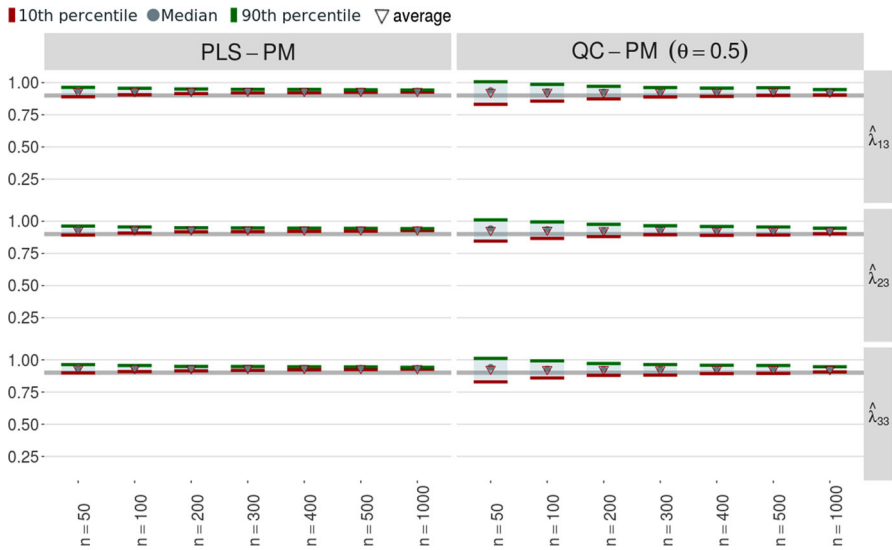


Fig. 8 Diminutive distribution charts of the loadings of scenarios belonging to *Group 3* with $n \in \{50, 100, 200, 300, 400, 500, 1000\}$, homogeneous blocks, path coefficients equal to 0.3 and exponential distributed data

the PLS-PM and QC-PM. As regards loadings (Fig. 8), the distributions are always more variable in QC-PM than in PLS-PM. Nevertheless QC-PM always manages to capture the true parameter value within the 90% of central values. Considering the average values of *RBias* and *RMSE* in all blocks (Table 6), better performance of QC-PM in terms of efficiency and unbiasedness is confirmed especially for larger sample sizes. Looking at the distribution of path coefficient estimates (Fig. 9), we note the ability of QC-PM to capture the positive skewness of the distribution used to generate the data: the variability of the estimates is smaller for $\theta = 0.25$ and larger in the right tail ($\theta = 0.75$) and the parameter is overestimated at the lowest quantile and underestimated at the highest quantile. For $\theta = 0.5$ both methods provide unbiased estimates. The *RBias* values in Table 7 confirm the reduction in the bias as n increases. More complex is the interpretation of the *RMSE* values, which combine bias and variability: the estimates at quantile 0.25, for example, are more biased but less variable than those at quantile 0.5, so the *RMSE* is affected by a kind of trade-off between variability and bias.

5 An application on Chronic Kidney Disease in diabetic patients

QC-PM potentialities are described through an artificial dataset which simulates a study on CKD in diabetic patients. The original study was proposed by Wang et al. (2016) who used real data to examine the potential risk factors of CKD through a quantile approach to factor-based SEM. In particular, data were generated mimicking the model and estimates obtained by Wang et al. (2016), since the original data were

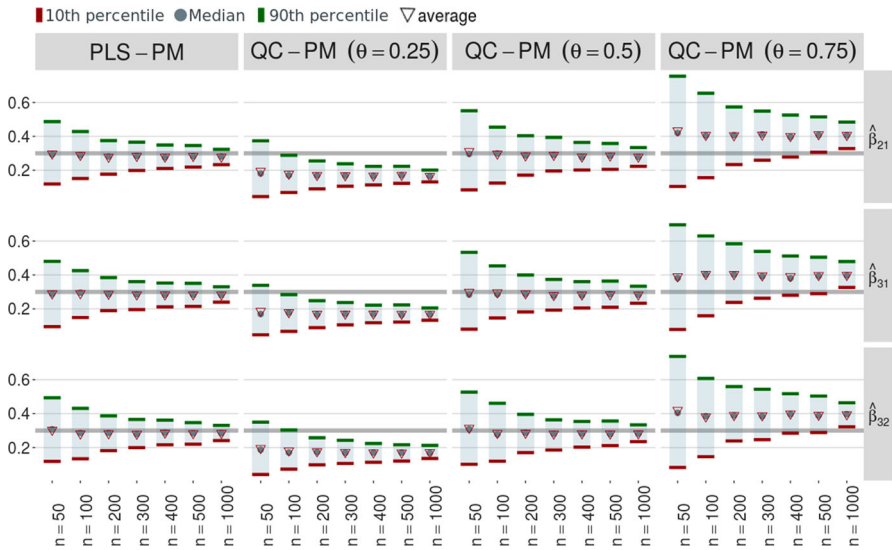


Fig. 9 Diminutive distribution charts of the path coefficients of *Group 3* of scenarios belonging to *Group 3* with $n \in \{50, 100, 200, 300, 400, 500, 1000\}$, homogeneous blocks, path coefficients equal to 0.3 and exponential distributed data

Table 6 Values of *RBias* and *RMSE* for the loading (averages for each block) of the scenarios belonging to *Group 3* with $n \in \{50, 100, 200, 300, 400, 500, 1000\}$, homogeneous blocks, path coefficients equal to 0.3 and Exponential distributed data (the levels of the varying factor, the sample size, are in the first column)

