A comparison of multiple regression analyses in Stata and WarpPLS

Augustine Tarkom

Texas State University, USA

Prash Gopal

Texas A&M International University, USA

Abstract

This paper illustrates a comparative analysis of multiple regression analysis using two different software. The software packages used are WarpPLS 8.0 and Stata 17. Multiple regression analyses performed with both software produce the same results. WarpPLS 8.0 has the added advantage over Stata owing to its graphic user interface that aids in model specification and visualization. Furthermore, it provides users with additional tools to visualize moderating effects. Both software have equal accuracy in terms of the results but differences in terms of what they offer users.

Keywords: Multiple Regression, Moderating Relationship, WarpPLS, Stata

Introduction

Regression analysis is a basic statistical technique used in various disciplines to investigate correlations between variables. This paper aims to provide a comprehensive comparative analysis of multiple regression analysis performance using two widely utilized software packages: WarpPLS and Stata. WarpPLS is an established software known for its robust variance-based and factor-based structural equation modeling (SEM) capabilities, which rely on partial least squares (PLS) and factor-based methods (Canatay et al., 2022). Although WarpPLS is primarily utilized for SEM (Amora, 2021; Morrow & Conger, 2021; Rasoolimanesh, 2022; Amora, 2023), it also offers robust regression analysis tools. This versatility makes WarpPLS an attractive option for researchers looking to seamlessly integrate SEM and regression analysis. While SEM using the PLS method is still employed today in a broad range of fields such as information systems, health, and organizational leadership (Kock, 2022a, 2014; Kock et al., 2019), the WarPLS software, which offers users a wide range of features, many of which are absent in other SEM software, has been inadequately utilized vis-à-vis its regression abilities (Kock, 2023, 2022b). Stata, a prominent software used in finance, economics, and sociology, is renowned for its wide-ranging regression modeling features (Sauerbrei et al., 2006). By

juxtaposing these two software packages, this paper highlights their respective similarities, strengths, ease of use, visualization capabilities, and the accuracy of the results obtained. Additionally, this comparison seeks to illuminate the often-overlooked regression capabilities of WarpPLS, a software primarily known for its structural equation modeling prowess, while contrasting these features with Stata's well-established regression tools. By doing so, we aim to broaden researchers' perspectives on the potential applications of WarpPLS beyond its typical use, particularly in the realm of multiple regression analysis.

Illustrative model

In WarpPLS 8.0 (Kock, 2020; Kock, 2022a), users can specify the regression model visually through a user-friendly graphical interface. This feature allows an intuitive understanding of the relationships between predictor and criterion variables, depicted by directional arrows (Figure 1). The software presents the analysis results in a diagrammatic and tabular form, thus facilitating straightforward interpretation and information assimilation. Robust inference is ensured by WarpPLS, which reports the regression coefficients' statistical significance at the $p < 0.001$ level (Figure 2).

