Multilevel analyses in PLS-SEM:
An anchor-factorial with variation diffusion approach

Ned Kock

Texas A&M International University, USA

Abstract

A multilevel analysis, in the context of structural equation modeling via partial least squares (PLS-SEM), can be seen as an analysis in which: (a) data is collected at the individual level from multiple groups, and (b) group membership is expected to influence data analysis results. In this paper we illustrate such an analysis employing WarpPLS, a leading PLS-SEM software tool. The analysis employs an anchor-factorial with variation diffusion approach.

Keywords: Multilevel Analysis; Instrumental Variable, Structural Equation Modeling; Partial Least Squares; WarpPLS.

Introduction

A multilevel analysis employing structural equation modeling via partial least squares (PLS-SEM) is characterized by data being collected at the individual level from multiple groups, where group membership is expected to influence data analysis results (Kock et al, 2017); notably path coefficients. In this paper we exemplify and discuss such an analysis utilizing WarpPLS 7.0 (Kock, 2020a; 2020b). The analysis employs an "anchor-factorial with variation diffusion" approach (Kock, 2020a).

Illustrative model and data

Figure 1 shows the illustrative model that is used as a basis for our discussion. It contains four latent variables: the degree to which individuals in a company uses task-specific technologies (TU); the education level of the individuals (ED); the problem-solving ability of the individuals (PS); and the job performance of the individuals (JP).

We employed the Monte Carlo simulation method (Kock, 2016) to create data: 300 cases, each case referring to an individual in a company. The existence of three companies was assumed in the data creation process, each of the companies representing a business area: farming, manufacturing, and technology. That is, even though the unit of analysis is the individual, the individuals came from three separate companies.

The model incorporates three predictions, which are based on past empirical research on related topics. The predictions are that job performance (JP) is significantly and positively associated with technology use (TU), education level (ED), and problem-solving ability (PS). Increases in these three variables are hypothesized to cause increases in JP.

Figure 1: Illustrative model used
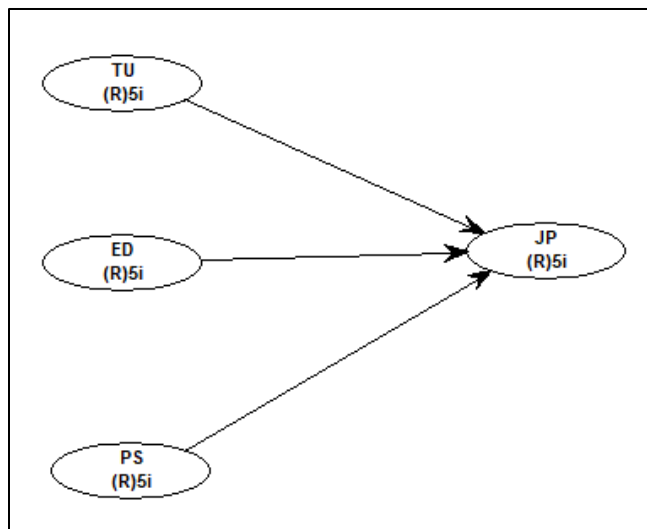
Note: model notation used is the same as that employed by Kock (2020b).

What characterizes a multilevel analysis?

A multilevel analysis is characterized by data being collected at the individual level from multiple groups, where group membership is expected to influence path coefficients. Consider our illustrative data. Each of the three companies from which data was collected consists of a group of individuals, with the data collected at the individual level. Since each company is of a different type (farming, manufacturing, or technology), company membership is expected to influence analysis results.

This situation creates an endogeneity problem (Kock et al, 2017), where a hidden variable is expected to influence an endogenous variable indirectly through its predictors. Figure 2 illustrates this problem. The hidden variable is company type (CO), which influences each of the three predictors: TU, ED and PS. As such, variation from CO flows indirectly into the endogenous variable JP. This effect, if strong, could significantly bias the path coefficients associated with the links: TU > JP, ED > JP, and PS > JP.

To solve this problem, we need to create an instrumental variable that incorporates the variation in CO that ends up in JP; via the intermediate effects on TU, ED, and PS. We then need to control for the effect of this instrumental variable with respect to JP. That is, we will add this instrumental variable into the model by making it point at JP.

An anchor-factorial with variation diffusion approach

In our example, the company type variable is categorical, and is stored as a data label column using three identifiers: "FARM", "MANU" and "TECH". These refer respectively to farming, manufacturing, and technology. We need to use the menu option "Explore categorical-numeric-categorical conversion", under the "Explore" menu options of WarpPLS, to perform a

categorical-to-numeric conversion and obtain the indicator for the instrumental variable CO. For this to be done, we first need to analyze the model without CO being added to it.

