Journal of Marketing Analytics
https://doi.org/10.1057/s41270-026-00466-2

SOFTWARE REVIEW

Multiple linear and logistic regression analysis: a SmartPLS 4 software tutorial

Vasilica-Maria Margalina[1] · **Charlotte Kreienbaum**[2] · **Joseph F. Hair**[3] · **Jan-Michael Becker**[4] · **Christian M. Ringle**[2,5]

Received: 9 February 2026 / Accepted: 10 February 2026

Abstract
This tutorial and case study provides a comprehensive, step-by-step guide to conducting multiple and logistic regression analyses using SmartPLS 4. Although SmartPLS is primarily known for partial least squares structural equation modeling (PLS-SEM), its latest version includes features that enable researchers to perform and visualize regression analyses effectively. The tutorial introduces the theoretical foundations of both standard multiple and logistic regression, outlines the main stages of model specification and estimation, and demonstrates how to implement these analyses within the SmartPLS environment. Emphasis is placed on key analytical decisions, such as model design, assumption testing, interpretation of coefficients, and goodness-of-fit measures. Practical examples and graphical outputs are included to illustrate the implementation process and to enhance understanding of the results.

Keywords Multiple linear regression · Logistic regression · Analysis · Evaluation · Application · SmartPLS

Introduction

Regression methods have become an integral component of many data analyses concerned with exploring the relationships between dependent and independent variables. Regression analysis is used in both academia and organizational settings for (1) calculating if one or more independent variables have a significant relationship with a single dependent variable, (2) estimating the relative contribution of different independent variables in explaining the variance of the dependent variable, and (3) making predictions (Sarstedt and Mooi 2019).

There are several regression analysis techniques that can be used to analyze the relationships between a single dependent variable and several independent variables. Multiple linear regression is the most widely used multivariate technique for both explanation and prediction (Hair et al. 2019). Prediction refers to the extent to which a set of independent variables can predict the values of a dependent variable. Explanation examines the strength of the effects (regression coefficients) of each independent variable on the dependent variable to understand their (relative) importance.

There are many statistical software applications that have incorporated multiple and logistic regression, such as SPSS and STATA. In these statistical software applications, users

✉ Christian M. Ringle
c.ringle@tuhh.de

Vasilica-Maria Margalina
vasilica.margalina@uab.ro

Charlotte Kreienbaum
charlotte.kreienbaum@tuhh.de

Joseph F. Hair
jhair@southalabama.edu

Jan-Michael Becker
jan-michael.becker@bi.no

1 Faculty of Economics, University "1 Decembrie 1918" of Alba Iulia, Alba Iulia, Romania

2 Institute of Management and Decision Sciences, Hamburg University of Technology, Am Schwarzenberg-Campus 4 (D), 21073 Hamburg, Germany

3 Cleverdon Chair of Business, Mitchell College of Business, University of South Alabama, Mobile, AL, USA

4 Department of Marketing, BI Norwegian Business School, Oslo, Norway

5 College of Business, Law and Governance, James Cook University, Townsville, Australia

Published online: 02 March 2026

usually need to navigate through different windows or write several syntaxes to obtain all the results as a table or graph, which is needed for regression analysis. This aspect of their graphical user interface (GUI) has changed very little in the last 30 years. Therefore, their GUI and usability do not fully align with current analytical software standards. In contrast, the statistical software SmartPLS 4 (https://www.smartpls.com/; Ringle et al. 2024), which was originally developed for partial least squares structural equation analysis (PLS-SEM), now includes a regression module with a modern GUI and other methods, such as necessary condition analysis (NCA; Becker et al. 2026; Dul 2016; Richter et al. 2020), importance-performance map analysis (IPMA; Ringle and Sarstedt 2016), path analysis and PROCESS analysis (Hayes 2022; Sarstedt et al. 2020), covariance based structural equation modeling (CB-SEM; Jöreskog 1978; Hair et al. 2025), and generalized structured component analysis (GSCA; Hwang and Takane 2004).

Prior reviews of the SmartPLS software highlight the software's intuitive design and ease of use (Sarstedt and Cheah 2019; Cheah et al. 2024; Hair et al. 2025). Several textbooks (e.g., Hair et al. 2026, 2024; Chua 2024), articles (Matthews et al. 2016; Sarstedt et al. 2024; Merkle 2025), and software reviews (Memon et al. 2021; Cheah et al. 2024; Sarstedt et al. 2024; Hair et al. 2025) have emphasized the potentials of the SmartPLS software and provided guidance on its use for a wide range of statistical techniques. However, the application of the SmartPLS software for regression analyses has not yet been documented.

Considering the above, this tutorial article provides guidance on how to conduct a regression analysis using the SmartPLS software. To do so, we draw on the case studies of multiple and logistic regression analysis used in Hair et al. (2019; i.e., Chap. 5 and Chap. 8), which is one of the most widely used textbooks on multivariate analyses in social sciences (e.g., Black and Babin 2019). The aim of our *Software Review* article is to help researchers reliably apply regression analyses by facilitating the use of several measures for assessing and interpreting the results.

In the following sections, we describe the case study and the methods used for regression model estimation, followed by a step-by-step description of multiple and logistic regression model estimation, and the assessment of results using SmartPLS. This software tutorial article concludes with additional observations and SmartPLS software extensions, which will be available in the near future.

Table 1 Multiple linear regression model (Hair et al. 2019)

Dependent variable	Independent variables	
Customer satisfaction (X_{19})	Product Quality	X_6
	E-Commerce Activities/Website	X_7
	Technical Support	X_8
	Complaint Resolution	X_9
	Advertising	X_{10}
	Product Line	X_{11}
	Salesforce Image	X_{12}
	Competitive Pricing	X_{13}
	Warranty and Claims	X_{14}
	New Products	X_{15}
	Ordering and Billing	X_{16}
	Price Flexibility	X_{17}
	Delivery Speed	X_{18}

Data and regression models

The case study used for the tutorial is based on a market segmentation study executed by HBAT, an actual company which manufactures and sells paper products, as described in Hair et al. (2019). The sample size is $N = 100$ and the dataset is disguised but represents actual product marketing situations. As displayed in Table A1 (in Appendix), information about perceptions of HBAT was collected for data warehouse classification variables (X_1-X_5; i.e., core firm characteristics and their association with HBAT), actual perceptions of HBAT (X_6-X_{18}; i.e., each respondent's perceptions of HBAT on a set of business functions), and outcome variables (X_{19}-X_{23}) reflecting respondent's purchase relationships with HBAT. The perception variables and most of the outcome variables are measured using a 0–10-point scale (i.e., with 0 = "Poor" and 10 = "Excellent"). The outcome variables can be used as the dependent variables in a linear multiple regression model example (Hair et al. 2019). For this example, we use customer satisfaction (X_{19}) as the dependent variable and the independent HBAT perceptions variables are shown in Table 1.

For the logistic regression model, we need a binary dependent variable. The dataset includes several binary warehouse classification variables that can be used as outcome variables (X_1-X_5; Table A1; in the Appendix). Following Hair et al. (2019), the aim of the logistic regression in this case study is for HBAT's team to assess the relative influence of each HBAT's performance perception variable on the observed differences between two customer groups. Like Hair et al. (2019), we use the binary variable Region (X_4) as the dependent variable in the logistic regression model example (Table 2). This variable classifies customer location into two categories, USA/North America (0) and outside North America (1). The independent variables are the perceptions of HBAT (Table 2), same as for the multiple linear regression.

Table 2 Logistic regression model (Hair et al. 2019)

Dependent variable	Independent variables	
Region (X_4)	Product Quality	X_6
	E-Commerce Activities/Website	X_7
	Technical Support	X_8
	Complaint Resolution	X_9
	Advertising	X_{10}
	Product Line	X_{11}
	Salesforce Image	X_{12}
	Competitive Pricing	X_{13}
	Warranty and Claims	X_{14}
	New Products	X_{15}
	Ordering and Billing	X_{16}
	Price Flexibility	X_{17}
	Delivery Speed	X_{18}

Multiple linear regression

The objective of the multiple linear regression analysis is to explain a single metrically measured dependent variable (criterion/outcome) using several independent variables (predictors/regressors). The dependent variable (DV) in a multiple linear regression model should typically be a metrically measured variable. In contrast, the independent variables (IVs) can be both metric and discrete variables. However, discrete variables need to be transformed into numerical contrasts (e.g., dummy coding, effect-coding, Helmert coding, etc.) before they can be included in the model. The most common is dummy coding where the non-metric (discrete) variables are transformed into a binary variable with two categories represented as 0 and 1.

For the data used in our case study, the dependent variable is customer satisfaction (a metric variable) and the independent variables are HBAT's performance perceptions (metric variables; Table 1). As shown in Table 1, there are 13 possible independent variables that could be included in the regression model. When selecting the variables for a multiple regression, the objective is always to identify the best regression model (Hair et al. 2019). Researchers should consider three key issues when selecting their variables: theoretical justification, measurement error, and specification error (not including a meaningful IV). Because omitting relevant variables can bias results, it is generally safer to include a potentially irrelevant variable than to exclude an important one. We focus on the application of regression to a theoretically established model. *Note*: The use of automated, algorithm-based variable selection has decreased because of the criticisms regarding its atheoretical nature and the lack of considerations of factors, such as multicollinearity, the presence of outliers and influential data, and the interpretability of results (Hair et al. 2019). Therefore, this tutorial focuses on applying regression to a theoretically established model and does not further consider algorithm-based variable selection methods.

For the estimation of the regression coefficients, there are multiple approaches. For example, the maximum likelihood estimation (ML) or the ordinary least squares estimation (OLS) are both widely used in social sciences research (Sarstedt and Mooi 2019). This tutorial focusses on multiple linear regression analysis using the OLS estimation method. The OLS estimation minimizes the sum of squared differences between the actual and predicted values of the dependent variable. To do so, we examine the vertical deviations between the predicted outcome ($\hat{y}_i$) and the actual outcome value (y_i), with the goal of minimizing the error in predicting the dependent variable Y. By squaring these differences between the actual and predicted values, OLS accounts for both negative and positive deviations from the regression line.