Figure 1: From the left, the base model displaying conceptual links between the variables followed by the results of the analysis performed using WarpPLS
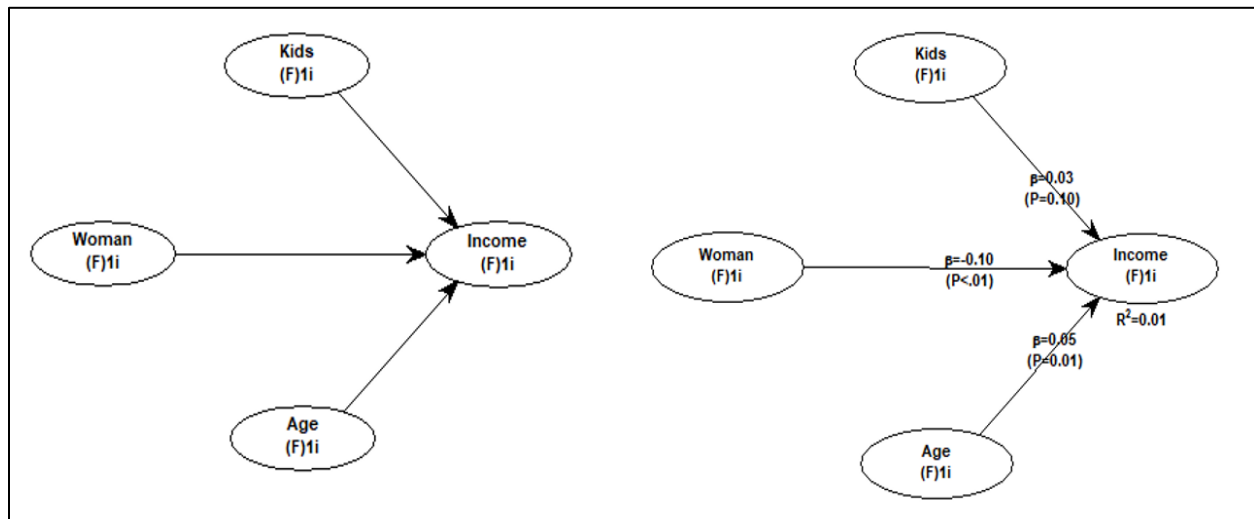

Figure 2: The regression coefficients and their P-values

Path coefficients

	Income	Woman	Age	Kids
Income		-0.101	0.053	0.030
Woman				
Age				
Kids				

P values

	Income	Woman	Age	Kids
Income		<0.001	0.013	0.101
Woman				
Age				
Kids				

On the other hand, Stata requires users to define the regression model using code. While this approach may not be as visually pleasing, it offers advanced users greater flexibility and

customization options. Stata presents the regression results in a tabular format (Figure 3), providing unstandardized and standardized coefficients (Jann ETH Zurich, 2005).

Both software packages yield consistent results for the base model, with negligible differences attributable to rounding standards. WarpPLS's graphical interface and diagrammatic representation of results can provide a more accessible and intuitive approach for users with less coding experience or those seeking a more visual grasp of the model.

Figure 3: The unstandardized regression results for the base model (above), and the standardized regression results for the base model (below)

```
. reg Income Woman Kids Age

      Source │       SS           df       MS            Number of obs   =      1,804
─────────────┼──────────────────────────────────         F(3, 1800)      =       9.00
       Model │   2.2958e+10          3  7.6526e+09        Prob > F        =     0.0000
    Residual │   1.5307e+12      1,800   850409499        R-squared       =     0.0148
─────────────┼──────────────────────────────────         Adj R-squared   =     0.0131
       Total │   1.5537e+12      1,803   861727621        Root MSE        =      29162


      Income │ Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
─────────────┼────────────────────────────────────────────────────────────────
       Woman │  -5958.239   1385.798    -4.30   0.000    -8676.18   -3240.298
        Kids │   1968.718   1643.915     1.20   0.231   -1255.464    5192.901
         Age │   90.85233   43.24067     2.10   0.036    6.045146    175.6595
       _cons │   28401.53    2217.73    12.81   0.000    24051.93    32751.13


.
end of do-file
```

```
. reg stdInc stdWom stdKids stdAge

      Source │       SS           df       MS            Number of obs   =      1,804
─────────────┼──────────────────────────────────         F(3, 1800)      =       9.00
       Model │   26.6415993         3  8.88053311        Prob > F        =     0.0000
    Residual │   1776.35841     1,800  .986865786        R-squared       =     0.0148
─────────────┼──────────────────────────────────         Adj R-squared   =     0.0131
       Total │   1803.00001     1,803  1.00000001        Root MSE        =      .99341


      stdInc │ Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
─────────────┼────────────────────────────────────────────────────────────────
      stdWom │  -.1009606   .0234819    -4.30   0.000    -.1470154   -.0549059
     stdKids │   .0300266   .0250728     1.20   0.231    -.0191482    .0792014
      stdAge │    .052564   .0250176     2.10   0.036     .0034975    .1016305
       _cons │   2.31e-09    .023389     0.00   1.000    -.0458723    .0458724


.
end of do-file
```

