Quantile Composite-based Path Modeling. An application to a study of chronic kidney disease

Pasquale Dolce[1], Cristina Davino[2], Domenico Vistocco[3]

[1]Dipartimento di Sanità Pubblica, Università degli Studi di Napoli Federico II
[2]Dipartimento di Scienze Economiche e Statistiche, Università degli Studi di Napoli Federico II
[3]Dipartimento di Scienze Politiche, Università degli Studi di Napoli Federico II, Napoli

INTRODUCTION

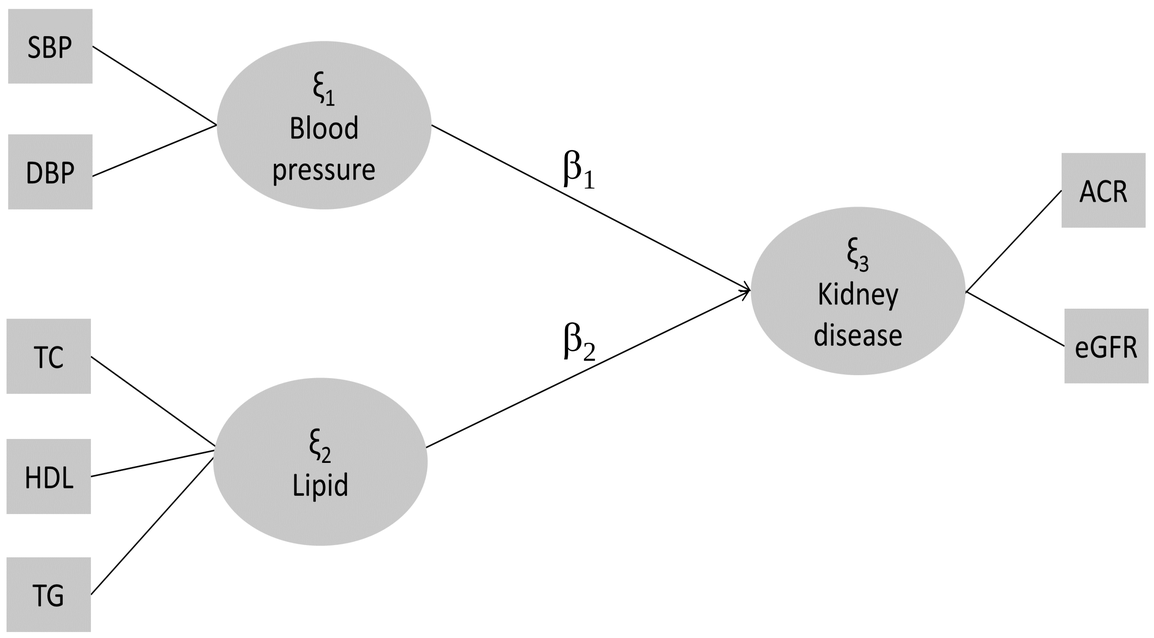
Composite–based path modeling aims to study relationships among a set of composites, i.e., linear combinations of observed or manifest variables (MVs). Prior knowledge is used to establish which MV is related to which composite and to specify the relations among composites. Partial Least Square Path Modeling (PLS–PM) [1] is the traditional method for composite–based path modeling. Being based on iterative alternating Ordinary Least Squares (OLS) algorithms, it focuses on the conditional means of the distributions. However, limiting to the mean may not reveal interesting effects at other locations of the outcome variable distributions, especially when response variables are highly skewed, the analysis is concerned also about the tail part of the distributions, heteroscedastic variance of the errors is present, distributions are characterized by outliers. In such cases, a quantile regression [2] approach to path modeling, which consider the entire distribution of outcome variables, is a valuable tool to complement the PLS-PM.

OBJECTIVE

To present an application of a recent proposal, called Quantile Composite-based Path Modeling (QC-PM) [3], on Chronic Kidney Disease (CKD) in diabetic patients, illustrating the potentialities of the method to detect heterogeneity in the variable relationships, and the benefit of using QC-PM as a supplement of PLS-PM.

