COLLANA DEL
DIPARTIMENTO DI ECONOMIA

NON-METRIC PLS PATH MODELING: INTEGRATION INTO THE LABOUR MARKET OF SAPIENZA GRADUATES

Francesca Petrarca

ISSN 2279-6916 Working papers
(Dipartimento di Economia Università degli studi Roma Tre) (online)

Working Paper n° 186, 2013

I Working Papers del Dipartimento di Economia svolgono la funzione di divulgare tempestivamente, in forma definitiva o provvisoria, i risultati di ricerche scientifiche originali. La loro pubblicazione è soggetta all'approvazione del Comitato Scientifico.

Per ciascuna pubblicazione vengono soddisfatti gli obblighi previsti dall'art. l del D.L.L. 31.8.1945, n. 660 e successive modifiche.

Copie della presente pubblicazione possono essere richieste alla Redazione.

esemplare fuori commercio
ai sensi della legge 14 aprile 2004 n.106

REDAZIONE:
Dipartimento di Economia
Università degli Studi Roma Tre
Via Silvio D'Amico, 77 - 00145 Roma
Tel. 0039-06-57335655   fax 0039-06-57335771
E-mail: dip_eco@uniroma3.it
http://dipeco.uniroma3.it

DIPARTIMENTO DI ECONOMIA

NON-METRIC PLS PATH MODELING: INTEGRATION INTO THE LABOUR MARKET OF SAPIENZA GRADUATES

Francesca Petrarca

Non-Metric PLS Path Modeling: integration into the labour market of Sapienza graduates

Francesca Petrarca

Abstract Non-Metric Partial Least Squares Path Modeling is a recent methodology based on the concept of Optimal Scaling applied to PLS Path Modeling algorithms. We adopted Non-Metric PLS Path Modeling to analyse a large administrative dataset containing nominal and ordinal variables using a specialized R package now available. We suggest a model in order to perform a preliminary quantitative study of the job success of Sapienza University of Rome graduates in terms of quality of work.

Keywords: Optimal Scaling, PLS Path Modeling, Structural Equation Models, Categorical Variables, Administrative Archive, Economics Education Research.

JEL classifications: C100, C390, A200.

1 Introduction

This paper has the following purposes:

1. to give a short presentation of the UNI.CO archive, [2], which is the up-to-date complete administrative database of the integration into the Italian labour market of Sapienza graduates;
2. to study indicators of job success and to estimate their relationship with educational and job curricula.
3. to model job success as a latent variable in PLS-PM framework;

Francesca Petrarca
University of Roma Tre, Department of Economics, Via Silvio D'Amico, 77, 00145, Rome, Italy.
e-mail: francesca.petrarca@uniroma3.it

4. to assess the effectiveness of Non-Metric approach [11] to Partial Least Square Path Modeling (PLS-PM) in the analysis of variables observed on different measurement scales.

The general framework is the study of the subordinate and para-subordinate employment offered to the Sapienza graduates by the Italian labour market. The main problem is to extract from the information contained in the database, indicators that show the possibility to obtain a good job position after graduation.
The goal is to define and measure the possibility of getting a good job position, i.e., satisfactory, well paid, stable over time, with possibility of improvements in career, consistent with university curriculum.
This study is based on the data of the UNI.CO archive[1] which contains the joint integration of the Sapienza graduates' archive and the Italian Ministry of Labour archive (known as *Compulsory Communication* (CO)). The integration of the two archives has produced a remarkable improvement in the quality of the information contained in the dataset with contributions entirely additional in respect of those provided singularly by each of them.
Measuring job qualifications is not an easy task, both in absolute or in relative terms [6]. In literature many indicators have been studied (e. g. index of job desirability [8], job quality index [10]).
In this explorative paper we are interested in studying two new composite indicators [1] that are related to the possibility of the success in terms of Sapienza graduates best employment status. These indicators quantify the concept of job success using the definition of *optimal* and *quasi-optimal contract* based on the ISCO classification of job quality and on a minimum continuative duration of the job. The two indicators that we want to study are:

- *An optimal contract*: a contract that offers a permanent and highly qualified position (by ISCO Classification) with an actual duration of at least 8 months.
- *A quasi-Optimal contract*: a contract that offers a highly qualified position (by ISCO Classification) with an actual duration of at least 8 months.

