Proceedings of the World Congress on Engineering 2010 Vol III
WCE 2010, June 30 - July 2, 2010, London, U.K.

PLS Path Modeling with Mode C Computational Experiments

Alba Martinez-Ruiz*, Tomas Aluja-Banet

Abstract—**Monte Carlo simulations and computational experiments were carried out to study the performance of partial least squares (PLS) path modeling with mode C. The empirical results are in line with the theoretical PLS framework. Inner relationships are underestimated and outer relationships overestimated.**

Keywords: partial least squares, structural equation models, Monte Carlo simulation

1 Introduction

Reflective relationships seek to represent variances and covariances between the manifest variables that are generated or caused by a latent variable. So, observed variables are treated as an effect of unobserved variables [2]. In a reflective measurement model, the manifest variables are measured with error (Figure 1(a)). Alternatively, formative relationships are used to minimize residuals in the structural relationship [9]. Here, manifest variables are treated as forming the unobserved variables, they are presumed to be error–free, and the construct is estimated as a linear combination of the manifest variables plus a disturbance term (Figure 1(b)). As in this case all variables forming the construct should be considered, the disturbance term represents all those non-modeled causes. Although formative measurement models were first discussed by [4] and [1], and a number of variables can be modeled in a better way through formative relationships, measurement variables have been traditionally modeled in a reflective mode. [9] and [5] pointed out that modeling formative modes using a covariance-based approach may lead to identification problems and Heywood cases. So, researchers may tend to define outer models as reflective. However, a number of researchers have pointed out that PLS path modeling overcomes the identification problems that arise when implementing a covariance-based approach [15, 17, 19]. That is because a PLS path modeling algorithm consists of a series of ordinary least squares (OLS) analysis. From a component-based approach, and "because the off-diagonal elements are not among the unknown parameters of the model and because the unobservables are explicitly estimated, there are no identification problems for recursive PLS models" [9, p. 443].

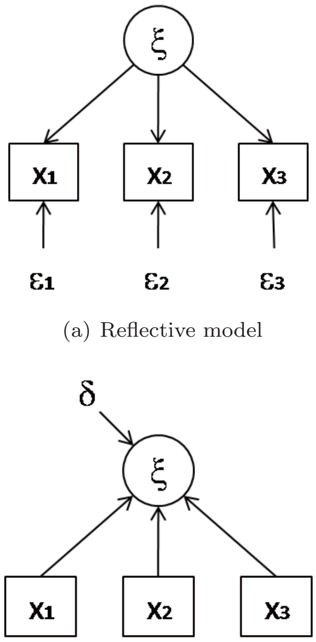
In Wold's PLS approach, a construct is completely determined by a linear combination of its indicators [17, 18, 19]. The procedure usually uses a Mode A or Mode B to model a structural equation model (SEM). Mode A or simple regression if the SEM includes reflective outer models. Mode B or multiple regression if formative outer models are included. However, "the algorithm is called PLS Mode C if each of Modes A and B is chosen at least once in the model" [18, p. 10]. To the best of our knowledge, there are only a small number of published articles that examine the performance of PLS path modeling algorithm in the presence of formative outer models, and they are not conclusive. Findings by [3] and [13] are quite different. For instance, Cassel et al. found that measurement relationships in formative outer models are overestimated, while Ringle et al. found that these relationships are underestimated. Thus, this paper aims to provide evidence regarding how well PLS path modeling performs if formative exogenous outer models are modeled using PLS Mode B and reflective endogenous latent variables are modeled using PLS Mode A. That is, PLS path modeling with mode C.

2 PLS Path Modeling

The PLS path modeling procedure –presented by Gerlach, Kowalski, and Wold in 1979– is a soft modeling technique and a data analytic tool for estimating structural equation models (SEM) and building a sequence of latent variables. PLS path modeling first estimates the unobservable variables and then the parameters with an aim toward maximizing the total variance and minimizing residuals of endogenous models regardless of the covariances among manifest variables. The structural or inner model describes relationships among constructs ξ_i by means of multiple regressions (Equation 1). ξ_i and ξ_j are the exogenous and endogenous latent variables, respectively, and β_{ji} are the path coefficients that measure the relationship among constructs. The condition imposed by Herman Wold is predictor specification,

*This research was supported by a grant from the Universidad Católica de la Ssma. Concepción, and the Comisión Nacional de Investigación Científica y Tecnológica (CONICYT), Chile. A. Martinez-Ruiz and T. Aluja-Banet are with Technical University of Catalonia, Department of Statistics and Operations Research, Jordi Girona 1-3, C5, 08034 Barcelona, Spain. Tel/Fax: 93-401-7986/5855 E-mails: alba.martinez-ruiz@upc.edu, tomas.aluja@upc.edu.