Figure 2: Multilevel model with endogeneity
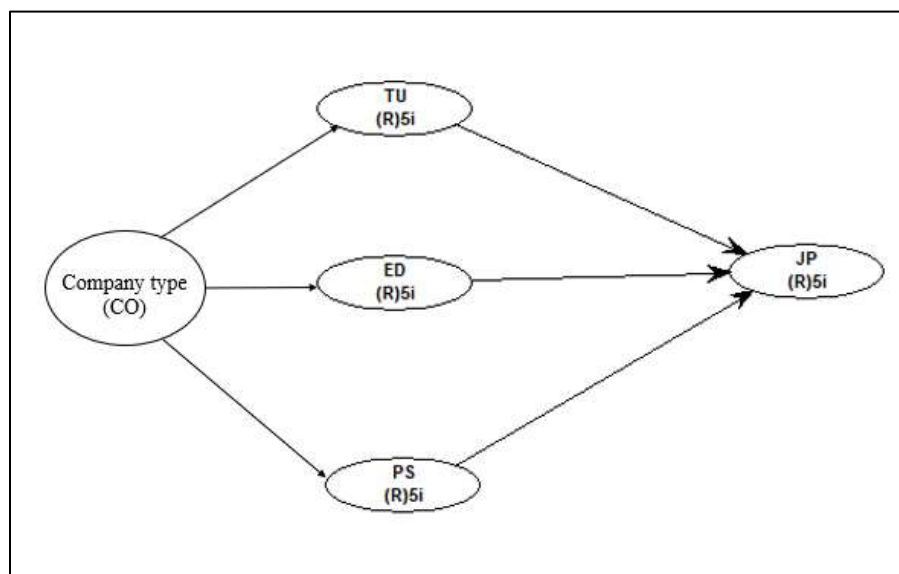
The conversion mode "anchor-factorial with variation diffusion" should be employed in cases like this, when the new instrumental variable is expected to be included in the model as a control variable. As noted by Kock (2020a), this option is a more sophisticated alternative to perform multilevel analyses than the group mean variable approach discussed by Kock & Hadaya (2018) in Appendix F of their article (see also: Kock et al, 2017).

Figure 3 shows the options chosen for categorical-to-numeric conversion in our illustrative example. The anchor latent variables are ED, TU, and PS. The correlation signs associated with each of these anchor latent variables reflect the expected signs of their relative relationships with the categorical variable. In this case, the signs are all set as positive, because technology use (TU), education level (ED), and problem-solving ability (PS) are all expected to be influenced in the same direction by company type. For example, "TECH" companies are expected to be associated with higher scores in these three variables than "MANU" and "FARM", and "MANU" associated with higher scores than "FARM". These expectations are based on theory and prior research.

The absolute correlations with the categorical variable, shown on the screen, are meant to give the researcher an idea of the strength of the associations among the quantified categorical variable, the anchor variables (ED, TU, and PS), and the endogenous variable that is indirectly affected (JP). The new instrumental variable will be created with a name like "c2n_CO", where the "c2n" part indicates that this new variable is the result of a categorical-to-numeric conversion of the data label variable noted as CO.

Figure 3: Categorical-to-numeric conversion

Conversion type		Categorical-to-numeric conversion mode			
Categorical-to-numeric	⌄	Anchor-factorial with variation diffusion			
Notes: start by selecting a conversion type; categorical variables are stored as label variables; each c					
variable; possible anchor latent variables are priority-ordered from left to right; numeric-to-categorical					

Categorical-to-numeric conversion					
**********************************					
Mode: Anchor-factorial with variation diffusion					
Current categorical variables					

Click to choose (only one):	CO				
Categorical variable to be converted					

Selected (click to remove):	CO	(3 catgs.)	(N=300)		
Possible anchor latent variables					

Click to choose one or more:	CO	ED	JP	TU	PS
Abs. correl. with categorical variable:	1.000	0.477	0.374	0.363	0.310
Latent variables to be used					

Selected (click to remove):	ED	TU	PS		
Correl. sign (click to switch):	+	+	+		

Once this new variable is created, it will be added to the dataset as a new standardized indicator. The scores of this new variable can be inspected through the option "View or save raw indicator data", under the "Data" menu option. Unlike raw scores obtained from unstandardized scales, these raw scores are created directly in standardized format. The scores can be inspected side-by-side with the corresponding data label values; the latter can be viewed through the "View or save data labels" option.