	BIAS			RMSE		
	Block X1	Block X2	Block X3	Block X1	Block X2	Block X3
<i>PLS-PM</i>						
$n = 50$	0.033	0.034	0.034	0.051	0.047	0.045
$n = 100$	0.037	0.039	0.037	0.039	0.040	0.039
$n = 200$	0.038	0.038	0.037	0.037	0.037	0.036
$n = 300$	0.038	0.038	0.038	0.036	0.036	0.036
$n = 400$	0.038	0.038	0.038	0.036	0.036	0.036
$n = 500$	0.039	0.038	0.038	0.036	0.035	0.035
$n = 1000$	0.039	0.038	0.038	0.035	0.035	0.035
<i>QC-PM</i>						
$n = 50$	0.032	0.031	0.032	0.081	0.086	0.085
$n = 100$	0.030	0.030	0.032	0.058	0.058	0.059
$n = 200$	0.029	0.029	0.029	0.047	0.046	0.046
$n = 300$	0.029	0.030	0.030	0.040	0.040	0.041
$n = 400$	0.029	0.029	0.030	0.037	0.037	0.037
$n = 500$	0.031	0.029	0.030	0.036	0.035	0.036
$n = 1000$	0.030	0.029	0.029	0.032	0.031	0.031

For each scenario, results for PLS-PM and QC-PM estimated at $\theta = 0.5$ are shown

Table 7 Values of *RBias* and *RMSE* for the Path coefficients of the scenarios belonging to *Group 3* with $n \in \{50, 100, 200, 300, 400, 500, 1000\}$, homogeneous blocks, path coefficients equal to 0.3 and Exponential distributed data (the levels of the varying factor, the sample size, are in the first column)

	BIAS			RMSE		
	β_{21}	β_{31}	β_{32}	β_{21}	β_{31}	β_{32}
<i>PLS-PM</i>						
$n = 50$	-0.007	-0.032	0.012	0.139	0.148	0.144
$n = 100$	-0.035	-0.028	-0.061	0.108	0.111	0.114
$n = 200$	-0.073	-0.040	-0.057	0.080	0.079	0.082
$n = 300$	-0.055	-0.054	-0.067	0.066	0.067	0.069
$n = 400$	-0.070	-0.054	-0.045	0.058	0.057	0.058
$n = 500$	-0.055	-0.054	-0.055	0.053	0.053	0.053
$n = 1000$	-0.070	-0.055	-0.051	0.041	0.039	0.038
<i>QC-PM $\theta = 0.25$</i>						
$n = 50$	-0.349	-0.376	-0.348	0.163	0.163	0.160
$n = 100$	-0.406	-0.403	-0.395	0.150	0.149	0.150
$n = 200$	-0.430	-0.432	-0.410	0.144	0.144	0.139
$n = 300$	-0.436	-0.430	-0.426	0.141	0.140	0.138
$n = 400$	-0.443	-0.430	-0.435	0.139	0.136	0.137
$n = 500$	-0.428	-0.435	-0.437	0.134	0.137	0.136
$n = 1000$	-0.449	-0.436	-0.429	0.137	0.134	0.132
<i>QC-PM $\theta = 0.50$</i>						
$n = 50$	0.039	-0.001	0.049	0.175	0.169	0.164
$n = 100$	-0.010	-0.013	-0.052	0.133	0.124	0.129
$n = 200$	-0.045	-0.031	-0.048	0.092	0.085	0.088
$n = 300$	-0.035	-0.065	-0.064	0.080	0.075	0.076
$n = 400$	-0.067	-0.055	-0.059	0.066	0.065	0.065
$n = 500$	-0.048	-0.052	-0.062	0.062	0.059	0.061
$n = 1000$	-0.071	-0.055	-0.059	0.047	0.042	0.042
<i>QC-PM $\theta = 0.75$</i>						
$n = 50$	0.446	0.303	0.401	0.277	0.247	0.273
$n = 100$	0.358	0.349	0.273	0.217	0.210	0.193
$n = 200$	0.351	0.347	0.300	0.171	0.169	0.159
$n = 300$	0.363	0.322	0.295	0.154	0.146	0.143
$n = 400$	0.335	0.307	0.329	0.140	0.130	0.133
$n = 500$	0.370	0.327	0.304	0.139	0.128	0.124
$n = 1000$	0.357	0.330	0.315	0.124	0.115	0.111

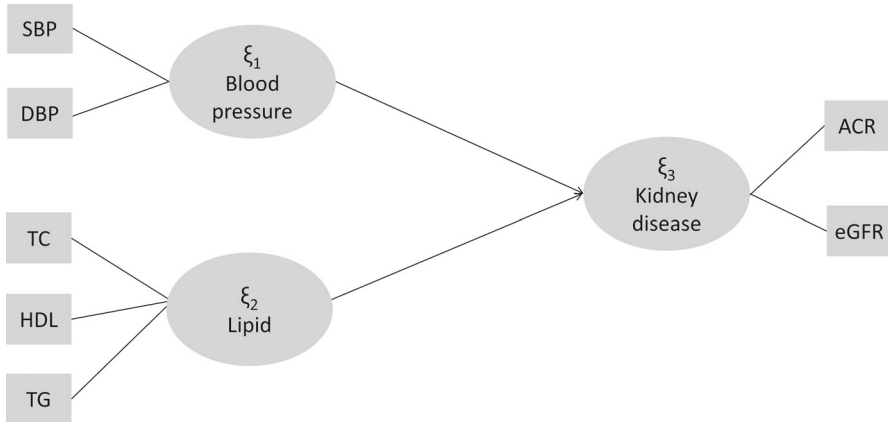


Fig. 10 Theoretical model for Chronic Kidney Disease data, following Wang et al. (2016)

not available. Even if we use artificial data, the involved variables and their relations are in line with the study by Wang et al., allowing a clear practical interpretation of results. Both studies are quantile based, even if Wang et al. exploited factor-based SEM while we focus on composite-based path modeling.

The main objective of this section is to show QC-PM in action, stressing its complementarity with the traditional methods for composite-based path modeling (PLS-PM), which focus only on conditional means. The advantage of using artificial data allows us to obtain a scenario where relationships among variables change with quantiles (i.e., there are different relations considering the different parts of the dependent variable distributions). Our main objective was not recovering parameters, but evaluate if QC-PM is able to detect this heterogeneity in the variable relationships.