The estimated regression coefficients are then used for explanatory purposes as they indicate the effect of the independent variables on the dependent variable (Hair et al. 2019). The coefficients directly obtained from the OLS procedure are also known as unstandardized coefficients and indicate the expected change of the dependent variable when the independent variable increases by one unit (Sarstedt and Mooi 2019). However, differences in the measurement scales of the independent variables, make the direct comparison of the unstandardized coefficients often impossible. By standardizing the regression coefficients, a common scale and variability is obtained, which usually consists in a mean of zero and a standard deviation of one (Hair et al. 2019). The absolute value of these standardized coefficients, also referred to as beta (β) coefficients, reflect the relative impact of a one-standard-deviation change in the independent variable on the dependent variable. Thus, standardized coefficients allow determining the independent variable that has the strongest relationship with the dependent variable.

In multiple linear regression, positive or negative signs of the coefficients indicate the direction of the relationships. A positive coefficient indicates that an increase in the independent variable is related to an increase in the predicted probabilities and vice versa for negative coefficients. In multiple linear regression a coefficient with a value of 0 indicates that the independent variable has no impact on the dependent variable. The intercept or constant of a linear regression model indicates the expected amount of the dependent variable when all independent variables are zero.

Logistic regression

The logistic regression model, in turn, works similarly to regular regression models, with the difference that the dependent variable is binary (dichotomous; LaValley 2008). Similar to multiple linear regression, the logistic regression estimates the relationship between the independent and dependent variables that best fit the data. Logistic regression essentially aims at predicting the probability of the dependent variable being 1 given the independent variables, which can be formally expressed by $\mathrm{P(Y = 1|X)}$. However, the estimation process has some differences due to the binary nature of the dependent variable. Unlike linear regression that aims at predicting metric outcomes, the predicted values in logistic regression (i.e., probabilities) should never be outside the range of 0 and 1.[1] For keeping the predicted values within the range of 0 and 1, the S-shaped (sigmoid) form of the logistic curve is used for the logistic transformation of the dependent variable (Fig. 1).

The logistic transformation process is performed in two basic steps. In the first step we have a linear model in which the log-odds (or logit) are linearly related to the predictor variables. The odds are the ratio of the probabilities of the two outcomes or events.

$$Odds = \frac{P(Y=1)}{P(Y=0)} = \frac{P(Y=1)}{1-P(Y=1)}$$

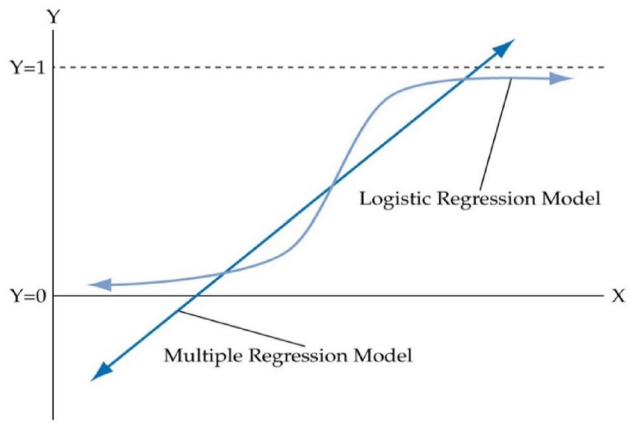


Fig. 1 Multiple linear regression and logistic regression curve (Hair et al. 2019; Chap. 8)

[1] In principle, it is also possible to use a multiple linear regression on binary outcomes. This approach is called a linear probability model. It suffers from the key limitation that the predicted values can be outside the 0–1 interval and therefore represent inadmissible outcomes (e.g., predicted probabilities of -20%). However, the literature also discusses several advantages of these linear probability models such as easy interpretation of the regression coefficients as percentage point change that the dependent variable equals 1 (e.g., Lee et al., 2025).

The natural logarithm of the odds (also called logit) is a metric outcome that can range from negative to positive infinity.

$$\ln(Odds) = \beta_0 + \beta_1 X_1 + \ldots + \beta_p X_p$$

In the second step these linear models can be transformed into a probability, which adheres to the threshold ranges of 0 and 1. The probability of the outcome being 1 (our focal prediction) given the predictor variable can therefore be expressed as

$$P(Y=1|X) = \frac{Odds}{1+Odds} = \frac{e^{\beta_0+\beta_1 X_1+\ldots+\beta_p X_p}}{1+e^{\beta_0+\beta_1 X_1+\ldots+\beta_p X_p}}$$

where $\frac{e^x}{1+e^x}$ is the sigmoid function. Because logistic regression employs this logistic transformation of the dependent variable, the interpretation of its coefficients differs from standard linear regression. Specifically, the coefficients represent the change in the log-odds (logit) for every unit increase in the independent variable. As these coefficients do not directly represent changes in the probability (but changes in the log-odds, which are hard to interpret), researchers often transform these values into odds ratios using the exponential function e^{β} to make these results more intuitive (Hair et al. 2019). Because these values are exponents, they are interpreted slightly different. Their impact is multiplicative, meaning that the coefficient's effect is not added to the dependent variable (the odds) but multiplied for each unit change in the independent variable.

Thus, a logistic regression coefficient of 0 corresponds to the odds ratio of 1, meaning that the odds will not change when changing the independent variable (they are multiplied with 1 and thus stay the same). Unlike the original coefficients, where the sign indicates direction, the odds ratio is therefore interpreted relative to 1: values greater than 1 reflect an increase in odds, while values between 0 and 1 reflect a decrease. Ultimately, while the two types of coefficients require different interpretations, both are essential for assessing the direction and predictive strength of the relationship. Therefore, even though the interpretation of the two types of coefficients is different, both are used to assess the direction as well as the predictive magnitude of the relationship.

Workflow for running the regression analysis

Hair et al. (2019) recommend following these six stages for both multiple and logistic regression analysis:

- Stage 1: Selecting the objectives of the regression analysis.

- Stage 2: Research design of the regression analysis.
- Stage 3: Testing the assumptions underlying regression analysis.
- Stage 4: Estimating the regression model and assessing overall fit.
- Stage 5: Interpreting the regression variate.
- Stage 6: Validating the results.

Stage 1 focuses on selecting the objective of the regression analysis, which may be explanation, prediction, or both. In this stage, the researcher should select the dependent and independent variables and specify the type of regression model (i.e., linear vs. logistic) based on the underlying theoretical model. Then, in Stage 2, the researcher must design the regression analysis by considering the sample size and the measurement of the variables. In multiple linear regression analysis, sample size affects the statistical power of significance testing and the generalizability of the results. In this case, logistic regression differs from multiple linear regression because it relies on maximum likelihood estimation (ML), which generally requires larger samples. As a result, logistic regression needs more observations than multiple regression. Although Hosmer and Lemeshow (2000) suggest sample sizes above 400, smaller samples can still work in practice. Researchers should also consider splitting the data into analysis and holdout samples to validate the model. However, the analysis portion must still meet the sample size requirements, which increases the total sample needed depending on the model's complexity (Hair et al. 2019).

In Stage 3, the researcher must examine the assumptions underlying the regression analysis. However, the assumptions are different for the two types of regression analyses. For multiple linear regression analyses (using OLS), there are five assumptions that must be checked (Gauss-Markov Theorem): linearity in parameters (the relationship between the dependent and independent variables does not need to be linear, only the parameters must enter linearly), random sampling from the population, no perfect multicollinearity, zero conditional mean (exogeneity; meaning the predictors do not contain systematic information about the error term), and homoscedasticity (meaning the variance of the error term is constant for all values of the predictor; Greene 2018; Wooldridge 2010). Logistic regression analysis has different assumptions, due to the binary nature of the dependent variable. For instance, the assumptions of normality and homoscedasticity do not apply for logistic regression. Assumptions that do apply for the logistic regression are no perfect multicollinearity, independent observations, and linearity of the independent variables with the log-odds but not with the probability of the outcome variable (Hosmer et al. 2013).

After checking whether all the assumptions are met, the regression model is estimated and model fit is assessed (Stage 4). Stages 5 and 6 focus on the interpretation and validation of the results. In Stage 5, the role played by each independent variable in the prediction of the dependent variable is examined by interpreting the regression coefficients. Finally, the regression model is validated for ensuring the generalizability of the results (Stage 6).

After creating the multiple and logistic regression models from Hair et al. (2019; Chap. 5 and Chap. 8), we now focus on Stage 4 and Stage 5, which deal with the assessment procedures for the model fit, as well as the significance and relevance of the variates. Table 3 provides a quick overview of relevant measures to assess model fit, check assumptions, and interpret the variate for multiple linear and logistic regression.

Table 3 Measures implemented in SmartPLS for multiple linear regression and logistic regression analysis

	Multiple linear regression	Logistic regression
Model fit	R^2	LogLikelihood
	R^2 adjusted	Deviance difference
	ANOVA	AIC
		BIC
		McFadden's R^2
		Cox and Snell's R^2
		Nagelkerke's R^2
		Confusion matrix
Interpreting the variate		
Significance	*t*-values	Wald test
Sign	Unstandardized and standardized coefficients	Coefficients and exponential coefficients
Relative importance	Standardized coefficients	
Assumptions		
Linearity	Predicted vs. residuals plot	Box-Tidwell Test*
	Predicted vs. actual values plot	
Independence of residuals	Residual autocorrelation plot	
	Durbin-Watson test	
Homoscedasticity	Predicted vs. residuals plot	
	Breusch-Pagan test	
Normality	QQ plot	
	Residual histogram	

*Box-Tidwell test is not yet included in SmartPLS. It tests the linearity between the logit and the independent variables.