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)



(a) Reflective model

(b) Formative model

Figure 1: Reflective and formative measurement models (focus on component-based approach)

$E(\xi_j/\xi_i) = \sum_i \beta_{ji}\xi_i$, that is, there is no linear relationship between predictor and residual. This condition implies that $E(\nu_j/\forall\xi_i) = 0$, and $cov(\nu_j, \xi_i) = 0$.

$$\xi_j = \beta_{j0} + \sum_i \beta_{ji}\xi_i + \nu_j \tag{1}$$

Manifest variables revealing or reflecting the effect of a construct are modeled as indicators of it in a reflective measurement model. Each manifest variable x_{jh} is related by simple ordinary least squares regression with the underlying construct ξ_j (Equation 2). The loadings λ_h determine the extent to which each indicator reflects a construct; ξ_j is a common factor with mean m, standard deviation one and it is indirectly observable by the manifest variables. The condition imposed by Herman Wold is predictor specification, $E(x_h/\xi) = \lambda_{h0} + \lambda_h\xi$. This condition implies that ϵ_h has zero mean, and it is uncorrelated with ξ_j. Moreover, the basic design of Herman Wold assumes that the covariance matrices of all ϵ_j are diagonal. As in a reflective model, where all the indicators of the block of variables reflect the same construct, there should be high collinearity among these variables. That is, the blocks of variables must be one-dimensional.

$$x_{jh} = \lambda_{jh0} + \lambda_{jh}\xi_j + \epsilon_{jh} \tag{2}$$

The latent variable is formed by a set of manifest variables as a linear function of them plus a residual in formative outer models (Equation 3). The weights π_h determine the extent to which each indicator contributes to the formation of the constructs. Each block of manifest variables may be multidimensional, and multicollinearity among indicators is not a necessary constraint. The condition imposed by Herman Wold is predictor specification $E(\xi/X_1, ..., X_{pj}) = \sum_h \pi_h x_h$. This condition implies that the residual δ has a zero mean, and it is uncorrelated with the manifest variables x_h. Since each construct is formed by a linear combination of the manifest variables, the sign of each weight π_h should be the same sign as the correlation between x_h and ξ [15].

$$\xi_j = \sum_h \pi_{jh}x_{jh} + \delta_j \tag{3}$$

2.1 PLS path modeling algorithm

The PLS path modeling algorithm is structured in three stages [17, 18, 19]. The first stage computes the case values of the latent variables; the second stage focuses on the inner and outer relationships; and in the third stage, location parameters of the latent variables, λ_{jh0} and β_{j0}, are estimated. Only the first stage is iterative. The algorithm for Wold's procedure is as follows.

The first stage. The algorithm starts choosing an arbitrary weight vector –outer weights– to first relate each latent variable with their own manifest variables. Usually this vector is a vector of ones. Each standardized latent variable Y_j –zero mean, unit variance– is computed as an exact linear combination of their own centered manifest variables:

$$Y_j = \sum w_{jh}x_{jh} \tag{4}$$

where w_{jh} are called the outer weights.

An auxiliary latent variable Z_j is introduced as a counterpart to the variable Y_j. Each Z_j is computed as a weighting sum of the latent variables which is related to:

$$Z_j \propto \sum e_{ji}Y_i \tag{5}$$

where e_{ji} are called the inner weights. There are three different weighting schemes that may be used to compute e_{ji}: the centroid, the factorial and the path weighting schemes. The first was introduced by Wold, and the last two by [11]. The simplest scheme is the centroid scheme where the e_{ji} are equal to the signs of the correlations between Y_j and the Y_i's. The inner weights are equal to the correlation between Y_j and Y_i when the factorial scheme is considered. The inner weights in a path weighting scheme are (a) equal to the regression coefficients of Y_i in the multiple regression of Y_j on all the Y_i related to the predecessor of Y_j, or (b) are equal to the correlation between the successor of Y_i and Y_j.