5.1 Data description

This application aims to study the effect of some risk factors on CKD. We started from the original path model in Wang et al. (2016) and removed the non significant predictors. In particular, the study investigates Type 2 diabetic patients who might have experienced CKD. Data consist of 300 patients. Diagnosis and staging of CKD were based on urinary albumin-creatinine ratio (*ACR*) and estimated glomerular filtration rate (*eGFR*). These two variables are the MVs of the outcome block named *Kidney disease*. The considered risk factors were *Blood pressure* and *Lipid*. The former was measured by systolic blood pressure (*SBP*) and diastolic blood pressure (*DBP*), while the latter by total cholesterol (*TC*), high-density lipoprotein (*HDL*), and triglycerides (*TG*). Therefore, the inner model underpinning our design and subsequent analyses consists of two exogenous constructs, *Blood pressure* (ξ_1) and *Lipid* (ξ_2), and one endogenous construct, *Kidney disease* (ξ_3). Figure 10 depicts the corresponding path diagram.

Data generation process exploits the classical covariance-based approach for SEM. As above mentioned, the results of the original study by Wang et al. (2016) represent the

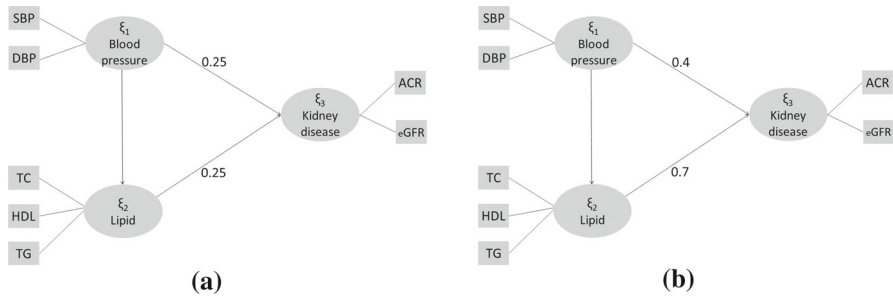


Fig. 11 Theoretical models for the simulated data: patients with low **(a)** and high **(b)** severity of kidney disease

starting point for the generation of artificial data. Therefore, the parameters of the SEM are set to the values of the model estimated in that study. In particular, *Blood pressure* was positively correlated with the severity of *Kidney disease*, and the correlation was stronger for higher quantiles. *Lipid* was found to be positively correlated with *Kidney disease* and, also for this variable, the correlation was stronger for higher quantiles. The resulting variance–covariance matrix characterizes the multivariate distribution used to generate data. The generation process was carried out using the software EQS 6.1 (Bentler 2006), computation and analysis using R (R Core Team 2020).

Heterogeneity in the inner model was introduced in the artificial data assuming that the exogenous constructs exert a different effect on the different parts of the endogenous construct distribution. It results that the path coefficients differ across quantiles. In order to generate data with these features, we supposed that two different populations exist, and for each population the model parameters are different. In particular we divided the patients in two groups. The first group was represented by patients with low severity of *kidney disease*, and thus the relationship between *kidney disease* and each of the two exogenous constructs is weaker (Fig. 11a). The second group was represented by patients with high severity of *kidney disease*: in such a case, the relationship between *kidney disease* and each exogenous constructs is stronger (Fig. 11b). In order to focus only on heterogeneity in the inner model, as in Wang et al. (2016), the loadings between constructs and the corresponding MVs were set all equal to 1 for both the populations.

The simulation procedure was articulated in the following three steps:

1. data were generated from a multivariate normal population, $\mathbf{X} \sim N(\mathbf{0}, \Sigma)$, where Σ is the population covariance matrix using the values in Fig. 11a for the parameters of the model. The sample size was set equal to 300. For each of the two MVs of *Kidney disease* block (*ACR* and *eGFR*), we removed the observations higher than the quantile 0.6 of the same MV. In other words, once the observations were sorted in non decreasing order with respect to the values on each MV, we kept the first 60% of observations, i.e. the first 180 observations. Then, the MVs of the endogenous block were transformed in order to have realistic values ranging from 1 to 6 for *ACR*, and from 50 to 90 for *eGFR*;
2. data were generated from a multivariate normal population, $\mathbf{X} \sim N(\mathbf{0}, \Sigma)$, where Σ is the population covariance matrix using the values in Fig. 11b for the parameters of the model. The sample size was set again equal to 300. For each of the

Table 8 Check for block unidimensionality and internal consistency

	MVs	Eig. 1st	Eig. 2nd	C.alpha	DH.rho
Blood pressure	2	1.73	0.266	0.847	0.99
Lipid	3	2.34	0.359	0.860	0.99
CKD	2	1.79	0.205	0.886	0.99

two MVs of the *Kidney disease* block, we kept the 40% of central observations around the MV mean (100 units), namely 20% on the left-neighborhood of the mean and the other 20% on its right-neighborhood. The two resultings MVs were transformed so to have values ranging from 6 to 10 for *ACR*, and from 90 to 120 for *eGFR*;

- the two data sets generated at the previous steps were stacked obtaining an unique data set with sample size 300. Note that the MVs of the exogenous blocks in the two models come from the same population, while obviously the same does not hold for the MVs of the endogenous block.

According to such data generation process, we expect that QC-PM provides estimates for the parameters of model (a) for quantiles smaller than 0.6, and estimates for the parameters of model (b) for quantiles larger than 0.6.

5.2 Results

This section describes a complete application of QC-PM, from the preliminary analysis to the evaluation of the goodness of fit. The aim is to illustrate the potential of the method along with the guidelines for the interpretation of the results.

An initial inspection of unidimensionality and internal consistency of blocks was performed. To check unidimensionality, we carried out a principal component analysis for each block of MVs. If a block is unidimensional, the first eigenvalue is expected to be the only one greater than 1 and much higher than the second one.