Testing moderating effects

WarpPLS allows users to specify moderating effects directly through the graphical interface (Kock, 2020; Hubona & Belkhamza, 2021), further enhancing the visual representation of the model (Figure 4). It also offers a range of visualization tools to explore the moderating effects, including 2D and 3D graphs in standardized and unstandardized scales (Figures 5, 6, and 7). The rocky (Figure 6) and smooth (Figure 7) formats are both generated through Delauny triangulations, except that the latter pictorially resembles a bed sheet. These visuals provide valuable insights into the nature and strength of the moderating relationships.

In Stata, users must specify moderating effects through code (Figure 8). While this approach may be more flexible for advanced users, it may present a steeper learning curve for those less familiar with coding. Stata generates a 2D graph to visualize the moderating effect, albeit with standardized values only (Figure 9).

Both software packages produce consistent results for the moderating effects analysis, with any minor discrepancies likely attributable to rounding conventions or visualization techniques. Regarding the results, both software have very little to separate them; any difference in usage could be attributed to the preference of users belonging to various fields. WarpPLS's graphical interface and diagrammatic representation of moderating links can provide more conceptual clarity for users seeking a more visual grasp of the mechanism.

Figure 4: From the left, the base model displaying conceptual links between the variables, followed by the moderating effect performed using WarpPLS
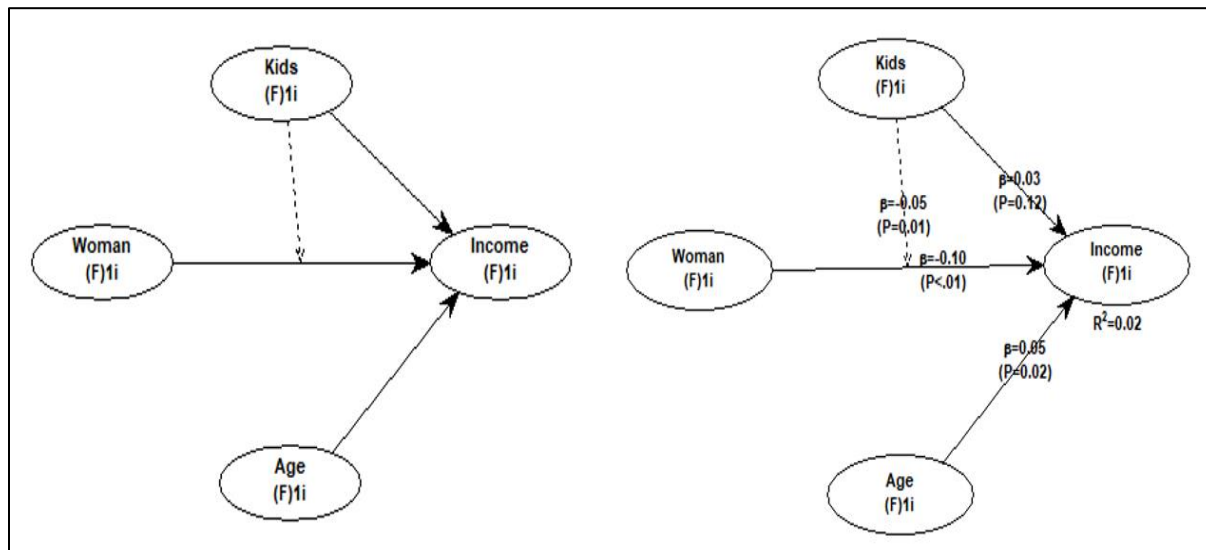
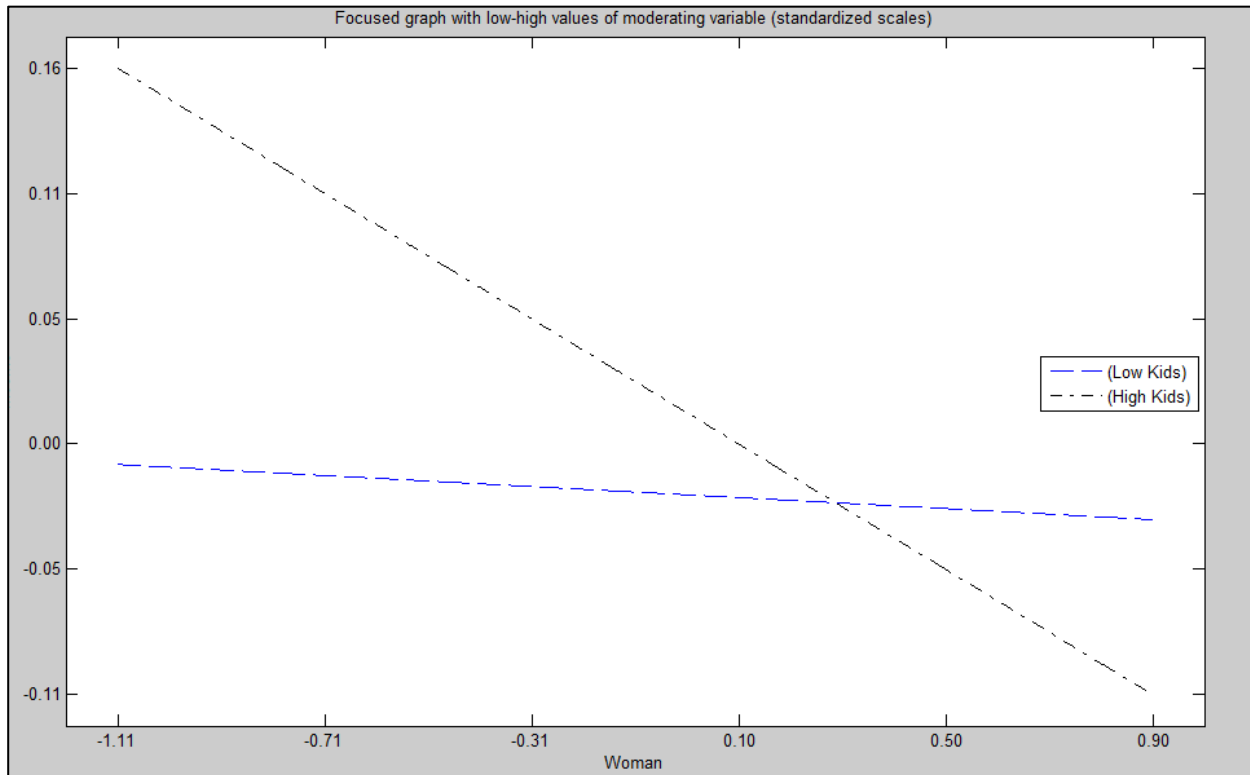**Figure 5: The 2D graph displaying the moderating effect (WarpPLS)**