Importing practice data in SmartPLS

To import the two projects for the case studies, open the **Workspace view** and go to **Sample projects**. Navigate to the **Regression and NCA** sample projects and tick the box next to **Regression model** and **Logistic regression** (Fig. 2).

The two sample projects, **Example – Regression model** and **Example – Logistic regression**, appear in the SmartPLS **Workspace** on the left side. Each project includes two models and two data sets. For illustrative purposes, we deleted the models except the **Multivariate Data Analysis book** (Hair et al. 2019) in the **Example – Regression model** and **Example – Logistic regression** projects. To do so, left-click on a model in the SmartPLS **Workspace** to select it. Next, right-click on the selected model to open a dialog with options for the **Workspace** and select the **Delete resource** option. Then confirm the deletion in the subsequent dialog box. Similarly, delete the data files except the **HBAT_splits_data [100]** in the **Example – Logistic regression** project and the **HBAT_regression_data [100]** in the **Example – Regression model** project. As a result, each of the two projects only contains one model and one dataset in the Workspace, as shown in Fig. 3.

Case study I: multiple linear regression

The first case study focusses on running the HBAT multiple linear regression model example presented by Hair et al. (2019; Chap. 5) in SmartPLS. To start with an analysis, double click on the **Multivariate Data Analysis book** model from the **Example - Regression model** project in the SmartPLS **Workspace**. Next, the **Modeling view** opens displaying the multiple linear regression model as shown in Fig. 4.[2] Besides the intercept, the model includes the dependent variable customer satisfaction (X_{19}) and the independent variables representing perceptions of HBAT's performance; these are: quality (X_6), e-commerce (X_7), complaint resolution (X_9), product line (X_{11}), and salesforce (X_{12}). The ratio of observations to independent variables is 20:1, which meets the minimum sample size guidelines recommended by Hair et al. (2019; Chap. 5).

To estimate the regression model in SmartPLS, click on the **Calculate** button in the menu bar and select **Regression analysis**. Now the SmartPLS software opens a dialog box in

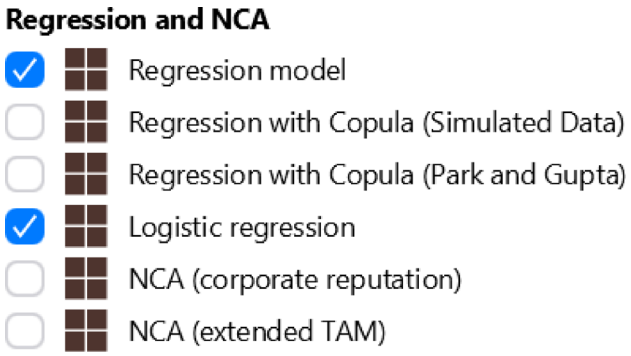


Fig. 2 Importing the sample projects in the SmartPLS software

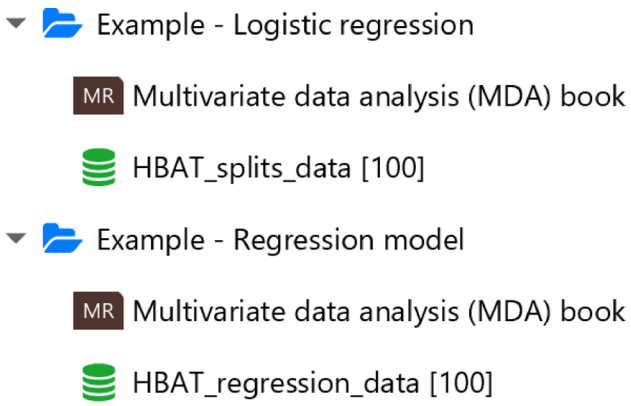


Fig. 3 Projects in the SmartPLS software

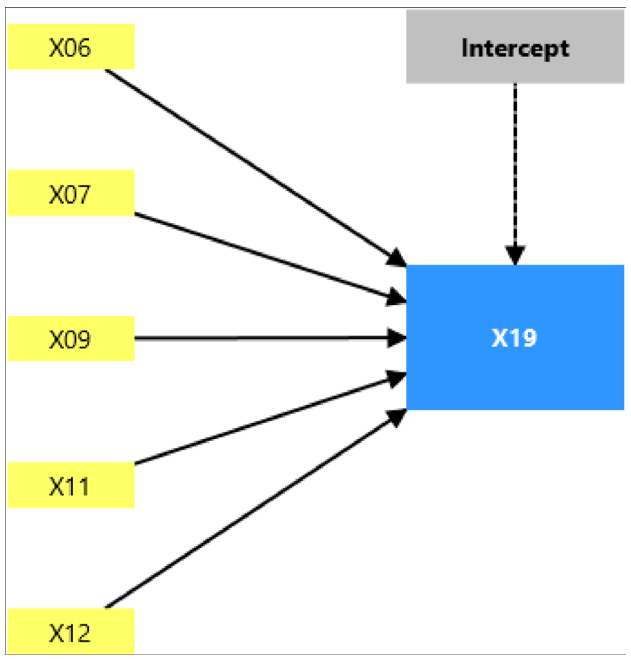


Fig. 4 Regression model (Hair et al. 2019; Chap. 5)

[2] *Note*: Unlike Hair et al. (2019), we renamed the variable names of the HABT dataset in SmartPLS. Specifically, under **Setup** in the SmartPLS Data view, we changed variable X1 to X9 into X01 to X09. This ensures an ascending ordering of indicators in all graphics and result tables (otherwise, for example, X12 would appear before X6). Alternatively, without renaming the variables, one could select in the SmartPLS Settings (Options) a sorting option that displays the constructs in the order in which they appear in the dataset.

Fig. 5 Regression analysis algorithm start dialog box in SmartPLS

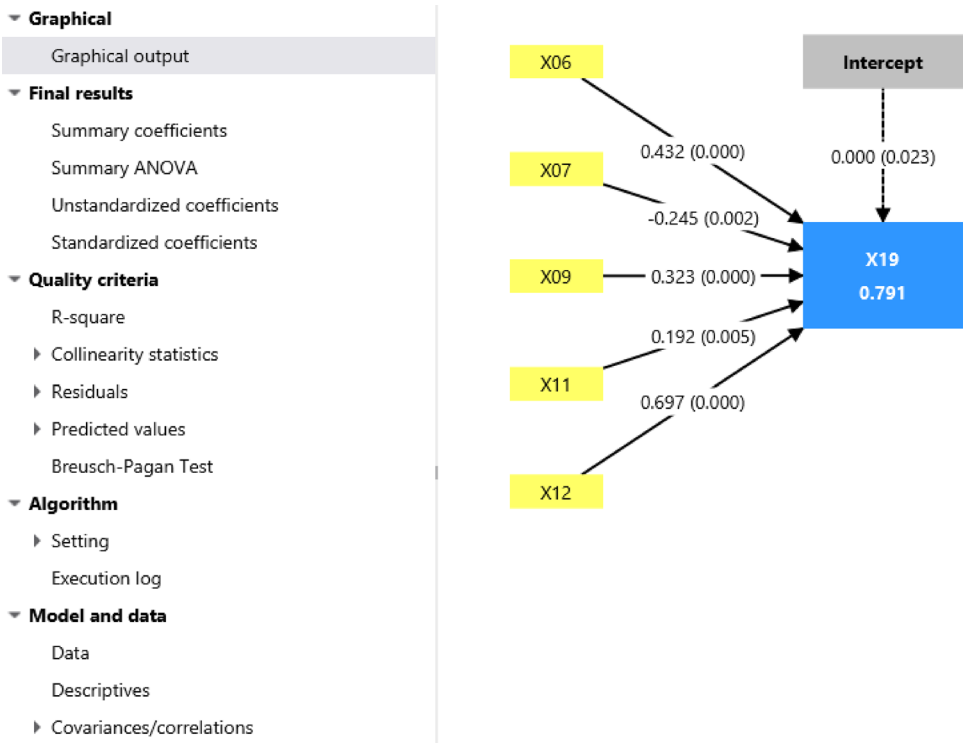


Fig. 6 Graphical output of the multiple linear regression

which the user can specify the test type (e.g., two tailed), the significance level (e.g., 0.05), and the standard error type (e.g., normal standard errors); the hyperlinks in the dialog box provide further explanations for each option and their alternatives. For this case study, we keep the default settings as shown in Fig. 5. Next, at the bottom of the **Regression analysis** dialog box, ensure that the box next to **Open report** has been ticked and click on the **Start calculation** button. The SmartPLS software then estimates the multiple linear regression model and, after completion, the default report results automatically opens.

The default report automatically opens the **Graphical output** (i.e., the regression model and its model estimation results; Fig. 6) At the bottom left of the SmartPLS **Results report** view, you can customize the results shown on the **Graphical output**. For the independent variables, we choose **Standardized coeffs. and p values**, while keeping **R-square** for dependent variable. The resulting **Graphical output** (Fig. 6) shows the standardized regression coefficients and the *p*-values in brackets; the value in the dependent variance X_{19} (i.e., 0.791) represents the amount of explained variance R^2. The left side of the **Results report** view shows a list of different regression analysis outputs provided by SmartPLS, which are ordered in five sections: **Graphical, Final results, Quality criteria, Algorithm**, and **Model and data.**

Under **Final results** we can select the **Summary ANOVA** results table (Fig. 7), which includes the following outcomes: (1) Regression sum of squares (111.205), which is the amount of variance explained by the model, and this value divided by the number of independent variables resulting in the mean square regression (22.241 = 111.205 / 5); (2) error sum of squares (29.422), which indicates the remaining unexplained variance, and this value divided by the number of independent variables resulting in the mean square error (0.313 = 29.422 / 94). Based on these outcomes, we obtain the *F*-ratio of 71.058 (= 22.241 / 0.313) at a significance level of 0.000. Hence, the regression model with its

Fig. 7 ANOVA results of the multiple linear regression analysis

	Sum square	df	Mean square	F	P value
Total	140.628	99	0.000	0.000	0.000
Error	29.422	94	0.313	0.000	0.000
Regression	111.205	5	22.241	71.058	0.000

	X19
R-square	0.791
R-square adjusted	0.780
Durbin-Watson test	2.449

Fig. 8 R^2 and Durbin-Watson test results

five predictors performs much better than a model with no predictors (i.e., if we used only the mean of the dependent variable X_{19} to predict its outcomes) and the large *F*-ratio indicates the model explains much more variance than the residual error.