The internal consistency of each block of MVs was instead evaluated through the Cronbach's α . Such index assumes equal population covariances among the indicators of one block, and such assumption is likely not met in empirical research. However, this index can be used as a lower bound for reliability (Benitez et al. 2020). We also consider Dijkstra–Henseler's ρ (Dijkstra and Henseler 2015) to evaluate composite reliability. Table 8 shows that all the blocks are unidimensional and internally consistent. The method used to obtain the artificial data in Sect. 5 provides equal loadings and therefore the values of Dijkstra–Henseler's ρ (DH.rho) are all equal to 0.99.

Model parameters were estimated through PLS-PM and QC-PM by setting the quantile in the iterative procedure to the median and considering a dense grid of quantiles in the inner model. The two panels in Fig. 12 show the different QC-PM path coefficient estimates across quantiles and the PLS-PM path coefficient estimates: Fig. 12a depicts the path coefficient connecting *Blood pressure* to *Kidney disease*, while Fig. 12b refers to *Lipid*. In particular, quantiles are represented on the horizontal axis and coefficients on the vertical axis. The horizontal solid lines represent the PLS-

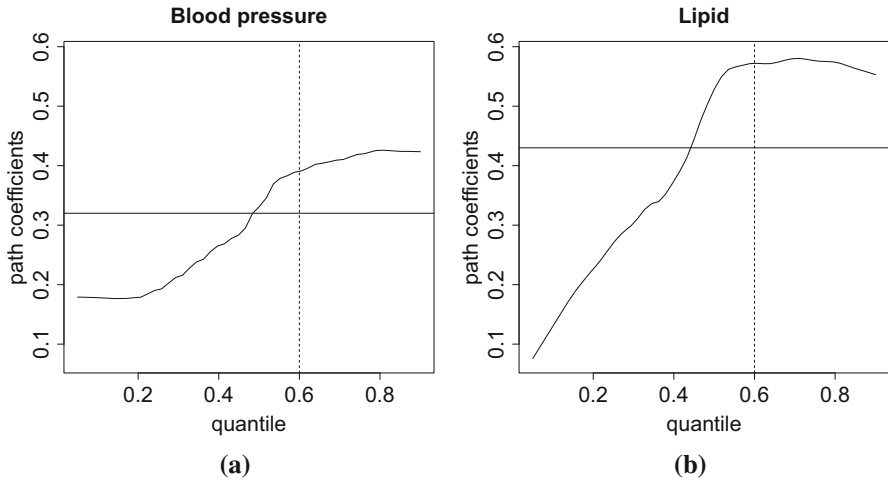


Fig. 12 Path coefficient estimates (y-axis) across quantiles (x-axis) linking *Blood pressure* (a) and *Lipid* (b) to *Kidney disease*. The horizontal solid lines represent the PLS-PM estimates and the vertical dotted line drawn at quantile 0.6 refer to the threshold used in the data generation process

Table 9 Path coefficients and corresponding standard error from a classical PLS-PM (first row) and from a QC-PM applied on the inner model for a selected set of quantiles ($\theta \in \{0.25, 0.50, 0.75\}$)

Quantile	Blood pressure (\pm standard error)	Lipid (\pm standard error)
PLSPM	0.320* (\pm 0.051)	0.426* (\pm 0.048)
0.25	0.189* (\pm 0.062)	0.263* (\pm 0.096)
0.50	0.291* (\pm 0.090)	0.547* (\pm 0.101)
0.75	0.416* (\pm 0.051)	0.588* (\pm 0.053)

* $p < 0.05$

PM estimates while the broken lines represent the QC-PM estimates over quantiles. The vertical dotted line drawn at quantile 0.6 in each figure refers to the threshold used in the data generation process (we expect that for quantiles smaller than 0.6, QC-PM produces estimates for the parameters specified in the model shown in the Fig. 11a, while for quantiles larger than 0.6, QC-PM produces estimates for the parameters specified in the model shown in the Fig. 11b). Figure 12 shows the ability of the QC-PM to detect the structure underlying the simulated data (Fig. 11). QC-PM was indeed able to distinguish the different effects in the different parts of the *Kidney disease* distribution: both path coefficients increase with quantiles and results are consistent with the true values specified in the population models shown in Fig. 11. Table 9 reports the path coefficient estimates (\pm standard errors) obtained using PLS-PM (first row) and QC-PM for the quantiles $\theta \in \{0.25, 0.50, 0.75\}$. Standard errors of estimates were obtained using bootstrap. All path coefficients were statistically significant. On the whole, except for coefficients at $\theta=0.75$, PLS-PM estimates are slightly more efficient than QC-PM ones. This is in line with theory: just like mean is more efficient than median, OLS regression estimates are usually more efficient than QR estimates. Both *Blood pressure* and *Lipid* have a positive impact on *Kidney disease*, which increases for

patients with higher levels of CKD. This positive and increasing effect is well-known in literature: hypertension, high presence of cholesterol, lipoprotein and triglycerides are all considered leading causes of CKD (Bakris and Ritz 2009).

The assessment of the models is carried out using the measures introduced in Sect. 3. It is worth to recall again that a direct comparison of the measures of fit of the two methods is not appropriate, since the two methods optimize different criteria. Hence, the objective is neither to compare PLS-PM and QC-PM results nor to identify the best model. Instead, we aim to illustrate how to use the measures define above for the assessment of QC-PM results.

With respect to the inner model, PLS-PM produces an R^2 equal to 0.214, while QC-PM provides *pseudo*- R^2 values increasing from lower to higher quantiles (0.051, 0.123, 0.185), for the quantiles $\theta \in \{0.25, 0.50, 0.75\}$. This result was expected and coherent with the data structure, as relationships among constructs increase with quantiles. The assessment of the outer model is carried out in two steps. Table 10 shows, for each block, the communality values related to each MV and to the whole block (in bold) both for PLS-PM and QC-PM. For the latter, obviously, estimates refer only to the median because, as specified in Sect. 2.3, quantiles are allowed to vary only in the inner model. Overall, the communality of blocks is satisfactory. From the average communality of each block (last row in each block—values in bold), each construct explains much of the variability of its own MVs. Considering the individual communality of each MV, we did not find much differences, coherently with the way data were generated (i.e., all loadings are equal). The global communalities are satisfactory showing a good fit of the outer model.