In the International Standard Classification of Occupations (ISCO) a highly qualified position is identified with ISCO1 (managers) and ISCO2 (intellectual and scientific professions). The threshold of at least 8 months comes from D.lgs.181/2000 that considers in a status of unemployment workers with a job contract of less than 8 months.

[1] The UNI.CO archive has been generated by an experimentation started in 2012 with the aim of establishing the integration of administrative archives. The results of the preliminary analysis of this new archive can be found in a first report (to be published) based on the Compulsory Communications archive for the study of labour demand for Sapienza graduates edited by the UNI.CO workgroup under the supervision of Giorgio Alleva. This group is composed of researchers coming from Sapienza University of Rome, Italian Ministry of labour and Italia Lavoro. For the Sapienza University: Pietro Lucisano, Carlo Magni, Silvia Massimi, Francesca Petrarca, Alessandro Sanzo, Bruno Sciarretta and Eleonora Renda. For the Italian Ministry of labour: Daniele Lunetta and Maurizio Sorcioni and for Italia Lavoro Giuseppe De Blasio. The workgroup was supported by a Scientific and Technical Committee of the Sapienza University composed of: Giorgio Alleva, Tiziana Catarci, Rosalba Natale and Cristiano Violani.

The concept of a good job is rather theoretical and it needs a quantification in order to be infered from data.
The class of the PLS methods [4, 5] seems to be the most suitable methodology to tackle this kind of problems because their capability to:

- quantify the latent variables (LVs) representing unobservable constructs;
- provide an estimate of the LVs for each observation;
- estimate cases affected by multicollinearity;
- work without distributional hypotheses.

The last point is important because in the social sciences it is often the case that the distributions of the variables are asymmetric and very far from the Gaussian distribution.
Our suggestion is to measure the concept of good job defining it as a latent variable in PLS-PM framework. Our analysis is based on a dataset of variables which are observed on different measurement scales (numerical, ordinal and nominal). An interesting possibility to address this issue has been recently offered by a new procedure called Non-Metric Partial Least Squares (NM-PLS) [11] which is based on the implementation of the optimal scaling method applied to PLS algorithms. NM-PLS extends the applicability of PLS methods to data measured on different measurement scales, as well as to variables linked by non-linear relationships. A distinctive feature of these algorithms is that they provide a new metric both to non-metric and to metric variables. In this paper we adopted the term non-metric data to refer to ordinal and nominal variables.
The structure of this paper is as follows. In Sec.2 we present a brief description of PLS-PM, the main feature of the NM-PLS and the assessment procedure adopted by this method. In Sec. 3 we discuss the dataset adopted and the model suggested. In Sec.4 we discuss the results taking into account the assessment of the model and the quantification procedure. Sec. 5 draws conclusions.

2 PLS-PM

The Partial Least Square can be viewed as a set of methods for analysing multiple relationships between various blocks of variables. In particular the most common application of PLS-PM is the calculation of indices to quantify some key concepts or constructs called *latent variables (LVs)* that cannot be measured directly. One can analyse these concepts combining and summarizing a set of information that in some way reflect the meaning of the concept. The latent variables are indirectly measured by means of variables which can be observed/measured called *manifest variables (MVs)*. The manifest variables are divided in blocks which reflect to some extent the latent construct they are associated with.
The PLS methods are part of Structural Equation Models (SEM) [3, 7] that include a number of statistical methodologies meant to estimate a network of causal rela-