Figure 6: The rocky 3D graph displaying the moderating effect
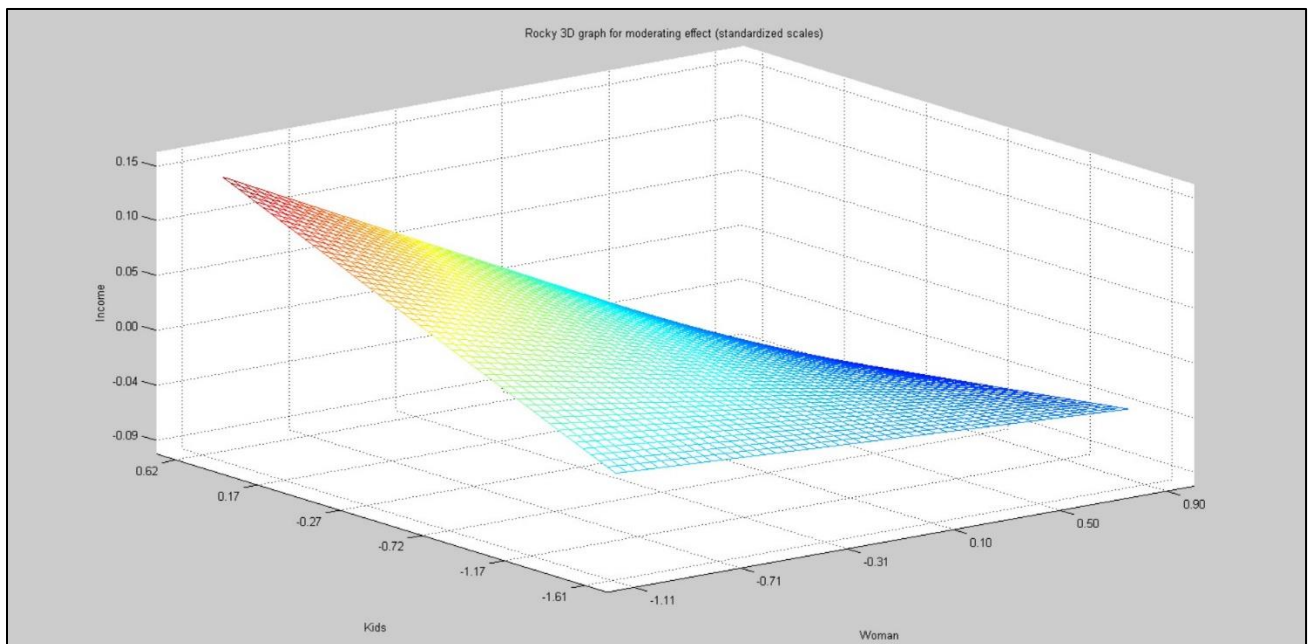
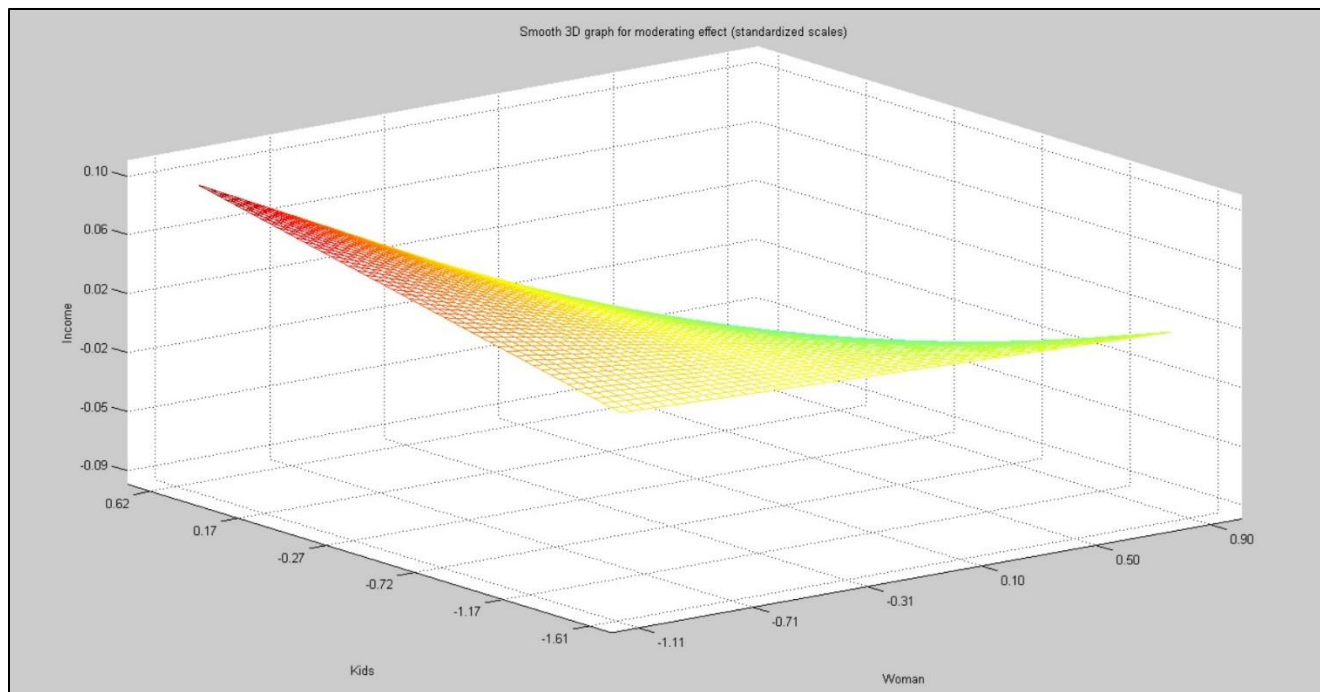**Figure 7: The smooth 3D graph displaying the moderating effect**