Finally, *Redundancy* values are reported in Table 11, PLS-PM on the first column and QC-PM on the subsequent columns. Results reveal a low ability of predictor constructs to explain the variability of the outcome MVs for low quantiles, while redundancies achieved almost moderate levels for high quantiles and for PLS-PM (Latan and Ramli 2013; Latan and Ghozali 2015).

Table 10 Communalities for PLS-PM and QC-PM ($\theta = 0.50$)

Construct	MV	PLSPM	QCPM at $\theta = 0.5$
Blood pressure	SBP	0.907	0.575
	DBP	0.821	0.683
		0.864	0.629
Lipid	TC	0.780	0.490
	HDL	0.809	0.584
	TG	0.745	0.534
		0.778	0.536
CKD	ACR	0.903	0.713
	eGFR	0.892	0.692
		0.897	0.703
Global		0.846	0.624

Table 11 Redundancy measures for PLS-PM and QC-PM applied on a selected set of quantiles ($\theta \in \{0.25, 0.50, 0.75\}$)

	QC-PM			
	PLS-PM	0.25	0.5	0.75
<i>ACR</i>	0.193	0.036	0.088	0.132
<i>eGFR</i>	0.191	0.035	0.85	0.128
<i>CKD</i>	0.192	0.036	0.086	0.130

6 Conclusions and insights for future works

The original proposal of QC-PM was presented for the first time at the 8th International Symposium on PLS and Related Methods (PLS'14) which took place in 2014 in Paris (www.pls14.org). The method was proposed with the aim to extend classical least squares methods for conditional mean to the estimation of conditional quantile functions in the context of composite-based path modeling. QC-PM complements the well-known and consolidated PLS-PM by exploring heterogeneous effects of explanatory constructs over the entire conditional distributions of the response constructs.

The present paper has formalized QC-PM and the iterative procedure for parameter estimation, starting from the simplest case of one block of MVs and moving until the general path model for multi-block data. In addition, a methodological variation in the estimation phase of the outer model is also proposed. The applicative potentialities of QC-PM, along with guidelines for the interpretation of results, were provided through the analysis of an artificial data set on CKD in diabetic patients. The example highlights how QC-PM can complement traditional methods for composite-based path modeling in presence of heterogeneity in the relationships among variables. The properties of the method across different scenarios were investigated through a simulation study. The simulation design took into account the factors that typically affect the results of composite-based path modeling methods: sample size, strength of the relationship within the blocks (homogeneous vs heterogeneous blocks), different levels of correlations between constructs and shape of distributions in the outcome blocks. Data were generated from composite-based populations. The comparison among the different scenarios was carried out in terms of *RBias* and *RMSE* of estimates obtained from 500 replications for each scenario. Several similarities between QC-PM and PLS-PM emerged comparing the performance of the two methods in all generated scenarios. Nevertheless, some differences were identified. However, it is worth to recall that the spirit of the simulation study is to show the properties of QC-PM rather than to provide a comparison with PLS-PM. In fact, QC-PM and PLS-PM are not alternative but complementary methods.

Simulations point out similar results for QC-PM and PLS-PM, both in terms of bias and *RMSE*. This confirms our insight to consider QC-PM as a supplementary method to PLS-PM, with similar features but able to assess relationships between variables in different parts of the distribution. However, it is noted that variability of the QC-PM estimates is always greater even though the bias is smaller (the true population parameter is always within the 90% range of the central values for large samples). Even if the convergence of estimates is confirmed as the sample size increases in the case of

heterogeneous blocks, both $RBias$ and $RMSE$ are always higher than the homogeneous case. The new element that emerges in the case of an asymmetric distribution is the ability of QC-PM to capture the positive skewness of the distribution used to generate the data. The variability of the path coefficient estimates is smaller for $\theta = 0.25$ and larger in the right tail ($\theta = 0.75$) and the parameter is overestimated at the lowest quantile and underestimated at the highest quantile.

From a methodological point of view, a promising extension of QC-PM will accommodate the case of observed or unobserved heterogeneity among observations. In the PLS-PM literature several contributions allow to treat both kind of heterogeneity (Sarstedt et al. 2016, 2011b; Lamberti et al. 2016; Sarstedt et al. 2011a; Esposito Vinzi et al. 2008). In the QR literature, Davino and Vistocco (2018) proposed an innovative approach to identify group effects through a quantile regression model. Future studies will be devoted to combine these approaches into QC-PM. Moreover, since a recent work by Davino et al. (2020) exploited the ability of QC-PM for in-sample prediction, future research will further evaluate the proposed approach from an out-of-sample prediction perspective.

A further development, albeit a minor one, will consider the implementation of another way of calculating outer weights based on a measure of quantile correlation. Several contributions in the literature extend the first proposal of quantile correlation (Li et al. 2014) introducing different alternatives to measure the linear correlation between any two random variables for a given quantile (Tang et al. 2021; Xu et al. 2020). The introduction of a descriptive measure such as quantile correlation into the process of calculating outer weights would have an interesting computational advantage over traditional modes requiring the estimation of regression models.

Funding Open access funding provided by Università degli Studi di Napoli Federico II within the CRUI-CARE Agreement.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

Appendix

Quantile Regression was proposed by Koenker and Basset (1978) as a complementary and robust approach to classical regression analysis. In their seminal paper, the authors remind that “*in statistical parlance the term robustness has come to connote a certain resilience of statistical procedures to deviations from the assumptions of hypothetical models*” and that the need for robust statistics alternative to least squares estimation dates back to the nineteenth century.