MATERIALS AND METHODS

Artificial data were generated mimicking a study proposed by Wang et al. [4] who examined the potential risk factors of CKD through a quantile approach to factor-based structural equation modeling. Data were simulated, considering the same model and the obtained estimates, because the original data were not available. The study investigates Type 2 diabetic patients who have experienced CKD. 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 measuring the composite Kidney disease (ξ_3). The considered risk factors were Blood pressure (ξ_1) and Lipid (ξ_2) . 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). The theoretical model is illustrated in the following figure.

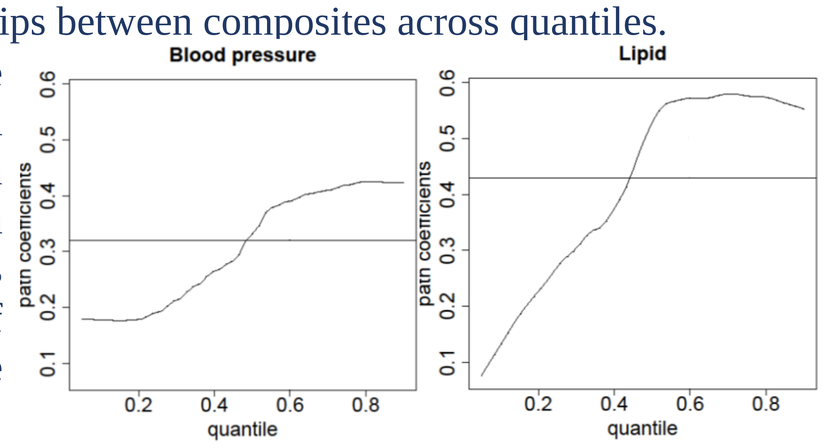


MATERIALS AND METHODS (Cont'd)

In the study by Wang et al. [4] Blood pressure and Lipid were positively correlated with the severity of Kidney disease, and these relations were stronger for higher quantiles. Thus, this heterogeneity was introduced in the artificial data assuming that ξ_1 and ξ_2 exert a different effect on the different parts of the ξ_3 distribution.

RESULTS

QC–PM was able to detect the structure underlying the data, since it properly distinguished the different effects in the different parts of the kidney disease distribution: both path coefficients increase with quantiles (according to the data generation process). The following figure represents the estimates of the path coefficients (β), which measure the relationships between composites across quantiles.



Both Blood pressure and Lipid have a positive impact on Kidney disease, which increases for patients with higher levels of CKD. As shown in the following table, the effect of Blood pressure on Kidney disease (β_1) estimated by PLS-PM was equal to 0.32 ($p<0.05$), while using QC-PM and considering the quantiles $\Theta \in \{0.25, 0.50, 0.75\}$, the estimate of β_1 was equal to 0.19 ($p<0.05$) for the Θ=0.25, 0.29 ($p<0.05$) for the Θ=0.50 and 0.42 ($p<0.05$) for the Θ=0.75. The PLS-PM estimated effect of Lipid on Kidney disease (β_2) was equal to 0.43 ($p<0.05$), while the QC-PM estimates of β_2 was equal to 0.26 ($p<0.05$) for the Θ=0.25, 0.55 ($p<0.05$) for the Θ=0.50 and 0.59 ($p<0.05$) for the Θ=0.75.

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

DISCUSSION

QC–PM complements the well-known and consolidated PLS–PM by exploring heterogeneous effects of explanatory composites over the entire conditional distributions of the response composites.

A description of QC–PM and the applicative potentialities of QC–PM was provided through the analysis of an artificial data set on chronic kidney disease in diabetic patients, highlighting the capability of the method in detecting the heterogeneity in the variable relationships.

BIBLIOGRAPHY

[1] Koenker, R. (2005). Quantile regression. Cambridge University Press, Cambridge, 2005.

[2] Tenenhaus, M., Vinzi, V. E., Chatelin, Y. M and Lauro C. (2005). PLS path modeling. *Comput. Stat. Data Anal*., 48(1): 159–205.

[3] Davino, C., Dolce, P., Taralli, S. and Vistocco M. (2020, in press). 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.*

[4] Wang, Y., Feng X. N. and Song X. Y. (2016). Bayesian Quantile Structural Equation Models, *Struct. Equ. Modeling*, 23(2): 246-258.