Next, we obtain the R^2 coefficient of determination and the *adjusted* R^2 under **Quality criteria** in the **R-square** results table (Fig. 8). With an R^2 value of 0.791 (i.e., the regression sum of squares of 111.205 divided by the total sum of squares of 140.628; Fig. 7), the regression model in Fig. 8 explains almost 80% of the variance of customer satisfaction (X_{19}). The *adjusted* R^2 of 0.780 indicates no overfitting of the model and that the results can be generalized to other samples from our population. *Note*: We address the Durbin-Watson test shown in Fig. 8 later when we discuss the outcomes for the regression residuals.

Next, we turn to the estimated coefficients. In SmartPLS, the model's unstandardized and standardized regression coefficients can be viewed by selecting **Final Results** under **Summary coefficients**. An unstandardized coefficient is considered statistically significant if its 95% confidence interval, defined by the lower bound at the 2.5th percentile and the upper bound at the 97.5th percentile (Fig. 9), does not include zero. In our analysis, all five unstandardized regression coefficients, as well as the intercept, are significant at the 0.05 level. *Note*: Although the intercept is statistically significant, its negative value is difficult to interpret. All variables are measured on a 0–10 scale (see Table A1 in the Appendix). Consequently, when all independent variables take the value zero, the dependent variable customer satisfaction (X_{19}) should also be zero and not −1.151, as suggested by the estimated intercept. Such a value lies outside the permissible range of the dependent variable. This would argue in favor of estimating the regression model without an intercept. Nevertheless, for consistency with the textbook example in Hair et al. (2019; Chap. 5), we proceed with a model estimation including an intercept.

Before interpreting the regression coefficients, it is necessary to assess potential collinearity issues. High levels of collinearity can substantially distort the estimated regression coefficients. The default report provides the variance inflation factor (VIF) as an overall indicator of collinearity and the condition index derived from the decomposition of the coefficient variance. To inspect these two diagnostics, we need to choose **Collinearity statistics** under **Quality criteria**. As shown in Figs. 10 and 11, all VIFvalues fall below the more conservative threshold of 3, and all condition indices are below the recommended threshold of 30 (Hair et al. 2019). These results indicate collinearity is not a critical concern.

On these grounds, we can examine the signs of the regression coefficients to determine the direction of the relationships. As shown in Fig. 9, all coefficients except one are positive, indicating that more favorable perceptions of HBAT are associated with higher predicted customer satisfaction. The negative coefficient for e-commerce (X7) indicates that higher perceptions of this variable are associated with lower customer satisfaction. Although this finding may appear counterintuitive, Hair et al. (2019, pp. 348–349) provide a detailed discussion and explanation of such results.

	Unstandardized coefficients	Standardized coefficients	SE	T value	P value	2.5 %	97.5 %
X06	0.369	0.432	0.047	7.820	0.000	0.275	0.463
X07	-0.417	-0.245	0.132	3.162	0.002	-0.679	-0.155
X09	0.319	0.323	0.061	5.256	0.000	0.198	0.439
X11	0.174	0.192	0.061	2.860	0.005	0.053	0.295
X12	0.775	0.697	0.089	8.711	0.000	0.598	0.952
Intercept	-1.151	0.000	0.500	2.303	0.023	-2.143	-0.159

Fig. 9 The multiple linear regression model's coefficients

	VIF
X06	1.373
X07	2.701
X09	1.701
X11	2.033
X12	2.880

Fig. 10 VIF values

In addition, the standardized coefficients provide information about the relative importance of each independent variable (Fig. 9). The standardized coefficient of 0.697 for salesforce image (X_{12}) indicates that it is the strongest predictor of customer satisfaction (X_{19}), followed by product quality (X_6), complaint resolution (X_9), e-commerce (X_7), and product line (X_{11}). Importantly, beyond statistical significance, all standardized coefficients exceed the commonly used relevance threshold of 0.10, indicating each regressor makes a meaningful contribution to explaining the dependent variable.

In evaluating the estimated equation, we consider not only the regressor' statistical significance and relevance but also whether the variables meet the assumptions of regression analysis. SmartPLS provides several diagnostic tools for assessing these assumptions and the quality of the regression model. For example, plotting the residuals (i.e., the difference between the observed and predicted values of the dependent variable) against the predicted dependent values is a basic method for detecting potential violations of linearity and residual homoscedasticity. In the SmartPLS report, this plot can be obtained via **Quality criteria** > **Predicted value** > **Predicted vs. residual** (Fig. 12). As shown, the residuals do not display any nonlinear pattern. Similarly, the plot of predicted versus actual values, accessed through **Quality criteria** > **Predicted value** > **Predicted vs. actual** (Fig. 13), also shows no nonlinear pattern, indicating the assumption of linearity is satisfied for the overall variable.

In addition, Fig. 12 shows no systematic pattern of increasing or decreasing residuals, indicating that heteroscedasticity is not a critical issue. In SmartPLS, heteroscedasticity can also be formally assessed using the Breusch-Pagan test, available via **Quality criteria** > **Breusch-Pagan test**. The results of this test, with a *p*-value greater than 0.05, further confirm the assumption of homoscedasticity is satisfied for our model (Fig. 14).

Next, to assess the independence of residuals, navigate to **Quality criteria** > **Residuals** > **Residual autocorrelation plot**. As shown in Fig. 15, no consistent pattern is observed. This finding is further supported by the results of the Durbin-Watson test (Fig. 8), which confirms the independence of the residuals.

Finally, the normality of the error term can be assessed using a QQ-plot or a residual histogram. These plots can be obtained in SmartPLS via **Quality criteria** > **Residuals**, then selecting either **QQ plot** or **Residual histogram**. As shown in Figs. 16 and 17, while a few values deviate from the reference line, the majority of residuals follow an approximately normal distribution.

Case study II: logistic regression

Our second case study demonstrates how to use SmartPLS to conduct a logistic regression. For the logistic regression analysis, double click on the **Multivariate Data Analysis**

	Eigenvalue	Condition index	X06	X07	X09	X11	X12	Intercept
0	5.858	1.000	0.001	0.000	0.001	0.001	0.000	0.000
1	0.073	8.935	0.035	0.045	0.017	0.090	0.060	0.000
2	0.037	12.661	0.244	0.001	0.380	0.014	0.003	0.022
3	0.015	19.668	0.078	0.059	0.409	0.778	0.014	0.118
4	0.010	24.543	0.532	0.054	0.054	0.041	0.275	0.653
5	0.007	28.647	0.110	0.842	0.140	0.076	0.648	0.207