Just one year earlier, Mosteller and Tukey (1977) sensed the need to identify more robust regression methods by stating that “*What the regression curve does is give a*

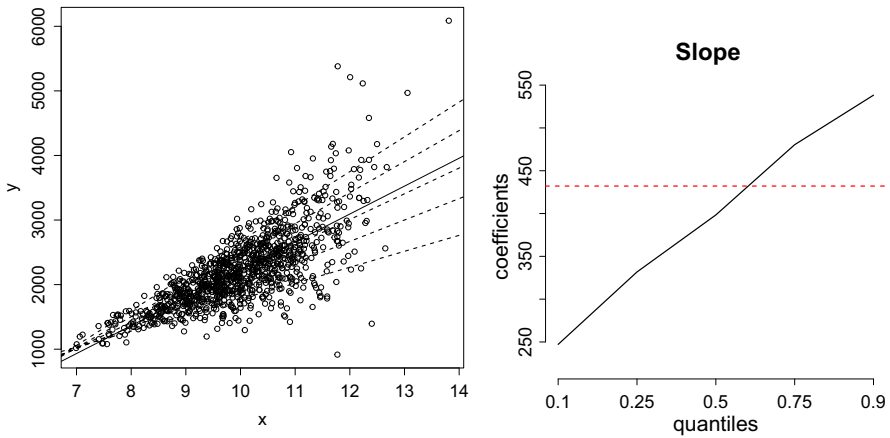


Fig. 13 Left-hand side: Scatterplot with simulated data, ordinary least square regression line (solid line) and quantile lines (dashed lines). Right-hand side: representation of QR slopes, dashed horizontal line corresponding to OLS slope

grand summary for the averages of the distributions corresponding to the set of X's. We could go further and compute several different regression curves corresponding to the various percentage points of the distributions and thus get a more complete picture of the set. Ordinarily this is not done, and so regression often gives a rather incomplete picture. Just as the mean gives an incomplete picture of a single distribution, so the regression curve gives a correspondingly incomplete picture for a set of distributions”.

Quantile Regression (QR) was introduced precisely for the purpose of going beyond the study of average effects in a regression model and to provide a description of the whole conditional distribution of a response variable in terms of a set of regressors. QR can be exploited in case of location, scale and shape shifts on the dependent variable but also when a monotone transformation of the response and/or the explanatory variables is advisable.

In order to show QR in action, we introduce an example dealing with a sample of $n = 10,000$ observations generated from the following model:

$$\begin{aligned}
 \mathbf{y} &= 1 + 2\mathbf{x} + (1 + \mathbf{x})\mathbf{e} \\
 \mathbf{x} &\sim N(10; 1) \\
 \mathbf{e} &\sim N\left((-1 + 20x); e^{x/3}\right).
 \end{aligned}$$

The scatterplot in Fig. 13 (left-hand side) shows the typical fan-like shape, sign that the amount of variability over the expected value of the dependent variable for a given value of the explanatory variable \mathbf{x} is not the same at every level of \mathbf{x} , but varies systematically with the level of \mathbf{x} .

Estimating a least-squares regression model (solid line in Fig. 13) would not fully capture the relationship between the two variables. In such a case, it is more interesting to explore the effect of the regressor at different parts of the distribution of the dependent variable.

A QR model for a given conditional quantile and for the i th observation can be formulated as follows:

$$y_i = \mathbf{x}_i \beta(\theta) + \epsilon_i(\theta)$$

$$Q_\theta(\hat{y}|\mathbf{X}) = \mathbf{X}\hat{\beta}(\theta)$$

where \mathbf{X} is the regressor matrix and \mathbf{x}_i a row of this matrix, \mathbf{y} the vector containing the dependent variable, $0 < \theta < 1$ is a generic quantile, $Q_\theta(\cdot|\cdot)$ denotes the conditional quantile function for the θ th quantile and ϵ is the error term such that $Q_\theta(\epsilon(\theta)|\mathbf{X}) = 0$.

QR offers a complete view of a response variable providing a method for modelling the rates of changes at multiple points (conditional quantiles) of its distribution (Koenker 2005; Davino et al. 2013) without requiring assumptions on the errors. Although different functional forms can be used, we will deal here only with simple linear regression models.

For each quantile, a regression line is estimated and, as a consequence, the estimated values of the response variable conditioned to given values of the regressors, provides the conditional quantiles of the dependent variable.

Going back to the example with simulated data, five QR lines have been estimated considering the following set of θ values: 0.1, 0.25, 0.5, 0.75, 0.9. The dashed lines in Fig. 13 (left-hand side) are obtained by estimating the effect of \mathbf{x} on the selected conditional quantiles of \mathbf{y} . They confirm the positive impact of the regressor but showing that this contribution increases as we move on the distribution of the \mathbf{y} variable.

In Fig. 13 (right-hand side), the slopes of the estimated quantile lines are graphically represented. The horizontal axis displays the different quantiles while the effect of the regressor is represented on the vertical axis. The dashed line parallel to the horizontal axis corresponds to the ordinary least squares coefficient. This graphical representation allows to visually catch the different effect of the regressor on the \mathbf{y} variable.

The parameter estimates in QR linear models have the same interpretation as those of any other linear model. The intercept measures the value of the dependent variable setting to zero all the regressors. Each slope coefficient $\hat{\beta}_i(\theta) = \frac{\partial Q_\theta(\mathbf{y}|\mathbf{X})}{\partial x_i}$ can be interpreted as the rate of change of the θ th conditional quantile of the dependent variable per unit change in the value of the i th regressor, holding constant the other regressors.

The conditional quantile estimator is obtained as a generalisation of the unconditional quantile estimator:

$$\hat{\beta}(\theta) = \underset{\beta}{\operatorname{argmin}} E [\rho_\theta(\mathbf{y} - \mathbf{X}\beta)] \tag{21}$$

where $\rho_\theta(\cdot)$ is an asymmetric absolute loss function which uses an unbalanced weighting system (weight equal to $(1-\theta)$ for the sum of negative deviations and weight equal to θ for the sum of positive deviations).