tionships, based on a theoretical model, linking two or more latent concepts, each measured by means of a number of observable indicators.
PLS-PM estimates the network of linear relations among the MVs and their own LVs, and among the LVs inside the model, through a system of inter-dependent equations based on simple and multiple regressions. The corresponding conceptual model can be represented by path diagrams where the LVs are represented by circles, the MVs by rectangles and the dependence relationships among the variables by arrows.
The difference between PLS-PM and SEM is that the first has been introduced as a component-based estimation procedure [14] and the second like a covariance-based LISREL-type (Linear Structural Relations) approach [9].
The PLS-PM is considered as a *soft-modeling* approach because it does not require strong assumptions with respect to the distributions, the sample size and the measurement scale. So the inferential approach is based on resampling technique that allow to obtain empirical distributions of the parameters. In the PLS-PM the outer weights, linking each MV to corresponding LV, are estimated by an iterative algorithm in which the latent variable scores are obtained through the alternation of the outer and inner estimations of the LVs. The PLS-PM consists of two sub-models:

- the structural model (or inner model) where the relationships among latent variables are established;
- the measurement model (or outer model) where the relationships between each variable and its block of manifest variables are established.

No formal proof of convergence of the general algorithm has been provided until now even though in some cases the PLS-PM loop is proven to converge monotonically, and the convergence is always reached in pratice.
The only two hypotheses underlying PLS models are:

- Each variable is measured on a interval (or ratio) scale;
- Relations between variables and latent constructs are linear and, consequently, monotone.

Therefore, standard PLS methods cannot handle data which are measured on a scale which does not have metric properties, nor non-linear relationships.
To overcome this problem a recent technique called Non-Metric Partial Least Squares (NM-PLS) algorithm has been set up. It consists in a new class of PLS algorithms that allow the PLS iteration to work as an optimal scaling algorithms, calculating iteratively both scaling and model parameters.

2.1 NM-PLSPM

The name Non-Metric are so called thanks to their capability to provide optimally scaled data ($\hat{\mathbf{x}}$) with a new metric structure, which does not depend on the metric properties of the raw data ($\mathbf{x}^*$). In other words, NM-PLS methods yield a metric to

non-metric data, and a new metric to metric data, linearizing the relations between variables and latent constructs, as required by the hypotheses of standard PLS models.
The NM-PLS algorithms optimize criteria under two sets of parameters: the model parameters and the scaling parameters constrained to the restrictions due to the scaling level chosen for each raw variable $\mathbf{x}^*$. In the NM-PLS framework the quantifications are not determined by an external criterion but are obtained by the optimal quantifications method with respect to a latent construct called *Latent Criterion (LC)* which is represented by an unknown vector (centered by construction), for which we use the generic symbol γ_{x^*}. For the NM-PLS, three levels of scaling are adopted according to measurement scale of the variables: nominal, ordinal and polynomial (or functional). A scaling (numeric) value [11] is assigned to each of the K categories (or distinct values) ϕ_k $(k = 1, \dots, K)$ of $\mathbf{x}^*$, such that:

- it is coherent with the chosen scaling level;
- it optimizes the model criterion.

In this way, each raw variable $\mathbf{x}^*$ is transformed as $\hat{\mathbf{x}} \propto \tilde{\mathbf{X}}\phi$ where $\phi' = (\phi_1, \dots, \phi_K)$ is the vector of optimal scaling parameters. The matrices $\tilde{\mathbf{X}}$ are the indicator matrices of the different categories of variables and they define a space in which the constraints imposed by the scaling level are respected. For example, at nominal scale level grouping property is preserved while ordinal scale level preserves grouping and order properties. The symbol $\propto$ means that the left side of the equation corresponds to the right side normalized to unitary variance. The raw data $\mathbf{x}^*$ are transformed by different real functions (scaling functions) $Q(\mathbf{x}^*\phi, \gamma_{\mathbf{x}^*})$, one for each scaling level, which generate the optimal scaled value $\hat{\mathbf{x}}$ for each observations. The scaling functions Q optimize the criterion