Once the auxiliary latent variables are estimated, the weights w_{jh} are recomputed. Recall that, in the itera-

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

tive process, these weights are used to estimate all latent variable scores as a linear combination of their own indicators. The procedure considers two ways of recomputing the outer weights, depending on the reflective or formative nature of the outer models: mode A and mode B. Usually, mode A is considered for recomputing the outer weights when outer models are reflective, and mode B is considered for recomputing the outer weights when outer models are formative. However, this rule is not mandatory. Depending on models, data characteristics and also the researcher's discretion, one mode or another will be more appropriate for a particular case. For mode A, the w_{jh} is the regression coefficient of Z_j in the simple regression of x_{jh} on the inner estimation of Z_j:

$$w_{jh} = cov(x_{jh}, Z_j) \qquad (6)$$

For mode B, the vector w_j of weights w_{jh} is the vector of the regression coefficient in the multiple regression of Z_j on the manifest variables $(x_{jh} - \widetilde{x}_{jh})$ related to the same latent variable Z_j:

$$w_j = (X_j' X_j)^{-1} X_j' Z_j \qquad (7)$$

The first stage is iterated until convergence.

The second stage. Once the algorithm converges, the latent variable scores estimated in stage 1 are used to estimate the inner and outer relationships by ordinary least squares regression without location parameters. If reflective blocks of variables are modeled, simple regression is used to estimate loadings (Equation 2). If formative blocks of variables are modeled, weights are estimated by ordinary multiple regression (Equation 3).

The third stage. The third stage focus on estimation of the location parameters, and the values of π_{jh0} and β_{j0} (Equation 1).

3 Monte Carlo Simulation Study

A Monte Carlo simulation study was designed to analyze the performance of PLS path modeling with mode C [12, 10]. The underlying population model considered a simple structure with three formative exogenous constructs and one reflective endogenous latent variable (Figure 2). Models with two, four, six and eight indicators per construct, and four different samples sizes (50, 100, 250, 500) were studied. Five hundred random data sets were generated for each of the 4×4 cells of the two-factor design. PLS path modeling with centroid scheme –as described in [15]– and bootstrapping were performed in R-project [14]. Five hundred replications (t) were made for each cell in the design. Results are provided in terms of the mean bias (accuracy, $\frac{1}{t}\sum_{i=1}^{t} E[\theta_i] - \theta$) and mean relative bias (MRB$= 100 * \frac{1}{t}\sum_{i=1}^{t} \frac{\theta - E[\theta_i]}{\theta}$, [7]).

The data were generated from a component-based model. We began generating standardized manifest variables x_{jh} for each formative outer model as independent normal data. Once the manifest variables were generated, we computed the exogenous constructs ξ_j and the endogenous latent variable η, so that the variance of the unobservable variables is one. The generated exogenous constructs are not collinear. The endogenous latent variable was calculated as linear combination of the exogenous constructs plus a disturbance term. Disturbance terms were computed as random normal data with a zero mean and the corresponding standard deviation. They were distributed independently of unobservable variables. Standardized observed variables y_i of reflective measurement models were generated as independent normal data. Errors of the reflective relationships were computed as random normal data with a zero mean and the corresponding standard deviation; they were also uncorrelated with the latent variable. To set the true population parameters for the models, we took into account different combinations of permissible values so as to see whether they are recovered by the PLS path modeling algorithm. Table 1 shows the true population values of weights, path coefficients and loadings. We consider large values for all the true loadings, at least 0.7 in the case of two manifest variables per construct. This ensures the unidimensionality of the block of variables and it satisfies the condition imposed by the PLS path modeling algorithm.