The most widespread algorithm for finding QR estimates is the one proposed in Koenker and d’Orey (1987) as a variant of the Barrodale and Roberts (1974) simplex algorithm. Although it is theoretically possible to extract infinite quantiles, a finite number is numerically distinct in practice. This is known as quantile process. A fairly

accurate approximation of the whole quantile process can be obtained using a dense grid of equally spaced quantiles in the unit interval (0, 1) (Davino et al. 2013).

References

- Amato S, Esposito Vinzi V, Tenenhaus M (2004) A global goodness-of-fit index for PLS structural equation modeling. Oral Communication to PLS Club, HEC School of Management, France, March, p 24
- Bakris GL, Ritz E (2009) The message for World Kidney Day 2009: hypertension and kidney disease: a marriage that should be prevented. *J Clin Hypertens* 11(3):144–147
- Barrodale I, Roberts FDK (1974) Solution of an overdetermined system of equations in the l_1 norm. *Commun Assoc Comput Mach* 17:319–320
- Benitez J, Henseler J, Castillo A, Schuberth F (2020) How to perform and report an impactful analysis using partial least squares: guidelines for confirmatory and explanatory IS research. *Inf Manag* 2(57):103168
- Bentler PM (2006) EQS 6 structural equations program manual. Multivariate Software, Encino, CA
- Chin WW (1998) The partial least squares approach to structural equation modeling. In: Marcoulides GA (ed) *Modern methods for business research*. Erlbaum, Mahwah, pp 295–358
- Davino C, Esposito Vinzi V (2016) Quantile composite-based path modelling. *Adv Data Anal Classif* 10(4):491–520
- Davino C, Vistocco D (2018) Handling heterogeneity among units in quantile regression. Investigating the impact of students' features on University outcome. *Stat Interface* 11:541–556
- Davino C, Furno M, Vistocco D (2013) Quantile regression: theory and applications. Wiley, Hoboken
- Davino C, Esposito Vinzi V, Dolce P (2016) Assessment and validation in quantile composite-based path modeling. In: Abdi H, Esposito Vinzi V, Russolillo G, Saporta G, Trincherà L (eds) *The Multiple facets of partial least squares methods*, chapter 13. Springer proceedings in mathematics and statistics. Springer, Berlin
- Davino C, Dolce P, Taralli S (2017) Quantile composite-based model: a recent advance in PLS-PM. A preliminary approach to handle heterogeneity in the measurement of equitable and sustainable well-being. In: Latan H, Noonan R (eds) *Partial least squares path modeling: basic concepts. Methodological issues and applications*. Springer, Cham, pp 81–108
- Davino C, Dolce P, Taralli S, Esposito Vinzi V (2018) A quantile composite-indicator approach for the measurement of equitable and sustainable well-being: a case study of the Italian provinces. *Social Indicators Research*, 136, pp 999–1029, Dordrecht, Kluwer Academic Publishers
- Davino C, Dolce P, Taralli S, Vistocco D (2020) Composite-based path modeling for conditional quantiles prediction. An application to assess health differences at local level in a well-being perspective. *Soc Indic Res*. <https://doi.org/10.1007/s11205-020-02425-5>
- Dijkstra T, Henseler J (2015) Consistent partial least squares path modeling. *MIS Q* 39(2):297–316
- Dolce P, Lauro CN (2015) Comparing maximum likelihood and PLS estimates for structural equation modeling with formative blocks. *Qual Quant* 49(3):891–902
- Efron B (1982) *The jackknife, the bootstrap, and other resampling plans*. SIAM, Philadelphia, p 38
- Efron B, Tibshirani RJ (1993) *An introduction to the bootstrap*. Chapman Hall, New York
- Esposito Vinzi V, Russolillo G (2013) *Partial least squares algorithms and methods*. Wiley Interdiscip Rev Comput Stat 5(1):1–19
- Esposito Vinzi V, Trincherà L, Squillacioti S, Tenenhaus M (2008) REBUS-PLS: a response-based procedure for detecting unit segments in PLS path modelling. *Appl Stoch Models Bus Ind* 24:439–458
- Esposito Vinzi V, Chin WW, Henseler J, Wang H (eds) (2010) *Handbook of partial least squares*. Springer, Berlin
- Fleishman AI (1978) A method for simulating non-normal distributions. *Psychometrika* 73:521–532
- Fornell C, Larcker DF (1981) Structural equation models with unobservable variables and measurement error: algebra and statistics. *J Mark Res* 18(3):328–388
- Furno M, Vistocco D (2018) *Quantile regression: estimation and simulation*. Wiley series in probability and statistics. Wiley, Hoboken
- Gould W (1997) sg70: interquantile and simultaneous-quantile regression. *Stata Tech Bull* 38:142
- Hair JF, Ringle CM, Sarstedt M (2011) PLS-SEM: indeed a silver bullet. *J Mark Theory Pract* 19(2):139–150
- Hair JF, Hult GTM, Ringle CM, Sarstedt M (2017) *A primer on partial least squares structural equation modeling (PLS-SEM)*, 2nd edn. Sage, Thousand Oaks