$$\arg\max_{\phi} \operatorname{cor}^2 \left(\tilde{\mathbf{X}}\phi, \gamma_{x^*}\right)$$

under the constraints chosen for the $\mathbf{x}^*$.
The resulting scaling values for the different $\mathbf{x}^*$ are the least square regression coefficients of $\mathbf{X}$ on γ which correspond to the average of $\gamma_{\mathbf{x}^*}$ conditioned to $\mathbf{x}^*$ categories. The geometric representation of the scaled variable $\hat{\mathbf{x}}$, normalized to unitary variance, can be obtained projecting $\gamma_{\mathbf{x}^*}$ on the space defined by the columns of $\tilde{\mathbf{X}}$.

2.2 Assessment of the model

In the PLS-PM frame, due to the fact that the model does not require distributional assumptions, the estimates of the parameter variability are obtained empirically by means of a bootstrap procedure. The validation of the quality of the model can also be studied by the evaluation of a few indicators that we briefly discuss in the following [4, 13].
For the measurement model the loadings represent the correlations between a latent

variable and its indicators whereas the communalities are the squared correlations. They represent the amount of variability explained by a latent variable (e. g. a loading greater than 0.7 means that more than $0.7^2 \approx 50\%$ of the variability in an indicator is explained by its latent variables). Therefore a value around 0.7 or more is usually considered good for the loadings. The average communality ($Av.C$) represents how much of the block variability is reproducible by the latent variable and the average variance extracted (AVE) represents the amount of variance that a latent variable captures from its manifest variables in relation to the amount of variance due to measurement errors. A good value of AVE index is at least 0.50 which means that 50% or more of the variance is accounted for.
For the structural model, the parameters generally taken into account are: the determination coefficients (R^2), the redundancy index and the average redundancy ($Av.R$). The R^2 represents the amount of variance in the endogenous latent variable explained by its independent latent variables. The redundancy index represents the amount of variance in an endogenous construct explained by its independent latent variables. High redundancy means high ability to predict. The average redundancy represents the percentage of the variance in the endogenous block that is predicted from the independent LVs associated to the endogenous LV. This index and the R^2 index are available only for the endogenous construct.
An index that takes into account the model performance in both the measurement and structural model and thus provides a single measure for the overall prediction performance is the GoF that the goodness of fit of the whole model. GoF is calculated as the geometric mean of the average communality and the average R^2 value.

3 Dataset and Model

In this paper we take into account a sub-set of the UNI.CO archive: we consider only the master degree graduates of the Sapienza University who belong to the engineering disciplinary sector. Moreover we consider only graduates that subscribed more than one contract during the three years after graduation (458 statistical units). In this preliminary study we propose a model in which the Job Success depends on the Educational and Job curricula. The set of manifest variables for each of the three latent variables representing Job Success, Educational Curriculum (Edu. Curr.) and Job Curriculum (Job Curr.) are described in Tab. 1. In the Job Success block are included as manifest variables only the two composite indicators: Optimal and Quasi-Optimal. In our model all the manifest variables are treated as reflective i. e. the LVs are to be considered as the cause of the MVs belonging to its own block. We performed a Non-Metric PLSPM analysis on the model by using the option centroid for the inner weight estimation (this choice only considers the sign of the correlations between a LV and its adjacent LVs). As shown in Fig.1 our model relates Job Success with Edu. Curr. and Job Curr. and also it analyses the relationship between Educational and Job curricula.

Table 1 Set of manifest variables for each latent variable.

