Advantages of nonlinear over segmentation analyses in path models

Ned Kock

Full reference:

Kock, N. (2016). Advantages of nonlinear over segmentation analyses in path models. *International Journal of e-Collaboration*, 12(4), 1-6.

Abstract

The recent availability of software tools for nonlinear path analyses, such as WarpPLS, enables e-collaboration researchers to take nonlinearity into consideration when estimating coefficients of association among linked variables. Nonlinear path analyses can be applied to models with or without latent variables, and provide advantages over data segmentation analyses, including those employing finite mixture segmentation techniques (a.k.a. FIMIX). The latter assume that data can be successfully segmented into subsamples, which are then analyzed with linear algorithms. Nonlinear analyses employing WarpPLS also allow for the identification of linear segments mirroring underlying nonlinear relationships, but without the need to generate subsamples. We demonstrate the advantages of nonlinear over data segmentation analyses.

Keywords: E-collaboration; nonlinear analysis; data segmentation; path modeling; Monte Carlo simulation.

Introduction

During most of the history of e-collaboration research (Kock, 2005) only linear quantitative analysis techniques have been available for use by empirical researchers. The recent availability of software tools for nonlinear path analyses, such as the software WarpPLS (Kock, 2010; 2013; 2015), enables e-collaboration researchers to take nonlinearity into consideration when estimating coefficients of association among linked variables.

Taking nonlinearity into consideration sometimes leads to results that are markedly different from the corresponding linear results, particularly when the underlying relationships take the form of U-curves (Kock, 2010; Kock & Gaskins, 2014). In such cases, the inclinations of the best-fitting lines for relationships associated with pairs of generic variables X → Y change from negative to neutral to positive for different values of X. That is, for different values of X, the path coefficients for the generic links X → Y assume different values, with markedly different interpretations.

Nonlinear path analyses can be applied to models with or without latent variables (Kock, 2014; Kock & Gaskins, 2014; Kock & Lynn, 2012; Kock & Mayfield, 2015; Kock & Verville, 2012), and provide advantages over data segmentation analyses, including those employing finite mixture segmentation techniques (a.k.a. FIMIX). The latter assume that data can be successfully segmented into subsamples, which are then analyzed with linear algorithms.

Nonlinear analyses employing the software WarpPLS also allow for the identification of linear segments emerging from a nonlinear analysis, but without the need to generate subsamples. In this paper we demonstrate the advantages of nonlinear over data segmentation analyses. These include a larger overall sample size for calculation of P values, and the ability to uncover very high segment-specific path coefficients.

Illustrative model and data

Our illustrative model contains only two variables, e-collaboration technology use (ETU) and new product quality (NPQ), and one causal link: ETU → NPQ. E-collaboration technology use (ETU) measures the extent to which a web-based e-collaboration tool has been used by teams of employees of a multinational company. The fictitious multinational company was assumed to develop and sell consumer products. Each team developed a new product, such as a new brand of toothpaste, whose market success was measured through the variable new product quality (NPQ). Both variables were measured through single indicators.

We employed the Monte Carlo method (Robert & Casella, 2005) to create sample data based on this model. The sample size was 250; meaning 250 rows of data, with each row referring to a new product development team. In addition to the two variables ETU and NPQ, we also created a numeric column and a column of text labels referring to three countries. The data sample was created based on the assumption that it was comprised of three subsamples, each coming from a country where the multinational company conducted operations. This is illustrated in Figure 1, where each of the data points represents a new product development team.

Data segmentation results

Table 1 shows the linear path coefficients and corresponding P values, for each of the three countries. The "View or change data modification settings" option in WarpPLS 5.0 allows users to run their analyses with subsamples defined by a range restriction variable, which we chose to be our numeric column referring to each of the countries by a number: 1 for Country1, 2 for Country2, and 3 for Country3. Using this option we were able to conduct linear analyses for each separate country without having to use different datasets in WarpPLS for each of the countries.

