3rd International Conference on Advanced Research Methods and Analytics (CARMA2020)
Universitat Politècnica de València, València, 2020
DOI: http://dx.doi.org/10.4995/CARMA2020.2020.11599

High order PLS path modeling to evaluate well-being merging traditional and big data: A longitudinal study

Francesca De Battisti, Elena Siletti
Department of Economics, Management and Quantitative Methods, Universita' degli Studi di Milano, Italy.

Abstract

We propose using high order partial least squares path modeling (PLS-PM) to define a synthetic Italian well-being index merging traditional data, represented by the Quality of Life index proposed by "Il Sole 24 Ore", and information provided by big data, represented by a Subjective Well-being Index (SWBI) performed extracting moods by Twitter. High order constructs allow to define a more abstract higher-level dimension and its more concrete lower-order sub-dimensions. These layered constructs have gained wide attention in applications of PLS-PM; many contributions in literature proposed their use to build composite indicators. The aim of the paper is to underline some critical issues in the use of these models and to suggest the implementation of a new adapted repeated indicator approach. Furthermore, following some recommendations proposed on the use of PLS-PM in longitudinal studies, we compare the situation in 2016 and 2017.

Keywords: *Well-being; big data;PLS-PM, SEM, hierarchical models.*

1. Introduction

Several contributions deal with the use of PLS-PM to assess a hierarchical construct model (Tenenhaus et al. 2005). Briefly, in Wold's original design it was expected that each construct would be necessarily connected to a set of observed variables (Wold 1982); on this basis, Lohmöller (1989) proposed the so-called hierarchical component model; recently, Wetzels et al. (2009) provided guidelines outlining four key steps to define a hierarchical construct model, while Becker et al. (2012) focused on the second-order hierarchical latent variable models, which are usually treated with reflective relationships, paying attention to formative relationships; finally, Sarstedt et al. (2019) deepened how to evaluate the results of higher-order constructs in PLS-PM using the repeated indicator and the two-stage approaches.

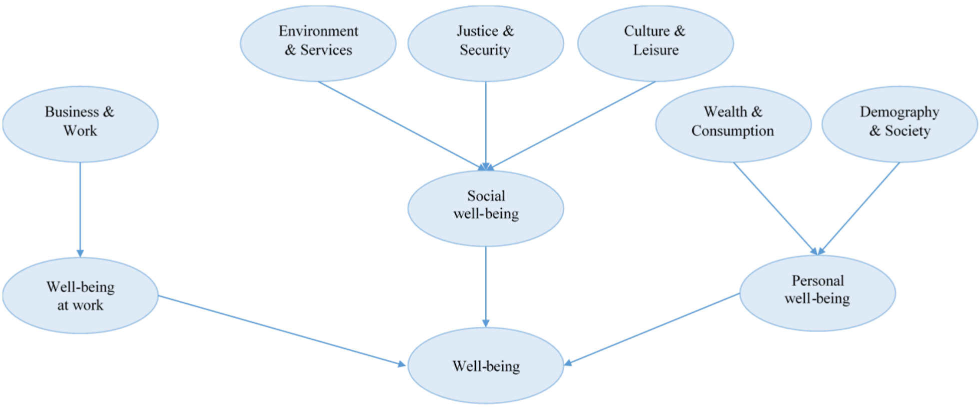

Figure 1. Structural model.

Since 1990, "Il Sole 24 Ore" has provided a yearly quality of life index (QoL) for the Italian provinces (NUTS- 3 in the European nomenclature of territorial units for statistics). Unfortunately, this index is strictly objective, observing only material aspects of quality of life. Following the Stiglitz et al. (2009) recommendation, "current well-being has to do with both economic resources, such as income, and with non-economic aspects of peoples' life (what they do and what they can do, how they feel, and the natural environment they live in)", we suggest to consider an overall index summarizing both the objective and subjective contents, integrating the QoL index with a perceived and subjective source coming from social networks big data: the SWBI index. This last is a multidimensional well-being measure auditing the social networks moods, proposed by Iacus et al. (2015). It is obtained exploiting the big amount of data offered by Twitter data and adopting a new human supervised technique of sentiment analysis (Ceron et al. 2016).