Fig. 11 Condition index

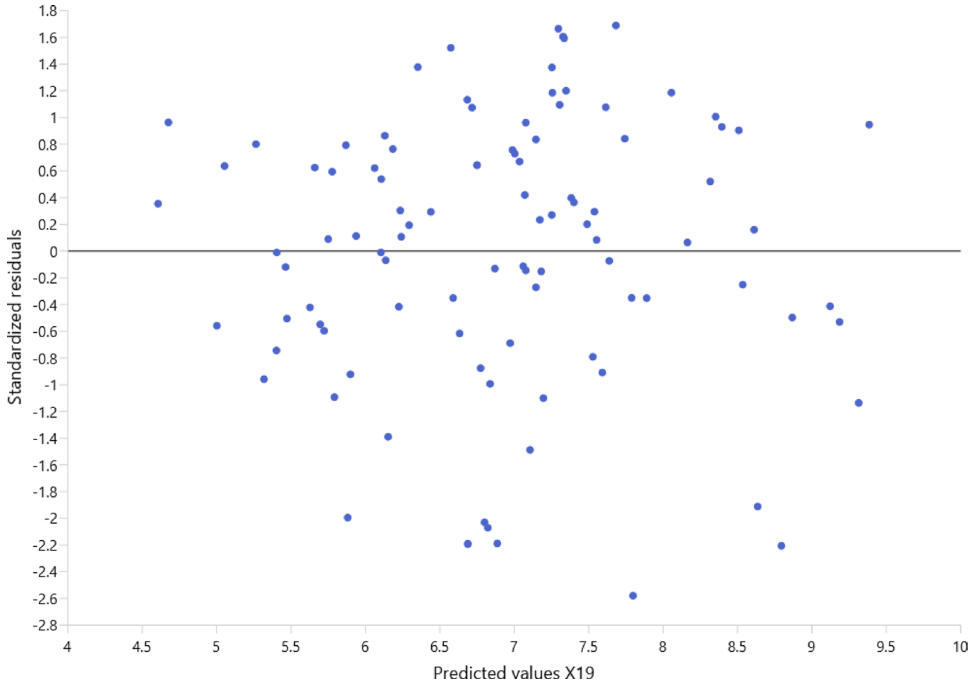


Fig. 12 Predicted vs. residual

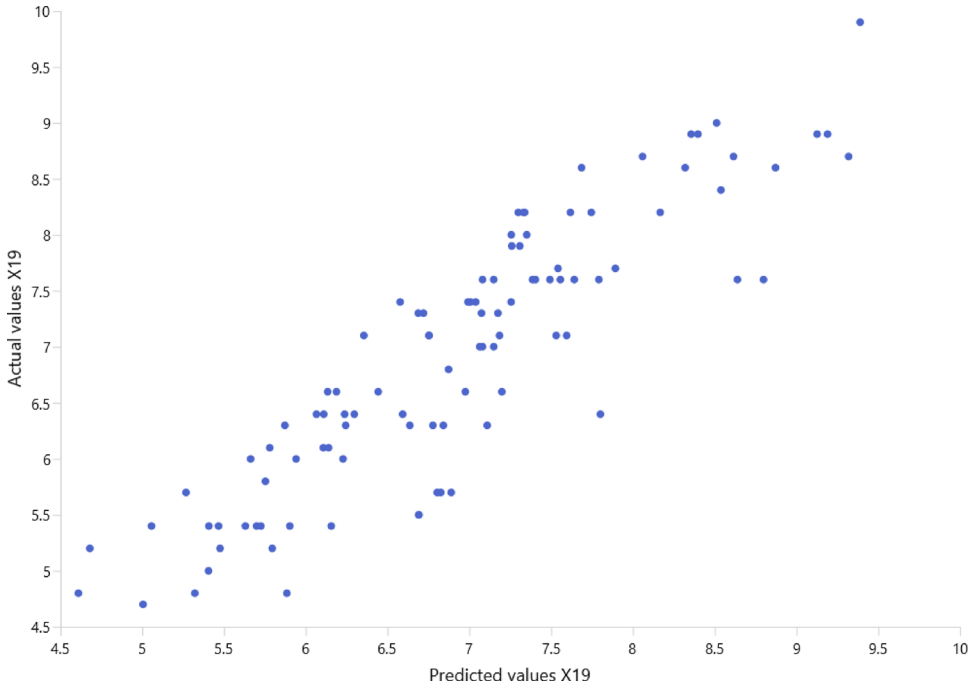


Fig. 13 Predicted vs. actual values

Fig. 14 Results of Breusch-Pagan test

	Test-Statistic	df	P value
Breusch-Pagan Test	4.037	5	0.544

book model located in the **Example – Logistic regression**. Figure 18 presents the logistic regression model as displayed in the Modeling view (in this case with the displayed results, which we obtain after model estimation). *Note*: Double click on the dependent X_4 variable and, under Indicator sort order, choose the option Import order to obtain the same order of the independent variables a displayed in Fig. 18.

SmartPLS represents the logistic regression model in the same manner as a multiple linear regression model. However, the dependent variable is now the binary variable

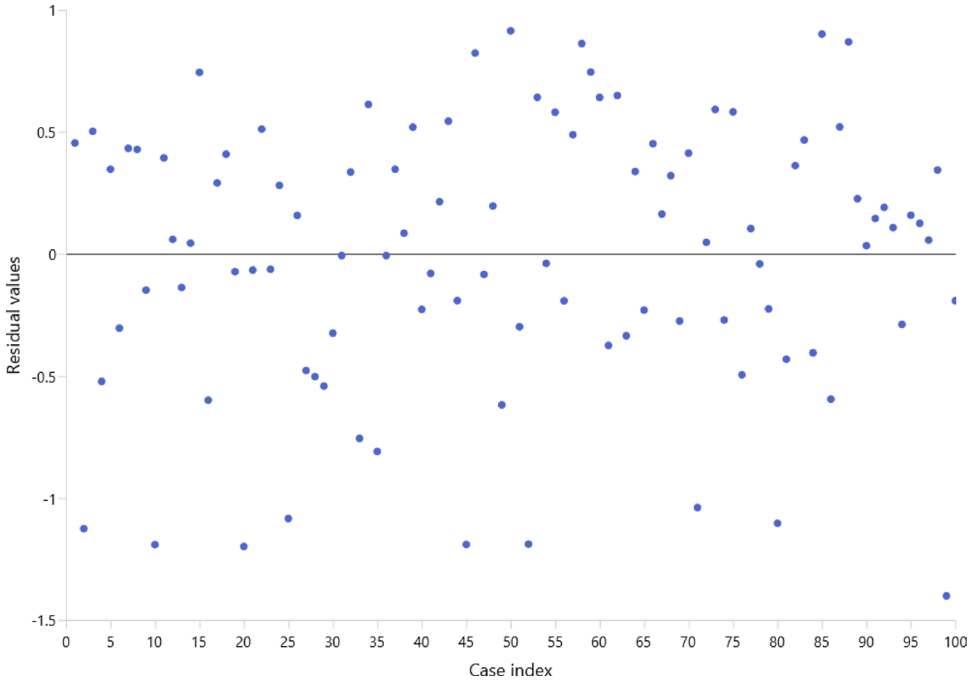


Fig. 15 Residual autocorrelation plot

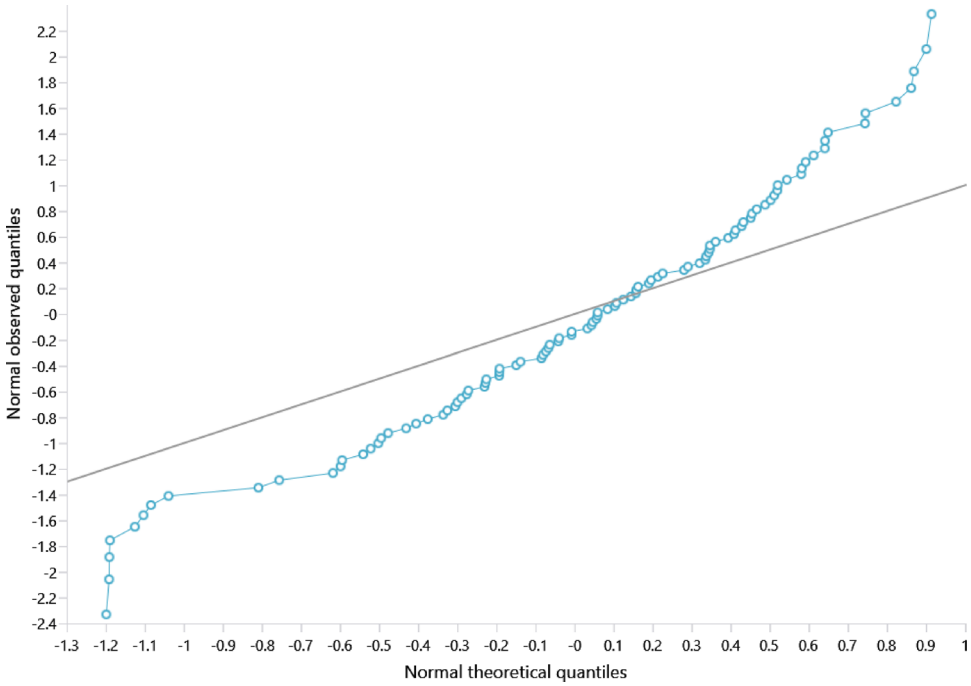


Fig. 16 QQ-plot

Region (X_4) with 0 indicating the customer location in the USA/North America and 1 outside North America. All 13 independent variables (X_6-X_{18}) represent perceptions of HBAT (Table A1 in the Appendix). In the logistic regression model example, we regress the binary dependent variable X_4 on the interval-scaled (quasi-metric) independent variables X_6-X_{18}. The data set of 100 observations is small and so it might be less efficient in model estimation. Hair et al. (2019; Chap. 8) further split the sample into 60 observations for estimation sample and 40 observations for the hold-out sample for predictive validation and variable selection. As our illustration will not focus on predictive variable selection, we will use the full sample of 100 observations for estimating the following results which therefore deviate from Hair et al. (2019).

For estimating the logistic regression model select **Calculate** > **Logistic regression**. A dialog box opens allowing us to specify the test type (one- or two-tailed), the significance level (e.g., 0.05), the maximum number of iterations for the maximum likelihood (ML) algorithm, and the

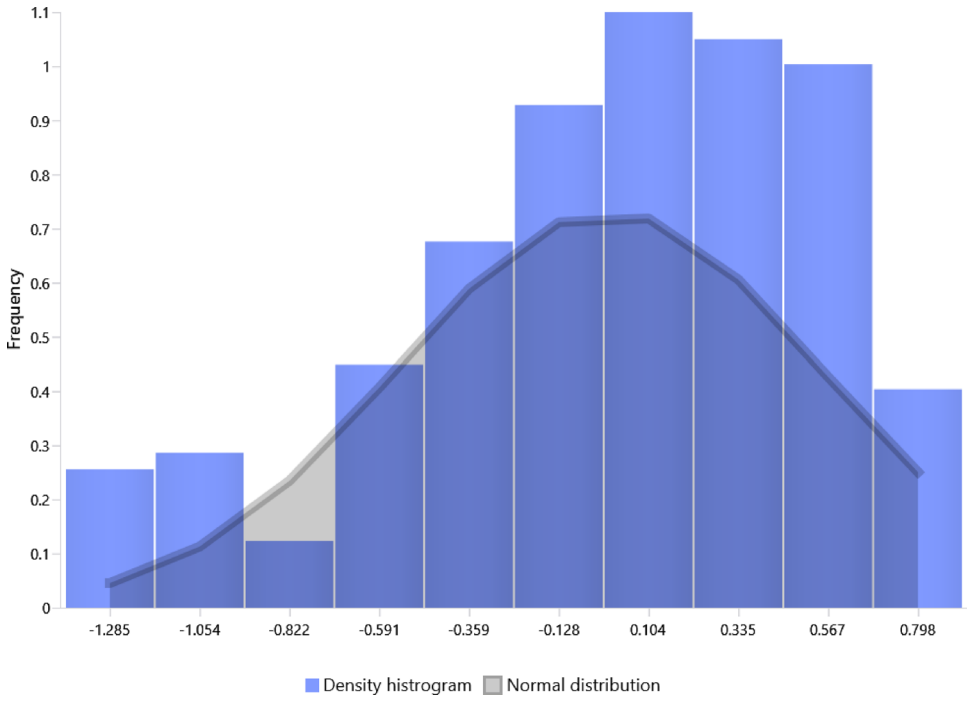


Fig. 17 Residual histogram

stopping criterion for the model estimation. We recommend you keep the default settings, as shown in Fig. 19. Next, click on the **Start calculation** button, after ensuring that the box next to **Open report** is ticked.