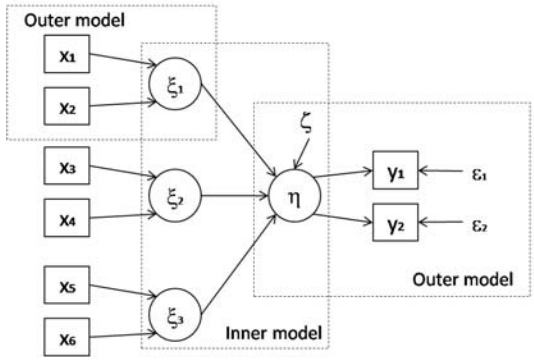


Figure 2: Inner and outer models of the simulated setups; outer models consider two, four, six, and eight indicators per construct.

4 Results

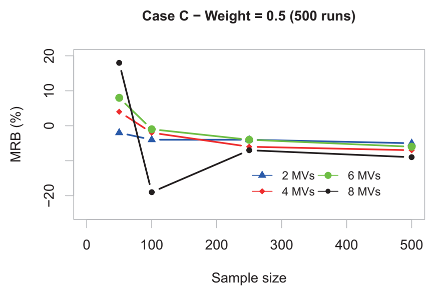
Some results are reported here. Figure 3 shows the mean relative bias of weights and loadings. The empirical results are in line with the theoretical PLS framework [8], and the true weights are overestimated by the PLS path modeling algorithm (Figure 3(a)). This confirms that PLS estimates are biased. Increasing the sample size, the bias decreases, and for N=500, PLS almost exactly recovers all the population values (small, moderate and large values). The largest MRBs are exhibited for models with the smallest sample sizes (N=50). In addition, the variability and mean square errors decrease by increas-

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

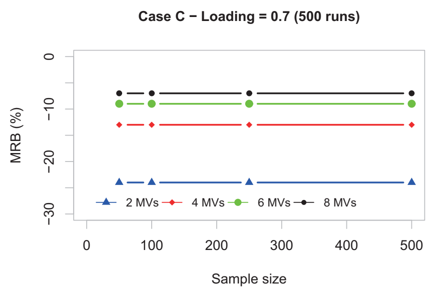
Table 1: Vector of true population values for weights, path coefficients and loadings; cases for two, four, six and eight indicators in each outer model.

MVs	Coefficient	True Values
2	Weights	(0.8,0.5)
		(0.4,0.8)
		(0.1,0.9)
	Path Coefficients	(0.5,0.4,0.6)[a]
	Loadings	(0.7,0.8)
4	Weights	(0.2,0.3,0.5,0.7)
		(0.2,0.4,0.6,0.5)
		(0.3,0.5,0.7,0.2)
	Path Coefficients	(0.5,0.4,0.6)[a]
	Loadings	(0.6,0.7,0.8,0.9)
6	Weights	(0.5,0.3,0.4,0.3,0.5,0.1)
		(0.2,0.4,0.6,0.4,0.2,0.3)
		(0.3,0.6,0.2,0.3,0.4,0.2)
	Path Coefficients	(0.5,0.4,0.6)[a]
	Loadings	(0.6,0.7,0.8,0.9,0.6,0.7)
8	Weights	(0.3,0.3,0.4,0.3,0.4,0.3,0.2,0.3)[b]
		(0.3,0.3,0.4,0.4,0.2,0.3,0.4,0.2)[b]
		(0.4,0.5,0.4,0.3,0.2,0.1,0.3,0.2)[b]
	Path Coefficients	(0.5,0.4,0.6)
	Loadings	(0.6,0.7,0.8,0.9,0.6,0.7,0.8,0.9)

[a] For N=50 the true path coefficient vector was (0.5,0.4,0.5).
[b] For N=50 and N=100 the true weight vectors were (0.3,0.1,0.4,0.3,0.4,0.3,0.2,0.2), (0.3,0.1,0.4,0.4,0.2,0.3,0.4,0.1), and (0.2,0.4,0.4,0.3,0.2,0.1,0.3,0.2).



(a) MRB for a weight of 0.5



(b) MRB for a loading of 0.7

Figure 3: Mean relative bias of a weight and a loading. Highlighting the influence of the sample size and the number of indicators per construct.

ing the sample size or increasing the number of manifest variables in all the simulated cases.