Figure 8: The standardized regression coefficients for the model with the moderating effect

```
. reg stdInc c.stdWom##c.stdKids stdAge

      Source |       SS           df       MS      Number of obs   =      1,804
-------------+----------------------------------   F(4, 1799)      =       8.00
       Model |  31.5169569         4  7.87923922   Prob > F        =     0.0000
    Residual |  1771.48306     1,799  .984704312   R-squared       =     0.0175
-------------+----------------------------------   Adj R-squared   =     0.0153
       Total |  1803.00001     1,803  1.00000001   Root MSE        =     .99232


      stdInc | Coefficient  Std. err.      t    P>|t|     [95% conf. interval]
-------------+----------------------------------------------------------------
      stdWom |  -.1016346   .0234582     -4.33   0.000    -.1476427    -.0556265
     stdKids |   .0276815   .0250674      1.10   0.270    -.0214829     .0768459
             |
   c.stdWom# |
   c.stdKids |  -.0519931   .0233666     -2.23   0.026    -.0978216    -.0061646
             |
      stdAge |    .049079   .0250392      1.96   0.050    -.0000299     .0981879
       _cons |   .0036305   .0234202      0.16   0.877    -.0423032     .0495642


.
end of do-file
```
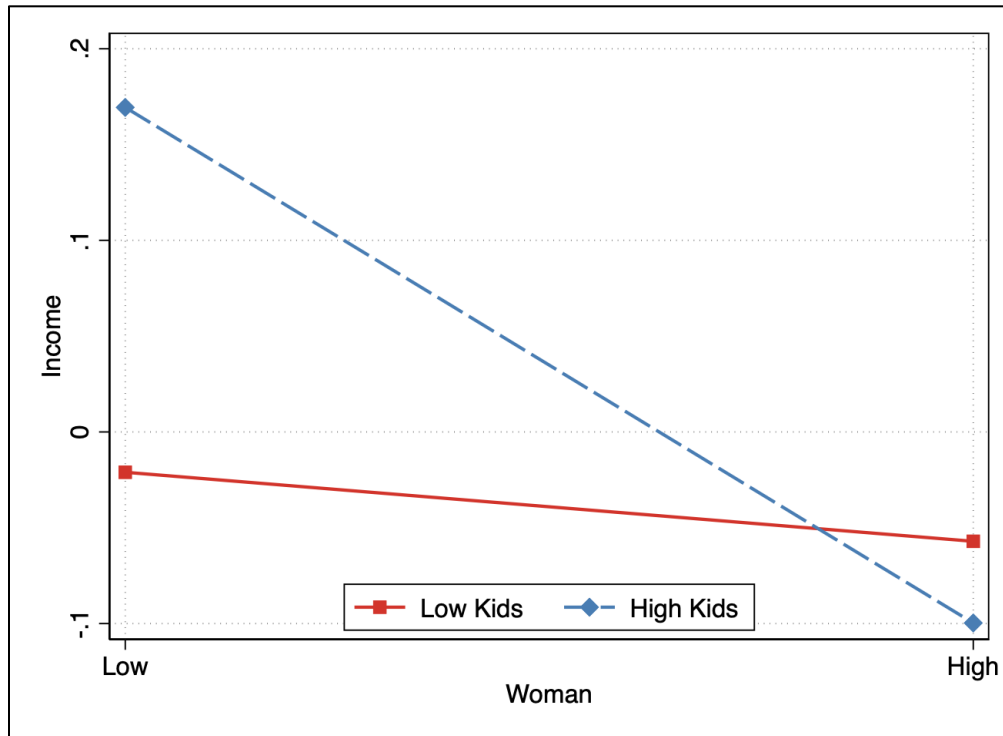
Figure 9: The 2D graph displaying the moderating effect (Stata)