After the computations finish, the **Graphical output** (Fig. 18) in the SmartPLS **Results report** view automatically opens the report. The **Graphical output** shows the logistic regression coefficients of the independent variables X_6 to X_{18} and the pseudo R^2 (i.e., displayed in the depend binary variable X_4). On the left side, the user can navigate through the different options of the report: **Graphical output, Final results, Model fit, Algorithm, Model and data**.

To evaluate the overall goodness of fit of the logistic regression model, open the results under **Model fit** > **Fit summary** in the **Results report**. The primary criterion of overall model fit is the likelihood value, which is conceptually analogous to the sums of squares used in multiple linear regression (Hair et al. 2019; Chap. 8). We evaluate model fit using the value of −2 times the log-likelihood (−2LL), where a value of 0 indicates a perfect fit. Accordingly, a lower −2LL value for the estimated model relative to the null model denotes superior fit. In SmartPLS, this statistic is reported in the Deviance (−2LL) column of the **Fit summary** table. As illustrated in Fig. 20, the estimated model yields a value of 30.789, which demonstrates a good model fit since it is substantially lower than the value of 133.750 obtained for the null model. The AIC and BIC values for the estimated model, which are likewise lower than those of the null model, further corroborate this conclusion.

Model fit can also be assessed with the help of the logistic regression model's pseudo R^2 values, which indicates how well the regression model accounts for the variation between the two groups of customers. More specifically, SmartPLS presents three different measures for pseudo R^2. **McFadden's** $\boldsymbol{R^2}$ compares the log-likelihood of the estimated model with that of the null model. Its values range from 0 to 1. A perfect fit has a R^2 value of 1, which is equivalent to a deviance of 0. The **Cox and Snell's** $\boldsymbol{R^2}$ is interpreted in the same way, with higher values indicating a greater fit. However, this measure of R^2 cannot reach the maximum value of 1. To overcome this limitation, **Nagelkerke's** $\boldsymbol{R^2}$ incorporates a modification in order to reach this maximum value. As we see in Fig. 20, the three measures of "pseudo" R^2 have values above 0.5 indicating that the logistic regression model explains the outcomes sufficiently well.

The third measure for the overall model fit is represented by the predictive accuracy of the model. In SmartPLS, we can use the confusion matrix to measure the predictive accuracy of predictive outcomes; click on **Final results** > **Predictions and probabilities** > **Confusion matrix** (Fig. 21). The high values of the percentages of 97.436 and 91.803 show high predictive accuracy of our model. This results in

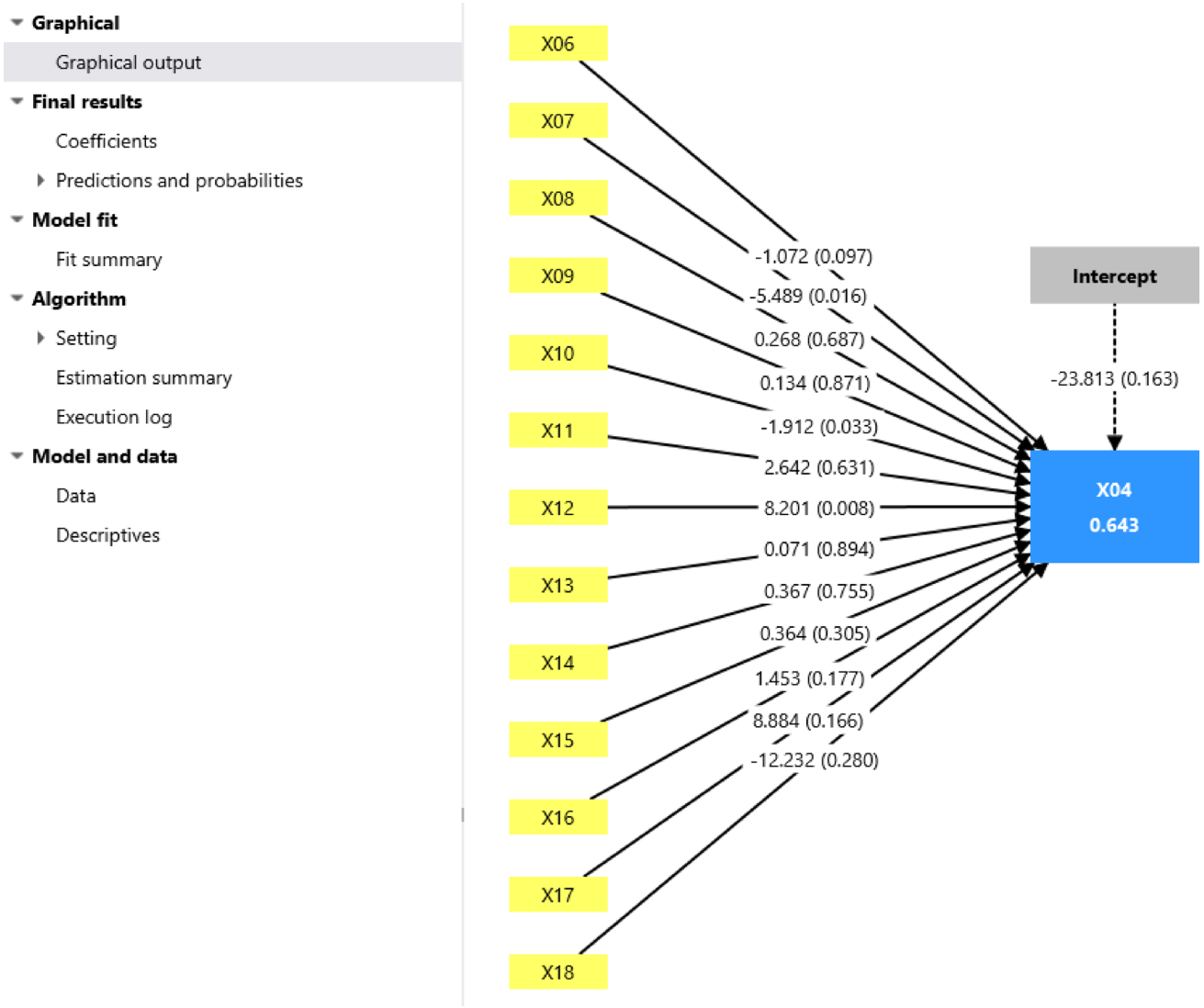
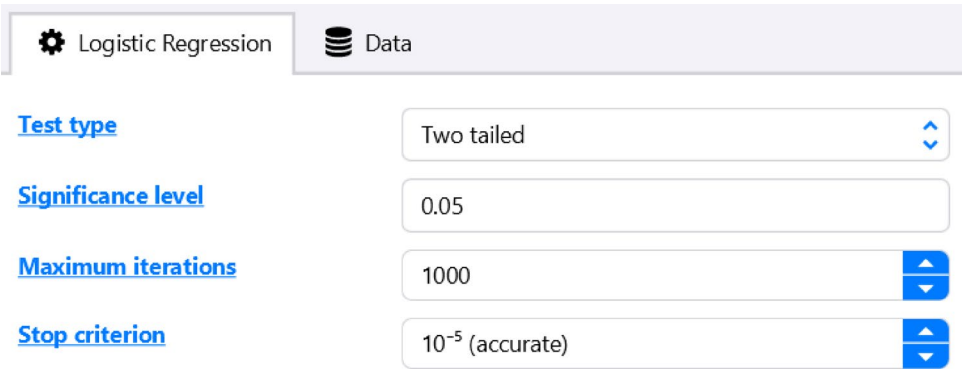


Fig. 18 Graphical results output for logistic regression model

Fig. 19 Logistic regression dialog box in SmartPLS



	LogLikelihood	Deviance	AIC	BIC	df	Cox and Snell's R-Square	Nagelkerke's R-Square	McFadden's R-Square
Null model	-66.875	133.750	135.750	138.355	99			
Estimated model	-15.394	30.789	58.789	95.261	86	0.643	0.872	0.770

Fig. 20 Summary of fit indices

Fig. 21 Confusion matrix

	Predicted value = 0.0	Predicted value = 1.0	Percentage correct
X04 = 0.0	38	1	97.436
X04 = 1.0	5	56	91.803

	Coefficients	SE	z-statistic	Wald	P value	Exp(Coefficients)
X06	-1.072	0.646	-1.661	2.758	0.097	0.342
X07	-5.489	2.269	-2.419	5.851	0.016	0.004
X08	0.268	0.666	0.403	0.162	0.687	1.308
X09	0.134	0.824	0.162	0.026	0.871	1.143
X10	-1.912	0.898	-2.130	4.537	0.033	0.148
X11	2.642	5.500	0.480	0.231	0.631	14.045
X12	8.201	3.109	2.638	6.957	0.008	3645.181
X13	0.071	0.532	0.134	0.018	0.894	1.074
X14	0.367	1.174	0.312	0.098	0.755	1.443
X15	0.364	0.355	1.026	1.052	0.305	1.439
X16	1.453	1.077	1.349	1.820	0.177	4.274
X17	8.884	6.410	1.386	1.921	0.166	7217.800
X18	-12.232	11.316	-1.081	1.168	0.280	0.000
Intercept	-23.813	17.052	-1.396	1.950	0.163	0.000

Fig. 22 The logistic regression model's coefficients and Wald's test

the typical member of group 0 (USA) having a probability of 97.436 of being correctly assigned to group 0 and the typical member of group 1 (non-USA) has a probability of 91.803 of being correctly assigned to group 1. This demonstrates the ability of the logistic model to create good separation between the two groups in terms of predicted probability, resulting in the excellent classification results.