- Hair JF, Risher JJ, Sarstedt M, Ringle CM (2019) When to use and how to report the results of PLS-SEM. *Eur Bus Rev* 31(1):2–24
- Hair JF, Howard MC, Nitzl C (2020) Assessing measurement model quality in PLS-SEM using confirmatory composite analysis. *J Bus Res* 109:101–110
- He XM, Zhu LX (2003) A lack-of-fit test for quantile regression. *J Am Stat Assoc* 98:1013–1022
- Henseler J, Ringle CM, Sarstedt M (2016) Testing measurement invariance of composites using partial least squares. *Int Mark Rev* 33(3):405–431
- Jöreskog KG (1978) Structural analysis of covariance and correlation matrices. *Psychometrika* 43(4):443–477
- Koenker R (2005) *Quantile regression*. Cambridge University Press, Cambridge
- Koenker R, Basset G (1978) Regression quantiles. *Econometrica* 46:33–50
- Koenker R, Basset G (1982) Robust tests for heteroscedasticity based on regression quantiles. *Econometrica* 50(1):43–61
- Koenker R, d'Orey V (1987) Computing regression quantiles. *Appl Stat* 36:383–393
- Koenker R, Machado JAF (1999) Goodness of fit and related inference processes for quantile regression. *J Am Stat Assoc* 94(448):1296–1310
- Kováč Š, Želinský T (2013) Determinants of the Slovak enterprises profitability: quantile regression approach. *Statistika* 93(3):41–55
- Lamberti G, Aluja TB, Sanchez G (2016) The pathmox approach for PLS path modeling segmentation. *Appl Stoch Models Bus Ind* 32(4):453–468
- Latan H, Ghozali I (2015) *Partial least squares: concepts, techniques and application using program Smart-PLS 3.0*, 2nd edn. Diponegoro University Press, Semarang
- Latan H, Ramli NA (2013) The results of partial least squares-structural equation modeling analyses (PLS-SEM). Available at SSRN. <https://doi.org/10.2139/ssrn.2364191>
- Li G, Li Y, Tsai C (2014) Quantile correlations and quantile autoregressive modeling. *J Am Stat Assoc* 110(509):233–245
- Lohmöller JB (1989) *Latent variable path modeling with partial least squares*. Physica-Verlag, Heidelberg
- Mosteller F, Tukey J (1977) *Data analysis and regression*. Addison-Wesley, Reading, MA
- OECD (2008) *Handbook on constructing composite indicators: methodology and user guide*. OECD, Paris
- Parzen MI, Wei L, Ying Z (1994) A resampling method based on pivotal estimating functions. *Biometrika* 18:341–350
- Paxton P, Curran PJ, Bollen KA, Kirby J, Chen F (2001) Monte Carlo experiments: design and implementation. *Struct Equ Model Multidiscip J* 8(2):287–312
- R Core Team (2020) R: a language and environment for statistical computing. R Foundation for Statistical Computing. Austria (<https://www.R-project.org/>), Vienna
- Rigdon EE (2013) Partial least squares path modeling. In: Hancock G, Mueller R (eds) *Structural equation modeling: a second course*, 2nd edn. Information Age, Charlotte, pp 81–116
- Rudis B (2019) *ggeonodist: create diminutive distribution charts*. R package version
- Sarstedt M, Henseler J, Ringle CM (2011a) Multi-group analysis in partial least squares (PLS) path modeling: alternative methods and empirical results. In: Sarstedt M, Schwaiger M, Taylor CR (eds) *Advances in international marketing*, vol 22. Bingley, Emerald, pp 195–218
- Sarstedt M, Becker J-M, Ringle CM, Schwaiger M (2011b) Uncovering and treating unobserved heterogeneity with FIMIX-PLS: which model selection criterion provides an appropriate number of segments? *Schmalenbach Bus Rev* 63:34–62
- Sarstedt M, Ringle CM, Gudergan SP (2016) Guidelines for treating unobserved heterogeneity in tourism research: a comment on Marques and Reis (2015). *Ann Tourism Res* 57:279–284
- Schlittgen R (2019) R package *sempls: simulation, estimation and segmentation of composite based structural equation models* (version 1.0.0). <https://cran.r-project.org/web/packages/cbsem/index.html>
- Schlittgen R, Sarstedt M, Ringle CM (2020) Data generation for composite-based structural equation modeling methods. *Adv Data Anal Classif*. <https://doi.org/10.1007/s11634-020-00396-6>
- Shmueli G, Sarstedt M, Hair JF, Cheah J-H, Ting H, Vaithilingam S, Ringle CM (2019) Predictive model assessment in PLS-SEM: guidelines for using PLSpredict. *Eur J Mark* (forthcoming)
- Stone M (1974) Cross-validatory choice and assessment of statistical predictions. *J R Stat Soc* 36:111–147
- Tang W, Xie J, Lin Y, Tang N (2021) Quantile correlation-based variable selection. *J Bus Econ Stat*. <https://doi.org/10.1080/07350015.2021.1899932>
- Tenenhaus M (2008) Component-based structural equation modelling. *Total Qual Manag Bus Excell* 19:871–886

- Tenenhaus M, Vinzi VE, Chatelin YM, Lauro C (2005) PLS path modeling. In: Computational statistics and data analysis, pp 159–205
- Vale C, Maurelli V (1983) Simulating multivariate nonnormal distributions. *Psychometrika* 48(3):465–471
- Wang Y, Feng XN, Song XY (2016) Bayesian quantile structural equation models. *Struct Equ Model Multidiscip J* 23(2):246–258
- Wold H (1966a) Nonlinear estimation by iterative least squares procedures. *Research Papers. Statistics* 630
- Wold H (1966b) Estimation of principal component and related models by iterative least squares. In: Krishnaiah PR (ed) *Multivariate analysis*. Academic Press, New York, pp 391–420
- Wold H (1975) From hard to soft modelling. In Wold H (ed) *Modelling in complex situations with soft information*. (Paper, Third World Congress on Econometrics; Toronto, Canada; 1975 August 21–26). (Research Report 1975:5). University, Institute of Statistics, Goteborg, Sweden
- Wold H (1982) Soft modeling: the basic design and some extensions. In: Jöreskog K, Wold H (eds) *Systems under indirect observation*, vol 2. North-Holland, Amsterdam, pp 1–54
- Wold H (1985) Partial least squares. In: Kotz S, Johnson NL (eds) *Encyclopedia of statistical sciences*, vol 6. Wiley, New York, pp 581–591
- Xu C, Ke J, Zhao X, Zhao X (2020) Multiscale quantile correlation coefficient: measuring tail dependence of financial time series. *Sustainability* 12(12):4908

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.