LVs	MVs	Description	Scale
Edu. Curr.	Age	Class of age at university graduation	Numerical
	Final grade	Final university grade	Numerical
	Average grade	Average graduation grade	Numerical
	Isee	Indicator of economic equivalent situation: it measures the economic status of the families	Ordinal
Job Curr.	N_cn	Class of number of job relationships	Ordinal
	gg_work	Class of number of worked days	Ordinal
	gg_isco12	Class of number of worked days with high professional position	Ordinal
	gg_al243	Class of number of worked days with an actual duration of the contract of at least 8 months	Ordinal
	gg_CTI	Class of number of worked days with a permanent contract	Ordinal
Job Success	Optimal	The graduate has got optimal contract	Nominal
	Quasi-optimal	The graduate has got a quasi-optimal contract	Nominal

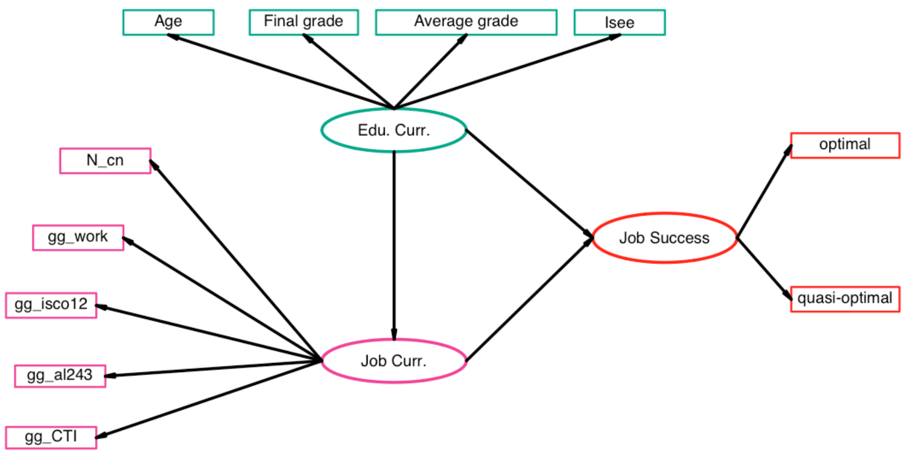

Fig. 1 Path Diagram depicting our model.

4 Discussion of the results

We performed a NM-PLSPM analysis on the model described in Sect. 3 using a code written in the R language by Russolillo and described in [11], details about the PLS-PM in R are given in [13, 15]. A new improved version of the PLS-PM R package, containing the non metric extension, will be available shortly when the