Figure 1: Underlying nonlinear relationship and country-specific patterns

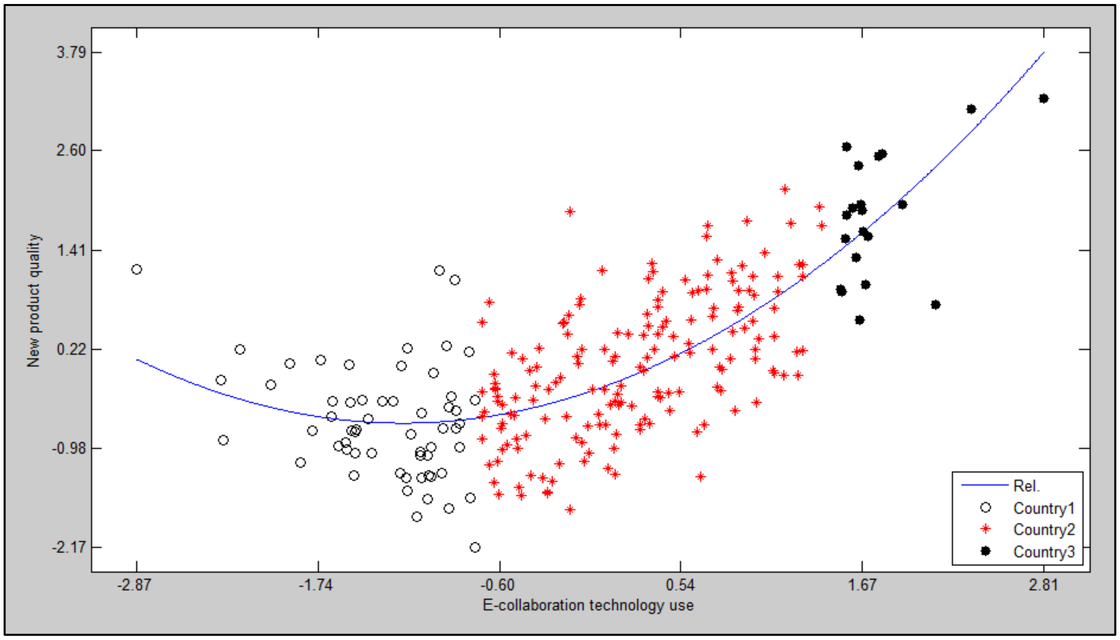

Table 1: Linear path coefficients and P values for each of the three countries

Country	Path coefficient	P value
Country1	-0.30	≤ 0.01
Country2	0.59	≤ 0.01
Country3	0.72	≤ 0.01

For Country1 the path coefficient for the link ETU → NPQ was statistically significant ($P \leq 0.01$) and negative ($\beta = -0.30$). This means that, for the new product development teams in Country1, increases in e-collaboration technology use (ETU) were associated with decreases in new product quality (NPQ). For Country2 the path coefficient for the link ETU → NPQ was statistically significant ($P \leq 0.01$) and positive ($\beta = 0.59$), meaning that in Country2 increases in e-collaboration technology use (ETU) were associated with increases in new product quality (NPQ). For Country3 the path coefficient for the link ETU → NPQ was statistically significant ($P \leq 0.01$) and also positive ($\beta = 0.72$), meaning that in Country3 increases in e-collaboration technology use (ETU) were associated with increases in new product quality (NPQ). The steepness of the positive relationship in Country3 was greater than in Country2.

Nonlinear analysis results

The "View focused relationship graphs with segments" options of WarpPLS 5.0 allow users to view graphs that focus on the best-fitting line or curve, that exclude data points to provide the effect of zooming in on the best-fitting line or curve area, and that show curves as linear segments. We used these options to show linear segments corresponding to each of the three countries, with their respective path coefficients and P values (see Figure 2).