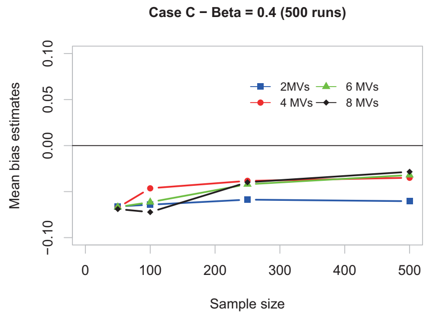
Simulations performed by [6] for PLS models with reflective relationships showed that, by themselves, neither the number of indicators nor the sample size substantively improve the quality of the estimates. Rather, it is necessary to increase both factors at the same time for an improvement in the quality of the estimates. Here, the simulations for PLS models with formative blocks of variables render the same aforementioned result. So, PLS path modeling is consistent and consistent at large. Nevertheless –and recalling that PLS algorithm computes the latent variables as an exact linear combination of the observed variables– the results suggest that estimates will improve by increasing the sample size more than increasing the number of observable variables, depending on the correlations between manifest variables. So, the researcher may suspect the type of relationship that she expects to find.

As can be seen in Figure 3(b), the estimates of loadings are very close to the true values in all cases, regardless of the sample sizes and number of manifest variables per construct. PLS path modeling overestimates the population values. Moreover, according to the results, a higher number of manifest variables seems to be more important than a higher sample size for decreasing the bias of the estimates in reflective outer models. This is in contrast with the formative relationships and coincides with the results found by other researchers [6]. This is clearly seen in Figure 3(b) where the mean relative bias for a loading of 0.7 strongly decreases when the number of indicators increases. So, this confirms that PLS estimates are consistent at large [6, 18].

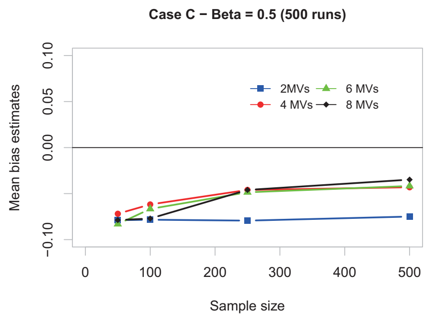
Results for estimates of inner relationships are quite conclusive. Figure 4 allows us to see how an increase in both the number of manifest variables and the sample size reduces the mean bias of path coefficients for all assumed true values. The algorithm underestimates the true path coefficients in all the analyzed cases. As the sample size increases, the estimates increasingly approach the true values and the biases decrease.

In accordance with expectations, when the models consider only two indicators, the mean bias is the largest and proves to be quite the same when sample size increases. The MRB ranges from 18% (N=50) to 15% (N=500). For models with four, six and eight indicators, the MRB decreases when sample size increases. Increasing the number of indicators per latent variable, it yields closer estimates to the true values; but this factor tends to have less influence on the quality of the estimates than the sample

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)



(a) True value = 0.4



(b) True value = 0.5

Figure 4: Mean bias of path coefficients. Highlighting the influence of the sample size and the number of indicators.

size does.

Summing up, in all analyzed cases, the results obtained are better when each outer model considers more manifest variables per construct. However, it is worth noting that the estimates are shown to be quite accurate and precise when the measurement models include only two indicators per construct. This suggests that PLS path modeling may be a robust alternative when estimating structural equation models with formative relationships and few indicators per construct.

5 Conclusions

For the studied model, the findings suggest that PLS path modeling offers a way to build "proper indices" for unobservable variables and to estimate the relationships between them. The procedure shows a tendency to overestimate outer relationships and underestimate inner relationships. It is worth noting that the estimates are shown to be robust when the measurement models include only two indicators per construct. It is true that when the number of observed variables and sample size increase, the quality of the PLS path modeling estimates increases. But when few indicators and a small sample size are considered, we can obtain acceptable estimates of the parameters. [16] have noted the same behavior in a reflective block of variables with two indicators. Finally, we think that the model simulated here represents a number of models that can be studied in real-world applications: those in which formative exogenous outer models are modeled using PLS Mode B and reflective endogenous latent variables are modeled using PLS Mode A. That is, PLS Mode C, in terms of Wold's approach.

References

[1] Blalock, H. *Causal Inferences in Non-experimental Research*, University of North Carolina Press, Chapel Hill, 1964.

[2] Bollen, K., Lennox, R., "Conventional Wisdom on Measurement: A Structural Equation Perspective," *Psychological Bulletin*, V110, N2, pp. 305-314, 1991.