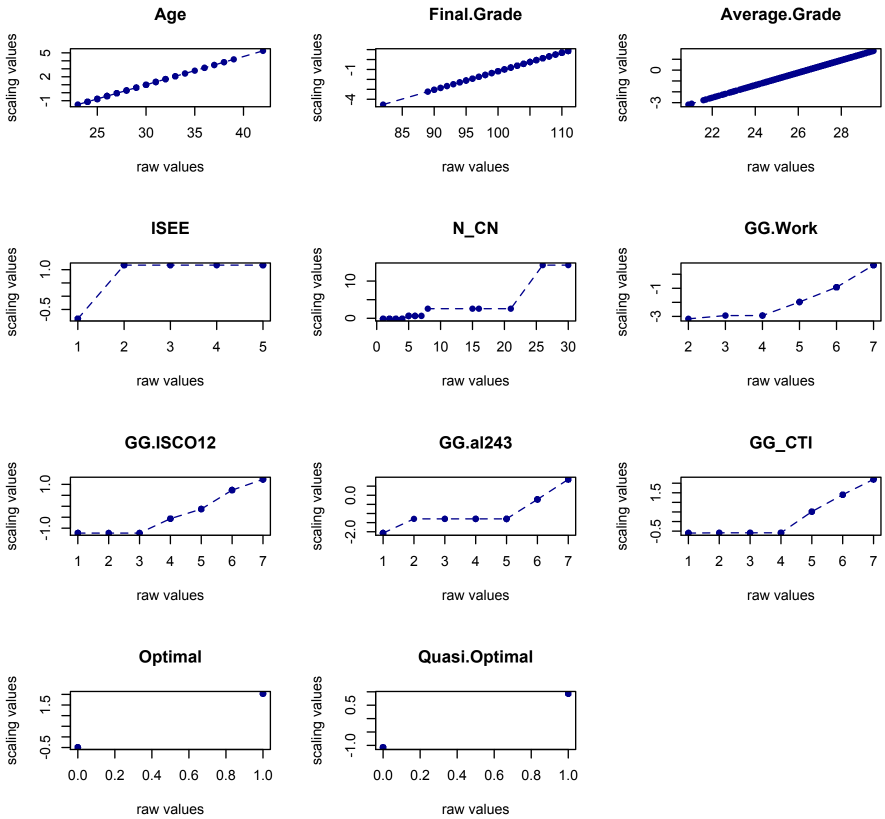

Fig. 2 Values of the original variable plotted versus corresponding optimal scaling values

present test phase will be completed[2].

The iterative algorithm of the Partial Least Square Path Modeling separately estimates the several blocks of the measurement model and then, in a second step, estimates the structural model coefficients. In the Non-Metric PLSPM the standard PLSPM procedure is combined with optimal scaling methods, during the cycles of iteration the model and the scaling parameters are alternately optimized in a modifies PLSPM loop where the quantification phase is added.

In our case the convergence of the algorithm has been achieved after only 9 cycles. In Fig. 2 we report, for all the variables, the plots of the raw values versus the scaling values obtained at the end of the convergence of the iterative procedure. These plots show that all the non-metric manifest variables are properly quantified using monotone transformations of the quantitative MVs.

---

[2] Russolillo, G. and Sanchez, G. (2013), private communication.

As matter of comparison, we have performed the basic analysis by PLSPM (without optimal scaling) replacing all the categorical variables with dummy variables, produces in many cases a non monotonic relationship between the dummy and the LV score with a non clear interpretation of the results. In the following we report the results coming from the application of the NM-PLSPM to our model.
We start now to examine the results of the outer model. The values of the validation

Table 2 Main results of the measurement (outer) model with corresponding 95% confidence intervals built by means of 1000 bootstrap samples. The weights and loadings (λ) are shown.

LVs	MVs	Weights	λ	Std.Error	perc.025	perc.975
Edu. Curr.						
	Age	-0.10	-0.80	0.12	-0.85	-0.70
	Final Grade	0.11	0.93	0.13	0.88	0.96
	Average Grade	0.12	0.94	0.13	0.89	0.96
	ISEE	0.04	0.29	0.16	-0.10	0.47
Job Curr.						
	N_CN	-0.06	-0.11	0.12	-0.28	0.14
	GG Work	0.23	0.78	0.03	0.70	0.82
	GG ISCO12	0.42	0.74	0.03	0.68	0.78
	GG al243	0.26	0.78	0.03	0.70	0.82
	GG_CTI	0.25	0.57	0.05	0.45	0.66
Job Success						
	Optimal	0.27	0.82	0.03	0.76	0.86
	Quasi Optimal	0.33	0.89	0.02	0.85	0.92

parameters of the outer model with corresponding 95% confidence intervals built by means of 1000 bootstrap samples are reported in Tab.2. In the block of Edu. Curr. we have for all the MVs high loadings with the exception of the manifest variable Isee (0.29). Thus we could consider removing this variable from the model. Moreover the empirical validation of the model shows that this value is not significant. The block of Edu. Curr. is positively affected by all the its MVs with the exception of Age that is negatively correlated. This is a trivial fact because the older the graduates the less is the study success.
In the block of Job Curriculum we find a similar situation to the one found in the previous block. Only for N_cn we have a very small values of the loadings (0.11). Also in this case the bootstrap procedure indicates a non significant value. We have checked that removing Isee and N_cn from the model, the GoF increases from 0.42 to 0.46. All the MVs in this block are positively correlated with the own LV with the exception of N_cn that is negative. In the block of Job Success we have high loadings for all the MVs.
The Optimal and Quasi Optimal indicators are discriminant to the construction of the Job Success block, see Tab.2. In fact, the weights of the MVs quantified at nominal scaling, which reflect the variability of the corresponding LV explained by the categories of the MVs, have high values particularly in the case of the Quasi Optimal indicator.