Figure 2: Linear segmentation of nonlinear relationship

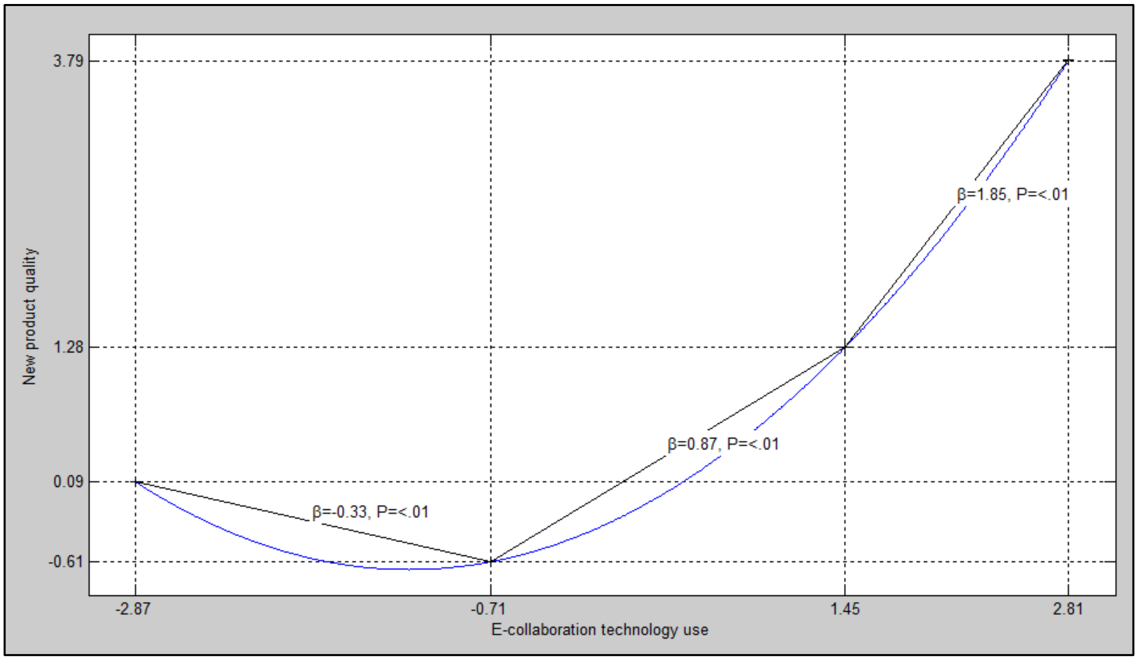

Table 2: Path coefficients and P values for each of the three countries

Country	Path coefficient	P value
Country1	-0.33	≤ 0.01
Country2	0.87	≤ 0.01
Country3	1.85	≤ 0.01

The number of segments shown in a graph depends on the absolute effect segmentation delta chosen by the user through the “Settings” menu option. This absolute effect segmentation delta is the change (or delta) threshold in the first derivative of the nonlinear function depicting the relationship before a new segment is started. For example, a delta of 0.1 means that in each segment the first derivative of the nonlinear function depicting the relationship does not vary more than 0.1.

Since the first derivative does not change in linear relationships, segmentation only occurs in nonlinear relationships. This graph segmentation option allows for the identification of unobserved heterogeneity without a corresponding reduction in sample size, providing an alternative to data segmentation approaches such as finite mixture segmentation (a.k.a. FIMIX).

In our analysis we chose an absolute effect segmentation delta of 1.2, leading to the results summarized in Table 2. This choice of absolute effect segmentation delta made the graph segments, corresponding to the values of e-collaboration technology use (ETU) shown on the graphs horizontal axis, coincide with the country subsamples. For Country1 the path coefficient for the link ETU → NPQ was statistically significant ($P \leq 0.01$) and negative ($\beta = -0.33$). For

Country2 the path coefficient for the link ETU → NPQ was statistically significant ($P \leq 0.01$) and positive ($\beta = 0.87$). For Country3 the path coefficient for the link ETU → NPQ was statistically significant ($P \leq 0.01$) and also positive ($\beta = 1.85$).

As we can see, the graph segmentation approach discussed in this section yielded greater path coefficients than the data segmentation approach discussed in the previous section. This was the case for each of the three countries, and is consistent with the existence of an underlying U-curve relationship for the link ETU → NPQ. In such a relationship path coefficients would be expected to grow quickly in strength as values of ETU grow, which happens as one goes from left to right in the graph. This is the reason for the much stronger path coefficients for the segments referring to Country2 and Country3.