Our proposal consists in mashing-up the different data using the higher order PLS-PM, that has been often applied to buid composite indices using traditional data (Cataldo et al. 2017, Lauro et al. 2018, Davino et al. 2017).

Table 1. : Indicators and themes of "Il Sole 24 Ore" for 2017 and 2016.

Wealth and consumptions	**Demography and society**	**Business and work**
Bank deposits per capita	Resident graduates	Registered enterprises (per 100 inhabitant)
Average monthly rent	Birth rate	Employment rate
Durable goods mean spending for family	Ageing index	Rate of youth unemployment (15-29)
Protests per capita	Internal migratory balance	Loans on deposit (%)
Monthly retirement benefits	Inhabitants for square Km	Exports in % of GDP
Real estate assets per capita (only in 2016)	Acquisition of Italian citizenship	Innovative start-ups per 1000 enterprises
Added value (per capita) (only in 2016)	Number of marriage separations (only in 2016)	Patent application (only in 2016)
GDP per capita (only in 2017)	Average number of education years (only in 2017)	Gender salary gap (%) (only in 2017)
Online shopping (only in 2017)		
Environment and services	**Justice and security**	**Culture and leisure**
Index on urban ecosystem	Home theft	Libraries
Social expenditure of local authorities per capita	Muggings and pick pocketing	Sportiness index
Broadband	Ultra-triennial pending lawsuits	Number of restaurants and pubs
Hospital emigration among regions	Robberies	Foreign traveler expenditure
Number of bank branches, ATM and POS	Scams and computer frauds	Non-profit association
Availability of municipal nursery schools (only in 2016)	Car thefts	Entertainment tickets (only in 2016)
Index of climate changes in temperature (only in 2016)	Index of cause disposal (only in 2016)	Number of cinemas (only in 2016)
Land consumption (only in 2017)	Contentiousness index (only in 2017)	Seats in cinemas (only in 2017)
Spending on drugs (only in 2017)		Number of entertainments (only in 2017)

2. Method

Our data refers to those employed by QoL and SWBI in 2017 and 2016 years. The structural model has been reported in Figure 1.

The definition of the measurement model involves separately three levels. As displayed in Table 1, at the lower order we consider the six themes proposed by “Il Sole 24 Ore”, each related to seven indicators.

At the second order of this hieralchical model, we consider the three macro-dimensions suggested by the New Economics Foundation (2012):

- *Personal well-being*
- *Social well-being*
- *Well-being at work*

In defining this measurement level, some issues should be taken into account. First of all, in models with more than two levels the use of the repeated indicator approach yields collinearity problems among constructs, this aspect especially occurs if we need to define them as formative. Furthermore, since the standard structure of QoL, defined in the “Il Sole 24 Ore” project, does not consider a second level, we introduce at this level a new adapted repeated indicator approach, using the related first order indicators and the SWBI components (Table 2). It is worth to notice that, in QoL, some indicators have been changed from one year to the next (Table 1). This is not a good practice in structural equation models, because the invariance of the construct measure has to be ensured. However, our purpose is not to criticize the procedure adopted in the definition of the constructs, but rather to propose a new method to aggregate the constructs and moreover to consider also subjective aspects.

Table 2. : Indicators for macro dimensions.

Personal well-being	**Social well-being**	**Well-being at work**
7 from Wealth and consumptions 7 from Demography and society	7 from Environment and services 7 from Justice and security 7 from Culture and leisure	7 from Business and work
Emotional well-being (SWBI) Satisfying life (SWBI) Vitality (SWBI) Resilience and self-esteem (SWBI) Positive functioning (SWBI)	Trust and belonging (SWBI) Relationships (SWBI)	Quality of job (SWBI)

For the last order, we measure the overall well-being index using all the fifty indicators, applying a traditional repeated indicator approach. All the constructs are defined by composite measures.

Following Davino et al. (2017), to estimate the outer weights of the model we use Mode A for all the higher-order constructs and Mode B for the first-order constructs. In order to estimate inner weights, we use the path weighting scheme.

A drawback in the use of the repeated indicator approach is that the variables with a higher number of corresponding indicators have a major impact on the correspondent higher-order construct. Following this consideration, we foresee that QoL data will impact more on the overall index than data from SWBI. This is not a worrying issue, since the authoritativeness of the QoL is not affected.