In our example this side-by-side inspection yields the following scores: -1.352 for "FARM", 0.096 for "MANU", and 1.246 for "TECH". Since these scores are standardized, they usually vary from -2 to 2, with a mean of 0 (zero). Therefore, we can see that the quantified categorical variable has an intuitively appealing relationship with the underlying company type, which ultimately influences the job performance of the individuals (JP) in the company: low for the farming company, average for the manufacturing company, and high for the technology company. This could be interpreted as job performance being higher in the technology company, average in the manufacturing company, and low in the farming company.

The next step in the analysis is to add a new latent variable to the model, which we refer to as "CO", with the new indicator "c2n_CO" as its sole indicator. This new latent variable is added to the model pointing at job performance (JP). We then run the analysis again. Figure 4 summarizes the results for our illustrative model.

The results control for multilevel effects via the latent variable CO, which quantifies the categorical variable that stores information about company type membership for the individuals from whom data was collected. Note that the path coefficient for the link CO > JP is small (with

a value of -0.06) and statistically non-significant. This means that the multilevel effects do not have a sizeable effect on the path coefficients for the links TU > JP, ED > JP, and PS > JP.

Figure 4: Results controlling for multilevel effects
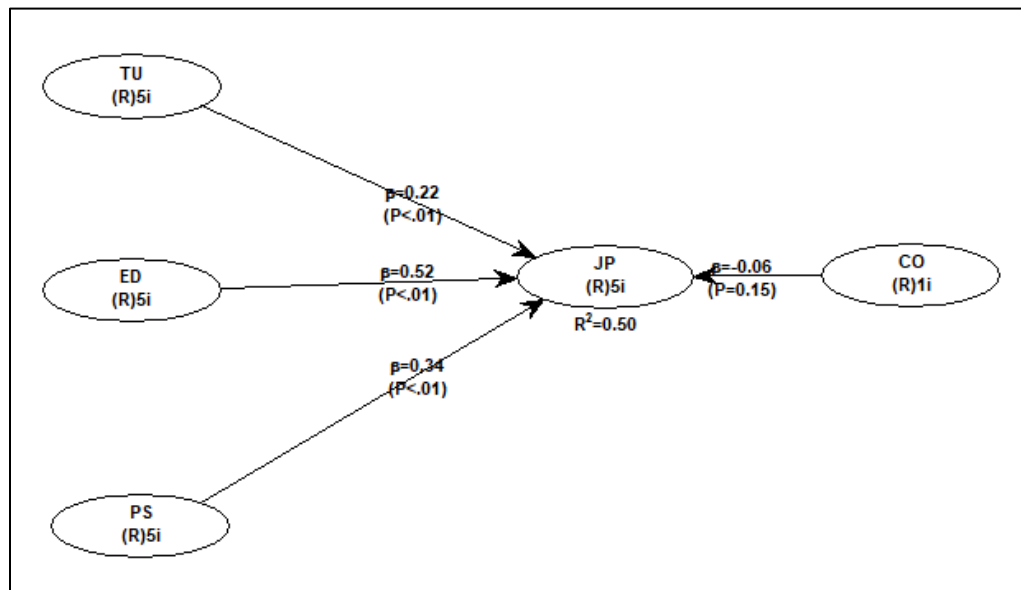
Conclusion

In this paper we illustrate a multilevel analysis with one categorical variable and one endogenous variable. The same procedure should be conducted in models with more than one categorical or endogenous variable, where expectations exist that each categorial variable influences one or more endogenous variables. These expectations should be based on theory and past research.

Acknowledgments

The author is the developer of WarpPLS, which has over 7,000 users in more than 33 countries at the time of this writing. He is grateful to those users for questions, comments, and discussions on topics related to the use of WarpPLS.

References

Kock, N. (2016). Non-normality propagation among latent variables and indicators in PLS-SEM simulations. *Journal of Modern Applied Statistical Methods*, 15(1), 299-315.

Kock, N. (2020a). *WarpPLS User Manual: Version 7.0*. Laredo, TX: ScriptWarp Systems.

Kock, N. (2020b). Full latent growth and its use in PLS-SEM: Testing moderating relationships. *Data Analysis Perspectives Journal*, 1(1), 1-5.

Kock, N., & Hadaya, P. (2018). Minimum sample size estimation in PLS-SEM: The inverse square root and gamma-exponential methods. *Information Systems Journal*, 28(1), 227–261.

Kock, N., Avison, D., & Malaurent, J. (2017). Positivist information systems action research: Methodological issues. *Journal of Management Information Systems*, 34(3), 754-767.