Note that the very high path coefficient for the link ETU → NPQ in Country3 would have been impossible to obtain through a linear analysis focusing only on the corresponding country-specific subsample. The reason for this is that in our simple model with only two variables any path coefficient coincides with the corresponding correlation coefficient, and a correlation coefficient cannot be greater than 1. Nevertheless, we know that the path coefficient is indeed very high ($\beta = 1.85$) for the Country3 segment because we created the data used in the analysis with the underlying relationship having a U-curve shape with a very steep gradient at its right end.

Discussion and concluding remarks

E-collaboration research (Kock, 2005) has a long and distinguished history, in part because of the emergence of the Internet and the corresponding growing importance of e-collaboration in organizations and society over the years. During most of this history, e-collaboration researchers conducting multivariate quantitative analyses have been restricted to the use of linear algorithms in their estimation of coefficients of association.

This situation has changed recently with the availability of user-friendly software tools for nonlinear analysis, such as the software WarpPLS (Kock, 2010; 2013; 2015). E-collaboration researchers, as well as researchers in other fields where multivariate statistical methods are employed, can now account for the existence of underlying nonlinearity when estimating coefficients of association among linked variables.

While the data and graph segmentation approaches discussed in the previous sections may look similar, they differ in at least one critical aspect. The graph segmentation approach, which is the type of segmentation enabled by the "View focused relationship graphs with segments" options of WarpPLS 5.0, does not involve splitting the sample into subsamples. The data segmentation approach, on the other hand, *does* involve splitting the sample into subsamples.

In this paper we demonstrated, employing a simulated dataset created via the Monte Carlo method, the advantages of nonlinear over data segmentation analyses. Particularly noteworthy are two advantages. The first is a larger overall sample size for calculation of P values, which enables researchers to uncover actual segment-specific effects that could otherwise be rendered non-significant due to a combination of underestimated path coefficients and small subsample sizes. The second advantage is the ability to uncover very high segment-specific path coefficients, which could otherwise be grossly underestimated.

Acknowledgments

The author is the developer of the software WarpPLS, which has over 7,000 users in more than 33 different countries at the time of this writing, and moderator of the PLS-SEM e-mail distribution list. He is grateful to those users, and to the members of the PLS-SEM e-mail distribution list, for questions, comments, and discussions on topics related to the use of WarpPLS.

References

Kock, N. (2005). What is e-collaboration. *International Journal of e-Collaboration*, 1(1), 1-7.

Kock, N. (2010). Using WarpPLS in e-collaboration studies: An overview of five main analysis steps. *International Journal of e-Collaboration*, 6(4), 1-11.

Kock, N. (2013). Using WarpPLS in e-collaboration studies: What if I have only one group and one condition? *International Journal of e-Collaboration*, 9(3), 1-12.

Kock, N. (2014). Advanced mediating effects tests, multi-group analyses, and measurement model assessments in PLS-based SEM. *International Journal of e-Collaboration*, 10(3), 1-13.

Kock, N. (2015). *WarpPLS 5.0 User Manual*. Laredo, TX: ScriptWarp Systems.

Kock, N., & Gaskins, L. (2014). The mediating role of voice and accountability in the relationship between Internet diffusion and government corruption in Latin America and Sub-Saharan Africa. *Information Technology for Development*, 20(1), 23-43.

Kock, N., & Lynn, G.S. (2012). Lateral collinearity and misleading results in variance-based SEM: An illustration and recommendations. *Journal of the Association for Information Systems*, 13(7), 546-580.

Kock, N., & Mayfield, M. (2015). PLS-based SEM algorithms: The good neighbor assumption, collinearity, and nonlinearity. *Information Management and Business Review*, 7(2), 113-130.

Kock, N., & Verville, J. (2012). Exploring free questionnaire data with anchor variables: An illustration based on a study of IT in healthcare. *International Journal of Healthcare Information Systems and Informatics*, 7(1), 46-63.

Robert, C.P., & Casella, G. (2005). *Monte Carlo statistical methods*. New York, NY: Springer.