2. Results

All the analysis has been carried on using the SEMinR library (Ray et al. 2019).

The PLS-PM path coefficients for the model in 2017 and 2016 are reported in Table 3. The behavior for 2017 and 2016 seams similar. *Personal well-being*, *Social well-being* and *Well-being at work* have a significant effect on the final synthetic well-being index (WB). The *Personal well-being* has the highest impact on WB, nevertheless the number of its indicators is lower than for *Social well-being*.

Analyzing the single macro-dimensions, we notice that on *Personal well-being* the *Wealth and consumptions* has an effect more than twice with respect to *Demography and society* in 2017 and fourfold in 2016. This result is very interesting, considering that in the construction of the “Il Sole 24 Ore” overall index each theme has the same weight. For the *Social well-being*, *Justice and security* has not a significant path in 2016.

Table 3. : Path coefficients for 2016 e 2017 (p-values < 0.001, except for *) and p-value of t-test for equality of paths.

Composite		Path coefficients 2017	Path coefficients 2016	*p*-value
WB	Personal well-being	0.452	0.435	0.245
	Social well-being	0.394	0.402	0.376
	Well-being at work	0.196	0.197	0.464
Personal well-being	Wealth and consumptions	0.720	0.812	0.017
	Demography and society	0.302	0.204	0.016
Social well-being	Culture and leisure	0.338	0.346	0.435
	Environment and services	0.320	0.416	0.039
	Justice and security	0.423	0.307*	
Well-being at work	Business and work	0.999	0.996	0.163

As a measure of goodness of fit of the model, we consider the redundancy (Lohmöller 1989) analyzing the convergent validity, because in hierarchical composite models the R^2 is very closed or equal to 1, as higher-order constructs are almost fully explained by their lower-order constructs. The redundancy index of WB is equal to 0.266 (2017) and 0.292 (2016). The redundancy indices for the sub dimensions are: for *Personal well-being* 0.323 (2017) and 0.337 (2016), for *Social well-being* 0.236 (2017) and 0.262 (2016), and for *Well-being at work* 0.340 (2017) and 0.380 (2016). The average redundancy for the overall model is equal to 0.276 (2017) and 0.300 (2016). These values are not high, but the model is complex and the redundancy values are generally small in PLS-PM; for these reasons, they are judged satisfactory. Furthermore, to validate this model we have to check the collinearity between indicators. The analysis performed at the first order highlights that the VIF values are acceptable, excluding the collinearity issue.

To deep the analysis, we propose to compare the results of 2016 and 2017, following the procedure suggested by Roemer (2016) on the use of PLS-PM in longitudinal studies. Especially , we refer to model type A.1, since our main research object is to investigate the evolution of effects over time and our panel data. After having estimated the model in two different years, we carry on a multigroup analysis, MGA (Henseler et al. 2009), to test the changes in the path coefficients over time; here the different "groups" are interpreted as the different points in time. The last column in Table 3 shows the p-values for the *t*-test used for MGA procedure. The path of *Wealth and consumption* and of *Demography and society* on *Personal well-being* are significantly different in 2016 and 2017. The same happens for the path of *Environment and services* on *Social well-being*

We compare the province rank from the new overall well-being index with the rank from the overall QoL index. The rank correlation indices are: for 2016 *Spearman* = 0.874 and *Kendall* = 0.743, for 2017 *Spearman* = 0.852 and *Kendall* = 0.705. Considering the subjective aspects of well-being slightly affects the classification of the Italian provinces on the basis of the well-being index. However, the different scores obtained are also due to the different methods applied in the aggregation of the dimensions (the same weight for the QoL and different relevance for PLS-PM).

Table 4. : Results of the test of significance of the changes in level of the constructs.