[3] Cassel, C., Hackl, P., Westlund, A., "Robustness of Partial Least-Squares Method for Estimating Latent Variable Quality Structures," *Journal of Applied Statistics*, V26, N4, pp. 435-446, 1999.

[4] Curtis, R., Jackson, E., "Multiple Indicators in Survey Research," *The American Journal of Sociology*, V68, N2, pp. 195-204, 1962.

[5] Chin, W.W., "Issues and Opinion on Structural Equation Modeling," *MIS Quarterly*, V22, N1, 1998.

[6] Chin, W. W., Newsted, P. R., "Structural Equation Modeling Analysis with Small Samples using Partial Least Squares," *Statistical Strategies for Small Sample Research*, Hoyle, R. H. (ed), Sage Publications, California, pp. 307-341, 1999.

[7] Chin, W. W., Marcolin, B. L., Newsted, P. R., "A Partial Least Squares Latent Variable Modeling Approach for Measuring Interaction Effects: Results from a Monte Carlo Simulation Study and an Electronic-mail Emotion/Adoption Study," *Information Systems Research*, V14, N2, pp. 189-217, 2003.

[8] Dijkstra, T., "Latent Variables and Indices: Herman Wold's Basic Design and Partial Least Squares," *Handbook of Partial Least Squares: Concepts, Methods and Applications*, Esposito Vinzi, V., Chin, W. W., Henseler, J., Wang, H. (eds), Springer, Heidelberg, 2010.

[9] Fornell, C., Bookstein, F. L., "Two Structural Equation Models: LISREL and PLS Applied to Consumer Exit-Voice Theory," *Journal of Marketing Research*, V19, N4, pp. 440, 1982.

[10] Gentle, J., *Random Number Generation and Monte Carlo Methods*, Springer, New York, 2003.

[11] Lohmöller, J. B., *Latent Variables Path Modeling with Partial Least Squares*, Physical–Verlag, Heildelberg, 1989.

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)

[12] Paxton, P., Curran, P., Bollen, K., Kirby, J., Chen, F., "Monte Carlo Experiments: Design and Implementation," *Structural Equation Modeling: A Multidisciplinary Journal*, V8, N2, pp. 287-312, 2001.

[13] Ringle, C.M., Götz, O., Wetzels, M., Wilson, B., "On the Use of Formative Measurement Specifications in Structural Equation Modeling: A Monte Carlo Simulation Study to Compare Covariance–based and Partial Least Squares Model Estimation Methodologies," Maastricht University, METEOR Research Memoranda RM/09/014, 2009.

[14] R Development Core Team, "R: A Language and Environment for Statistical Computing," *R Foundation for Statistical Computing*, Vienna, Austria, http://www.R-project.org, 2007.

[15] Tenenhaus, M., Esposito Vinzi, V., Chatelin, Y. M., Lauro, C., "PLS Path Modeling," *Computational Statistics & Data Analysis*, V48, pp. 159-205, 2005.

[16] Vilares, M., Almeida, M., Coelho, P., "Comparison of Likelihood and PLS Estimators for Structural Equation Modeling: A Simulation with Customer Satisfaction Data," *Handbook of Partial Least Squares: Concepts, Methods and Applications*, Esposito Vinzi, V., Chin, W. W., Henseler, J., Wang, H. (eds), Springer, Heidelberg, 2009.

[17] Wold, H., "Model Construction and Evaluation when Theoretical Knowledge is Scarce: Theory and Application of Partial Least Squares," *Evaluation of Econometric Models*, J. Kmenta, J., Ramsey, G. (eds), Academic Press, New York, pp. 47-74, 1980.

[18] Wold, H., "Soft Modeling: The Basic Design and Some Extensions," *Systems Under Indirect Observation*, Jöreskog, K. G., Wold, H. (eds), North-Holland, Amsterdam, pp. 1-54, 1982.

[19] Wold, H., "Partial Least Squares," *Encyclopedia of Statistical Sciences*, Kotz, S., Johnson, N.L. (eds), Wiley, New York, V6, pp. 581-591, 1985.

ISBN: 978-988-18210-8-9
ISSN: 2078-0958 (Print); ISSN: 2078-0966 (Online)