Next, we examine the logistic coefficients to assess their statistical significance, their sign, and the impact that each variable has on predicted probability of group membership. In the default report, you now need to click on **Final results** > **Coefficients** to display the results in Fig. 22. In logistic regression analysis, we use the Wald test to measure statistical significance for each estimated coefficient, which represents changes in the log-odds. Thereby we implicitly test the effect of an independent variable on predicting the group membership.

As shown in Fig. 22, the logistic regression coefficients of X_7 (-5.489), X_{10} (-1.912), and X_{12} (8.201) are significant at the 0.05 level. The sign of the logistic coefficients shows the direction of the relationship. In this case, the negative sign of the coefficient for the relationship between the independent X_7 and X_{10} variables and the dependent X_4 variable indicates that the likelihood of observing a one (i.e., $P(Y = 1|X)$ decrease when X_7 and X_{10} are increasing (or put differently the probability that the customer will be categorized as residing in the USA/North America increases). In contrast, the coefficient of X_{12} is positive indicating a positive relationship, thus increasing the probability that the customer will be categorized as residing outside North America.

However, as discussed when introducing the logistic regression model, the logistic regression coefficient cannot be interpreted as a direct change in probability. Thus, a one-unit change in the independent variable X_7 reflects a -5.489 change in the log-odds (logit), but not in the probability of the outcome being one. While we can conclude a directional

change in probability, the exact change in probability depends on the values of the other independent variables.

As discussed before, the exponentiated coefficients (odds ratio) can also be used to assess the size and direction of the relationship. Values above 1 indicate a positive relationship and below 1 a negative relationship. Hence, the exponential coefficients of 0.004 (X_7) and 0.148 (X_{10}) show negative relationships, while the outcome of 14.045 for X_{12} determines a positive relationship. For example, when X_{12} increases by one unit the initial odds are multiplied with 14.045, which implies a substantial increase. Similarly, we can use the exponentiated coefficient minus one, which expresses the percentage in change in odds. In our example, this means that an increase of one point decreases the odds by 99.6% for X_7, decreases the odds by 85.2% for X_{10}, and increases the odds by 3644.2% for X_{12}.

Observations and conclusions

Based on the case studies of multiple and logistic regressions by Hair et al. (2019), this article offers a comprehensive guide on how to conduct a regression analysis using the SmartPLS software. We show how multiple and logistic regression models are estimated and assessed in SmartPLS, starting with the statistical significance of the overall model, followed by the relative contribution of regression coefficients for the explanation and prediction of the dependent variable, and finally checking the regression assumptions. As such, this article provides comprehensive guidelines for scholars wanting to carry out their regression analyses with the SmartPLS software.

SmartPLS offers a user-friendly interface for regression analysis with results presented in easily readable tables and graphs. Extending our demonstrations of the multiple linear regression and the logistic regression analyses in SmartPLS, researchers can also access additional analyses (e.g., Hair et al. 2024; Hair et al. 2026), such as necessary condition analysis (Sarstedt et al. 2024); mediator and moderator analysis (i.e., by adding an interaction term); moderated mediation; and conditional process analysis using PROCESS (e.g., Hayes 2022). If researchers want to analyze relationships between latent variables they can perform more advanced analyses, such as PLS-SEM, CB-SEM, or GSCA.

Based on the guidelines of Hair et al. (2019) for performing multiple linear regression analysis, future extensions of SmartPLS could support additional graphs to test assumptions and identifying influential observations, such as the standardized partial regression plots and residual-versus-leverage plots. The logistic regression algorithm implemented in SmartPLS is currently in the beta stage; therefore, the inclusion of additional features, such as receiver operating characteristic (ROC) curve analysis (e.g., Hanely & McNeil 1982) would enable researchers to more comprehensively assess the model's predictive accuracy. Moreover, implementing the Box–Tidwell test would allow for a systematic evaluation of the linearity assumption (Box & Tidwell 1982).

One aspect not further addressed in this article, but of substantial importance for the estimation of regression models, is the assessment and treatment of endogeneity. To address this issue, Park and Gupta (2012) introduced the instrument variable (IV)-free Gaussian copula approach. Subsequent publications have elaborated on this method and generated additional insights into the instrument-free framework (e.g., Becker et al. 2022; Eckert and Hohberger 2023; Hult et al. 2018; Park and Gupta 2024; Quian et al. 2025; Yang et al. 2025). Most notably, Liengaard et al. (2025) proposed an extended framework for handling endogeneity using the Gaussian copula approach, which is implemented in SmartPLS and can be fully utilized by applied researchers. In addition, SmartPLS supports single and multiple mediation, moderation, and conditional process analyses (Cheah et al. 2021; Hayes 2022; Sarstedt et al. 2020) using linear regression models. Further, researchers can gain more from their regression analysis results by applying the IPMA to their regression models (Ringle and Sarstedt 2016), which is also fully implemented in SmartPLS. Finally, unobserved heterogeneity represents a validity threat to regression model results (e.g., Becker et al. 2013; Jedidi et al. 1997). Latent class segmentation (e.g., by using the finite mixture approach; Becker et al. 2015; Wedel and DeSarbo 2002) allows us to address this critical issue. By using single-item constructs, researchers can use the FIMIX-PLS approach (Hahn et al. 2002; Sarstedt et al. 2011) implemented in SmartPLS to apply finite mixture latent class segmentation not only for more complex PLS path models, but also for simple regression models. Thereby, they can either indicate that unobserved heterogeneity does not represent a critical validity threat to regression results or, alternatively, reveal suitable segments for further group analyses. We encourage researchers to use this expanded portfolio of additional and complementary analyses offered by SmartPLS to execute richer and better validated regression model analyses results in their research.

Appendix

Table 4 HBAT data (Hair et al. 2019)

Variable	Name	Survey question	Scale
Data warehouse classification variables			
X_1	Customer Type	Length of time a particular customer has been buying from HBAT	1–3 scale: 1 = less than 1 year; 2 = between 1 and 5 years; 3 = longer than 5 years
X_2	Industry Type	Type of industry that purchases HBAT's paper products	Binary: 0 = magazine industry; 1 = newsprint industry
X_3	Firm Size	Employee size	Binary: 0 = small firm, fewer than 500 employees; 1 = large firm, 500 or more employees
X_4	Region	Customer location	Binary: 0 = USA/North America; 1 = outside North America
X_5	Distribution System	How paper products are sold to customers	Binary: 0 = sold indirectly through a broker; 1 = sold directly
Perceptions of HBAT			
X_6	Product Quality	Perceived level of quality of HBAT's paper products	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_7	E-Commerce	Activities/Website Overall image of HBAT's website, especially user-friendliness	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_8	Technical Support	Extent to which technical support is offered to help solve product/service issues	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_9	Complaint Resolution	Extent to which any complaints are resolved in a timely and complete manner	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_{10}	Advertising	Perceptions of HBAT's advertising campaigns in all types of media	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_{11}	Product Line	Depth and breadth of HBAT's product line to meet customer needs	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_{12}	Salesforce	Image Overall image of HBAT's salesforce	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_{13}	Competitive Pricing	Extent to which HBAT offers competitive prices	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_{14}	Warranty and Claims	Extent to which HBAT stands behind its product/service warranties and claims	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_{15}	New Products	Extent to which HBAT develops and sells new products	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_{16}	Ordering and Billing	Perception that ordering and billing is handled efficiently and correctly	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_{17}	Price Flexibility	Perceived willingness of HBAT sales reps to negotiate price on purchases of paper products	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_{18}	Delivery Speed	Amount of time it takes to deliver the paper products once an order has been confirmed	0–10 scale: 0 = "Poor" and 10 = "Excellent"
Purchase outcomes			
X_{19}	Customer Satisfaction	Customer satisfaction with past purchases from HBAT	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_{20}	Likelihood of Recommending HBAT	Likelihood of recommending HBAT to other firms as a supplier of paper products, measured on a 10-point graphic rating scale	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_{21}	Likelihood of Future Purchases from HBAT	Likelihood of purchasing paper products from HBAT in the future, measured on a 10-point graphic rating scale	0–10 scale: 0 = "Poor" and 10 = "Excellent"
X_{22}	Percentage of Purchases from HBAT	Percentage of the responding firm's paper needs purchased from HBAT	100-point percentage scale
X_{23}	Perception of Future Relationship with HBAT	Extent to which the customer/respondent perceives his or her firm would engage in strategic alliance/partnership with HBAT	Binary: 0 = Would not consider; 1 = Yes, would consider strategic alliance or partnership

https://www.smartpls.com

Funding Open Access funding enabled and organized by Projekt DEAL.

Declarations

Acknowledgements This article uses and discusses the statistical software SmartPLS (https://www.smartpls.com). Christian M. Ringle and Jan-Michael Becker acknowledge a financial interest in SmartPLS..

References

Becker, J.-M., A. Rai, C. M. Ringle, and F. Völckner. 2013. Discovering unobserved heterogeneity in structural equation models to avert validity threats. *MIS Quarterly* 37(3):665–694.

Becker, J.-M., N. F. Richter, C. M. Ringle, and M. Sarstedt. 2026. Must-have, or maybe not? A sensitivity-based extension to necessary condition analysis. Journal of Business Research, 206:115920.

Becker, J.-M., C. M. Ringle, M. Sarstedt, and F. Völckner. 2015. How collinearity affects mixture regression results. *Marketing Letters* 26(4):643–659.

Becker, J.-M., D. Proksch, and C. M. Ringle. 2022. Revisiting Gaussian copulas to handle endogenous regressors. *Journal of the Academy of Marketing Science* 50:46–66.