Constructs	Mean difference (2017-2016)	*t*-value	*p*-value
WB	-18656.38	-25.51	0.000
Personal well-being	-20462.18	-24.67	0.000
Social well-being	-759.05	-11.61	0.000
Well-being at work	53.03	57.77	0.000
Wealth and consumptions	-18167.35	-24.34	0.000
Demography and society	-156.01	-15.90	0.000
Culture and leisure	324.73	11.46	0.000
Environment and services	12.25	10.62	0.000
Justice and security	-49.58	-1.64	0.105
Business and work	67.60	58.93	0.000

Finally, being also interested in the change in the level of the constructs over time, we conduct a paired sample *t*-test, performed calculating the non-standardized scores (Table 4). All the constructs have significantly changed during the time. In particular, the WB has got worse from 2016 to 2017, as *Personal* and *Social well-being*. Instead *Well-being at work* has improved.

3. Conclusion

The aim of the paper is to propose a synthetic well-being index using high order PLS-PM merging traditional and Twitter big data. To our knowledge, this is one of the first attempts to merge objective and social network data at provincial level. The novelty of the proposal is also due to the choice of the PLS-PM, with the suggestion of adapting the repeated indicator approach. The insiders could be interested in applying a new approach to take into consideration simultaneously traditional and big data, merging them with suitable weights. The findings are interesting and highlight that considering subjective aspects has an impact on the overall evaluation. Some issues will deserve an in-depth analysis: the estimation of the outer weights in formative-formative hierarchical models and the extension of the multigroup approach to compare more than two situations.

References

Becker J.M., Klein K. & Wetzels M. (2012). Hierachical Latent Variable Models in PLS-SEM: Guidelines for Using Reflective-Formative Type Models. *Long Range Planning*, 45, 359-394.

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

Ceron A., Curini L. & Iacus S. (2016). iSA: A fast, scalable and accurate algorithm for sentiment analysis of social media content. *Information Sciences*, 105-124.

Davino C., Dolce P. & Taralli S. (2017). Quantile Composite-Based Model: A Recent Advance in PLS-PM, in Latan H. & Noonan R. (eds.) *Partial Least Squares Path Modeling. Basic concepts, methodological issues and applications*, (pp.81-108), DOI 10.1007/978-3-319-54069-3_5, Berlin: Springer.

Henseler J., Ringle C.M. & Sinkovics R.R. (2009). The use of partial least squares path modeling in international marketing. *Advances in International Marketing*, 20, 227-319.

Iacus S.M., Porro G., Salini S. & Siletti E. (2015). Social networks, happiness and health: from sentiment analysis to a multidimensional indicator of subjective well-being. ArXiv e-prints 1512.01569.

Lauro N.C., Grassia M.G. & Cataldo R. (2018). Model Based Composite Indicators: New Developments in Partial Least Squares-Path Modeling for the Building of Different Types of Composite Indicators. *Social Indicators Research*, 135, 421-455.

Lohmöller J.B. (1989). *Latent Variable Path Modeling with Partial Lest Squares*. New York: Springer,

New Economics Foundation (2012). *The Happy Planet Index: 2012 Report. A global index of sustainable well-being.*

Ray S., Danks N. & Velasquez Estrada J.M. (2019). SEMinR: Domain-Specific Language for Building PLS Structural Equation Models. R package version 0.7.0.

Roemer E. (2016). A tutorial on the use of PLS path modeling in longitudinal studies. *Industrial Management and Data Systems*, 116 (9), 1901-1921.

Sarsted M., Hair J.F., Cheah J.H., Becker J.M. & Ringle C.M. (2019). How to specify, estimate, and validate higher-order constructs in PLS-SEM. *Australian Marketing Journal*, 27, 197-211.

Stiglitz J.E., Sen A. & Fitoussi J.P. (2009). Report by the Commission on the Measurement of Economic Performance and Social Progress, Tech. Rep. INSEE.

Tenenhaus M., Vinzi V.E., Chaterlin M., & Lauro, C. (2005). PLS path modeling. *Computational Statistics and Data Analysis*, 48, 159-205.

Wetzels M., Odekerken-Schröder G. & van Oppen C. (2009). Using PLS Path Modeling for Assessing Hierachical Construct Models: Guidelines and Empirical Illustration. *MIS Quarterly*, 33 (1), 177-195.

Wold, H. (1982). Soft modeling: The basic design and some extensions. In K. Jöreskog & H. Wold (Eds). *Systems under indirect observation*, 2, (pp.1-54). Amsterdam: North-Holland.