Table 3 Results of the structural (or inner) model with corresponding 95% confidence intervals built by means of 1000 bootstrap samples. The R^2 and the path coefficients (β) are shown.

Paths	R^2	β	Std.Error	perc.025	perc.975
Edu. Curr. $\rightarrow$ Job Curr.	0.09	0.30	0.04	0.211	0.37
Edu. Curr.$\rightarrow$ Job Success	0.54	-0.07	0.03	-1.08	0.02
Job Curr. $\rightarrow$ Job Success		0.75	0.02	0.72	0.79

Results of the structural model with corresponding 95% confidence intervals built by means of 1000 bootstrap samples are reported in Tab.3. The path coefficient from Edu. Curr. to Job Curr. is moderately small (0.30) indicating a feeble influence of educational curriculum on job experiences. In the case of the regression of Job Success in respect of Edu. Curr. and Job Curr. we see that, while the Job Curr. influences the Job Success very much (0.75), the Edu. Curr. has a small coupling with Job Success and a negative sign (-0.07). In other words, it is the case in which Edu. Curr. is correlated with Job Curr. but it is almost uncorrelated with Job Success. It is a situation in which Edu. Curr. acts as a suppressor variable. It is also interesting to also note that the indirect effect of Edu. Curr. to Job Success i. e. the path: Edu. Curr.$-$Job Curr.$-$Job Success, gives a positive contribution of 0.17 which is not negligible.
The bootstrap intervals for the path coefficient of Edu. Curr. to Job Success contain the value zero, so this coefficient is not significant a 5% confidence level, see Tab.3. The results of this regression suggest to analyse a simpler inner model in which Edu. Curr. is linked with Job Curr. and Job Curr. with Job Success.This path follows the natural temporal sequence from Edu. Curr. to Job Curr. and then to Job Success of a standard student. We have checked this model and the results are substantially unchanged.
In Tab.3 we also reported the values of R^2 of the endogenous latent variables for each regression in the structural model. We have $R^2 = 0.09$ for the regression where the endogenous variable is Job Curr. and a higher value $R^2 = 0.54$ in the case of the endogenous variable Job Success. In order to evaluate these values, it should be taken into account that high values of R^2 are not expected because our endogenous manifest variables (Optimal and Quasi Optimal) in the block of Job Success are binary and they are analysed together with nominal, ordinal and numerical variables. The values of the main goodness indices obtained from our model are reported in

Table 4 Results of the main indices for the evaluation of the model. Average Communalities (Av.C), Average Redundancy (Av.R), AVE and GoF are shown.

LVs	Type	Av.C	Av.R	AVE
Edu. Curr.	Exogenous	0.62		0.62
Job Curr.	Endogenous	0.42	0.03	0.42
Job Success	Endogenous	0.73	0.40	0.73
GoF	0.42			

Tab. 4. The average redundancy for Job Success indicates that Edu. and Job Curricula predict 40% of the variability Job Success indicators whereas the average

redundancy for Job Curriculum indicates that Edu. Curriculum predicts lower value of 3% of the variability of Job Curriculum. The AVE index shows good values for all our constructs except for Job Curriculum. Finally, we obtained that the whole prediction power of the model is GoF=0.42.