Conclusion

In conclusion, both WarpPLS and Stata are powerful software packages capable of performing multiple regression analyses with high accuracy. The choice between the two, however, may depend on the specific needs and preferences of the user. WarpPLS excels in its user-friendly graphical interface, intuitive model specification, and comprehensive visualization tools, making it an attractive option for researchers seeking a more visual and interactive approach to regression analysis. Its ability to visualize moderating effects through 2D and 3D graphs in standardized and unstandardized scales is a notable strength. Moreover, WarpPLS's core strength lies in its robust variance-based and factor-based structural equation modeling capabilities, which rely on partial least squares (PLS) and factor-based methods. On the other hand, Stata offers greater flexibility and customization through its code-based approach, catering to advanced users of statistical analysis with specialized requirements. Its tabular output provides a more detailed view of the regression coefficients. Ultimately, both software packages generate multiple regression analysis results that are consistent and reliable for multiple regression analysis, and their respective strengths align with various user preferences and research objectives.

References

Amora, J. T. (2021). Convergent validity assessment in PLS-SEM: A loadings-driven approach. *Data Analysis Perspectives Journal*, *2*(3), 1-6.

Amora, J. T. (2023). On the validity assessment of formative measurement models in PLS-SEM. *Data Analysis Perspectives Journal*, *4*(2), 1-7.

Canatay, A., Emegwa, T., Lybolt, L. M., & Loch, K. D. (2022). Reliability assessment in SEM models with composites and factors: A modern perspective. *Data Analysis Perspectives Journal*, *3*(1), 1-6.

Hubona, G., & Belkhamza, Z. (2021). Testing a moderated mediation in PLS-SEM: A full latent growth approach. *Data Analysis Perspectives Journal*, *2*(4), 1-5.

Jann, B. (2005). Making regression tables from stored estimates. *The Stata Journal*, *5*(3), 288-308.

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

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

Kock, N. (2022a). *WarpPLS User Manual: Version 8.0*. Laredo, TX: ScriptWarp Systems.

Kock, N. (2022b). Testing and controlling for endogeneity in PLS-SEM with stochastic instrumental variables. *Data Analysis Perspectives Journal*, *3*(3), 1-6.

Kock, N. (2023). Using logistic regression in PLS-SEM: Dichotomous endogenous variables. *Data Analysis Perspectives Journal*, *4*(4), 1-6.

Kock, N., Mayfield, M., Mayfield, J., Sexton, S., & De La Garza, L. M. (2019). Empathetic leadership: How leader emotional support and understanding influences follower performance. *Journal of Leadership & Organizational Studies*, *26*(2), 217-236.

Morrow, D. L., & Conger, S. (2021). Assessing reciprocal relationships in PLS-SEM: An illustration based on a job crafting study. *Data Analysis Perspectives Journal*, *2*(1), 1-5.

Rasoolimanesh, S. M. (2022). Discriminant validity assessment in PLS-SEM: A comprehensive composite-based approach. *Data Analysis Perspectives Journal*, *3*(2), 1-8.

Sauerbrei, W., Meier-Hirmer, C., Benner, A., & Royston, P. (2006). Multivariable regression model building by using fractional polynomials: Description of SAS, STATA and R programs. *Computational Statistics & Data Analysis*, *50*(12), 3464-3485.