5 Conclusions

We have presented one of the first statistical analysis based on the data of the UNI.CO archive which is the more complete administrative archive of the Sapienza graduates available to date.
In this study, the NM-PLS have demonstrated a great adaptability to handle a large dataset with numerical, nominal and ordinal variables therefore confirming that the NM-PLS approach makes the PLS methodology even more flexible. The manifest variables are properly quantified by the optimal scaling technique that is adopted in this new procedure and it is implemented in the new R package.
The model studied in this paper to the aim of analysing the job success, taking into account the fact that the database is large and that it contains non-metric variables, gives a satisfactory representation of the data variability.
The high values of the measurement model have confirmed that the Optimal and Quasi Optimal indicators are discriminant to the construction of the Job Success block. We have seen that two variables (Isee and C_cn) can be removed without reducing the capacity of the model to explain the variance and also the structural inner model can be reduced to a model with a simpler structure where the path among the LVs becomes Edu. Curr., Job Curr. and Job Success. We have found that the age of graduation influences negatively the final job success, therefore the early conclusion of the scholastic career positively affects the success in the labour market. The overall frame that arises from this study is that the scholastic path of Sapienza engineering graduates does not seem to have a great direct influence to the aim of getting a satisfactory job.
However the model adopted in this preliminary work, that it has been chosen for its simplicity, probably does not contain the needed flexibility to explain the large quantity of information contained in UNI.CO archive. A model with more complex structure capable to recognize new composite indicators of the job success is under study.

Acknowledgements I would like to warmly thank profs. G. Alleva, J. Mortera, G. Russolillo, and S. Terzi for their invaluable help during my PhD studies and the writing up of this paper, special thanks to prof. S. Terzi for reviewing the paper. I would like to thank prof. G. Saporta for the warm hospitality at the CNAM in Paris. I would like to thank Dr. G. Sanchez for discussions about the new R package.

References

1. Alleva, G., Petrarca, F. (2013), New indicators for investigating the Integration of Sapienza graduates into the labour market, working papers n. **120**/2013 del Dipartimento Memotef, ISSN 2239-608X.
2. Alleva, G., Petrarca, F., Renda, E., Lucisano, P. and Magni, C., (2012), Potenzialità della matrice UNI.CO per lo studio delle caratteristiche della domanda di lavoro dei laureati della Sapienza: primi risultati e possibili sviluppi, Workshop on “Monitoring of the dynamics of the professional graduates”. Italia Lavoro, 24th September 2012.
3. Bollen, K. A. (1989), Structural equations with latent variables. New York: Wiley.
4. Esposito Vinzi, V., Chin, W., Henseler, J., Wang, H., (2010), eds., Handbook of Partial Least Squares (PLS): concepts, methods and applications. Berlin, Heidelberg, New York: Springer.
5. Esposito Vinzi, V., Russolillo, G., (2013), Partial Least Squares algorithms and methods. WIREs Comput Stat 2013, 5:1-19. doi:10.1002/wics.1239.
6. Fabbris, L., (2012), ed., Indicators of higher education effectiveness, Milano, McGraw-Hill.
7. Kaplan, D., (2000), Structural equation modeling: foundations and extensions. Thousands Oaks, California: Sage.
8. Jencks, C., Perman, L., Rainwater, L., (1988), What is a good job? A new measure of labor-market success. The American Journal of Sociology. **93**(6), 1322–1357.
9. Jöreskog, K. (1970), A general method for analysis of covariance structure, Biometrika **57**, 239-251.
10. Leschke, J, Watt, A., (2008), Job quality in Europe.WP 2008-07, ETUI-REHUS, Brussels.
11. Russolillo, G., (2012), Non-Metric Partial Least Squares. Electronic Journal of Statistics. Vol. **6**, 1641–1669.
12. Russolillo, G. and Trinchera, L., (2010), On the use of structural equation models and PLS path modeling to build composite indicators, working paper n. **30**/2010 Dipartimento degli Studi sullo Sviluppo Economico, Università degli Studi di Macerata.
13. Sanchez, G., PLS path modeling with R. http://www.gastonsanchez.com/PLS_Path_Modeling_with_R.pdf, http://www.gastonsanchez.com/pathdiagram. Last access, 15th September 2013.
14. Tenenhaus, M., (1998), La Régression PLS: théorie et pratique, Technip, Paris.
15. Sanchez, G., Trinchera, L.: R package of PLSPM. http://cran.r-project.org/web/packages/plspm/index.html. Last access, 15th September 2013.