Black, W. C., and B. J. Babin. 2019. Multivariate data analysis: Its approach, evolution, and impact. In *The great facilitator: Reflections on the contributions of Joseph F. Hair, Jr. to marketing and business research*, ed. B. J. Babin, and M. Sarstedt. 121–130. Cham: Springer.

Box, G. E. P., and P. W. Tidwell. 1962. Transformation of the independent variables. *Technometrics* 4(4):531–550.

Cheah, J.-H., C. Nitzl, J. L. Roldán, G. Cepeda Carrión, and S. P. Gudergan. 2021. A primer on the conditional mediation analysis in PLS-SEM. *ACM SIGMIS Database* 52:43–100.

Cheah, J.-H., F. Magno, and F. Cassia. 2024. Reviewing The SmartPLS 4 software: the latest features and enhancements. *Journal of Marketing Analytics* 12:97–107.

Chua, Y. P. 2024. *A step-by-step guide to SMARTPLS 4: ata analysis using PLS-SEM, CB-SEM, process and regression*. Kuala Lumpur: Researchtree Education.

Dul, J. 2016. Necessary condition analysis (NCA): logic and methodology of necessary but not sufficient causality. *Organizational Research Methods* 19(1):10–52.

Eckert, C., and J. Hohberger. 2023. Addressing endogeneity without instrumental variables: an evaluation of the Gaussian copula approach for management research. *Journal of Management* 49(4):1460–1495.

Greene, W. H. 2018. *Econometric analysis*. 8th ed. Harlow: Pearson.

Hahn, C., M. D. Johnson, A. Herrmann, and F. Huber. 2002. Capturing customer heterogeneity using a finite mixture PLS approach. *Schmalenbach Business Review* 54(3):243–269.

Hair, J. F., W. C. Black, B. J. Babin, and R. E. Anderson. 2019. *Multivariate data analysis*. 8th ed. Boston, MA: Cengage Learning.

Hair, J. F., B. J. Babin, C. M. Ringle, M. Sarstedt, and J.-M. Becker. 2025. Covariance-based structural equation modeling (CB-SEM): a SmartPLS 4 software tutorial. *Journal of Marketing Analytics* 13:709–724.

Hair, J. F., M. Sarstedt, C. M. Ringle, and S. P. Gudergan. 2024. Advanced issues in partial least squares structural equation modeling (PLS-SEM). 2nd ed. Thousand Oaks, CA: Sage.

Hair, J. F., G. T. M. Hult, C. M. Ringle, and M. Sarstedt. 2026. *A primer on partial least squares structural equation modeling (PLS-SEM)*. 4th ed. Thousand Oaks, CA: Sage.

Hanley, J. A., and B. J. McNeil. 1982. The meaning and use of the area under a receiver operating characteristic (ROC) curve. *Radiology* 143(1):29–36.

Hayes, A. F. 2022. *Introduction to mediation, moderation, and conditional process analysis: a regression-based approach*. 3rd ed. New York: Guilford Press.

Hosmer, D. W., S. Lemeshow, and R. Y. Sturdivant. 2013. *Applied logistic regression*. 3rd ed. Hoboken, NJ: Wiley.

Hult, G. T. M., J. F. Hair, D. Proksch, M. Sarstedt, A. Pinkwart, and C. M. Ringle. 2018. Addressing endogeneity in international marketing applications of partial least squares structural equation modeling. *Journal of International Marketing* 26(3):1–21.

Hwang, H., and Y. Takane. 2004. Generalized structured component analysis. *Psychometrika* 69(1):81–99.

Jedidi, K., H. S. Jagpal, and W. S. DeSarbo. 1997. Finite-mixture structural equation models for response-based segmentation and unobserved heterogeneity. *Marketing Science* 16(1):39–59.

Jöreskog, K. G. 1978. Structural analysis of covariance and correlation matrices. *Psychometrika* 43(4):443–477.

LaValley, M. P. 2008. Logistic regression. *Circulation* 117:2395–2399.

Liengaard, B. D., J.-M. Becker, M. Bennedsen, P. Heiler, L. N. Taylor, and C. M. Ringle. 2025. Dealing with regression models' endogeneity by means of an adjusted estimator for the Gaussian copula approach. Journal of the Academy of Marketing Science 53:279–299.

Lee, M. J., G. Lee, G., J. Y. Choi. 2025. Linear probability model revisited: why it works and how it should be specified. Sociological Methods & Research 54(1):173–186.

Matthews, L., M. Sarstedt, J. F. Hair, and C. M. Ringle. 2016. Identifying and treating unobserved heterogeneity with FIMIX-PLS: Part II – a case study. *European Business Review* 28(2):208–228.

Memon, M. A., T. Ramayah, J.-H. Cheah, H. Ting, and T. H. Cham. 2021. PLS-SEM statistical program: a review. *Journal of Applied Structural Equation Modeling* 5(1):i–xiii.

Merkle, A. C. 2025. Item-level correction: detecting, removing, and reporting common methods variance. *Journal of Marketing Analytics* 13:405–423.

Park, S., and S. Gupta. 2012. Handling endogenous regressors by joint Estimation using copulas. *Marketing Science* 31(4):567–586.

Park, S., and S. Gupta. 2024. A review of copula correction methods to address regressor–error correlation. *Impact at JMR* . Available at: https://www.ama.org/marketing-news/a-review-of-copula-correction-methods-to-address-regressorerror-correlation/.

Qian, Y., A. Koschmann, and H. Xie. 2025. EXPRESS: practical guide to endogeneity correction using copulas. *Journal of Marketing*, forthcoming.

Richter, N. F., S. Schubring, S. Hauff, C. M. Ringle, and M. Sarstedt. 2020. When predictors of outcomes are necessary: guidelines for the combined use of PLS-SEM and NCA. *Industrial Management & Data Systems* 120(12):2243–2267.

Ringle, C. M., and M. Sarstedt. 2016. Gain more insight from your PLS-SEM results: the importance–performance map analysis. *Industrial Management & Data Systems* 116(9):1865–1886.
Ringle, C. M., S. Wende, and J.-M. Becker. 2024. *SmartPLS 4*. Bönningstedt: SmartPLS.
Sarstedt, M., and J.-H. Cheah. 2019. Partial least squares structural equation modeling using SmartPLS: a software review. *Journal of Marketing Analytics* 7:196–202.
Sarstedt, M., and E. Mooi. 2019. *A concise guide to market research: the process, data, and methods using IBM SPSS Statistics*. 3rd ed. Berlin: Springer.
Sarstedt, M., J. F. Hair, C. M. Ringle, and M. Howard. 2020. Beyond a tandem analysis of SEM and PROCESS: use of PLS-SEM for mediation analyses. *International Journal of Market Research* 62(3):288–299.
Sarstedt, M., N. F. Richter, and C. M. Ringle. 2024. Combined importance–performance map analysis (cIPMA) in partial least squares equation modeling (PLS-SEM): a SmartPLS tutorial. *Journal of Marketing Analytics* 12:746–760.
Wedel, M., and W. S. DeSarbo. 2002. Market segment derivation and profiling via a finite mixture model framework. *Marketing Letters* 13(1):17–25.
Wooldridge, J. M. 2010. *Econometric analysis of cross section and panel data*. 2nd ed. Cambridge, MA: MIT Press.
Yang, F., Y. Qian, and H. Xie. 2025. Addressing endogeneity using a two-stage copula generated regressor approach. *Journal of Marketing Research* 62(4):601–623.

Vasilica-Maria Margalina is an associate professor in the Department of Business Administration and Marketing at the Faculty of Economics at the University "1 Decembrie 1918" of Alba Iulia (Romania). She previously worked as a professor at CESINE Centro Universitario of Santander (Spain) and as a researcher at Technical University of Ambato (Ecuador). She has conducted research in the field of consumer and organizational behavior with contributions published in journals such as Studies in Higher Education, Revista Latina de Comunicación Social and International Journal of Human Capital and Information Technology (IJHCITP). Her research interest also covers business analytics and behavioral research methods. She often presents seminars on research techniques and multivariate data analysis for universities in Latin America and Spain.

Charlotte Kreienbaum is a research fellow at the Institute of Management and Decision Sciences at the Hamburg University of Technology (Germany) and a rising PhD student whose research interests lie in organizational behavior and business analytics. Trained in psychology, Charlotte conducts research at the intersection of human behavior and data-driven decision making, with a particular interest in how organizations can optimize performance, well-being, and strategic outcomes. By combining quantitative methods with a behavioral science foundation, her work advances evidence-based insights for decision making in complex organizational and societal systems. Her scholarly contributions include a publication in the Journal of Occupational Health Psychology and ongoing projects advance empirical and methodological insights for management and applied behavioral research.

Joseph F. Hair Jr. is Cleverdon Chair of Business, and director of the PhD degree in business administration, Mitchell College of Business, University of South Alabama. He previously held the Copeland Endowed Chair of Entrepreneurship and was Director of the Entrepreneurship Institute, Ourso College of Business Administration, Louisiana State University. Joe was recognized by Clarivate Analytics from 2018 through 2024 for being in the top 1% globally of all business and economics professors based on his citations and scholarly accomplishments, which exceed 550,000 over his career. He has authored more than 145 editions of his books, including Multivariate Data Analysis (8th edition, 2019; cited 240,000+ times), PLS-SEM Primer (4th edition; 2026); Advanced PLS (2nd edition, 2024); MKTG (14th edition, 2024); Essentials of Business Research Methods (5th edition, 2024); Essentials of Marketing Analytics (2nd edition, 2025); and Essentials of Marketing Research (6th edition, 2026). He also has published numerous articles in scholarly journals and was recognized as the Academy of Marketing Science Marketing Educator of the Year. As a popular guest speaker, Professor Hair often presents seminars on research techniques, multivariate data analysis, and marketing issues for organizations in Europe, Australia, China, India, and South America. He has a forthcoming book on Sales Analytics (Sage 2026).

Jan-Michael Becker is a Professor in the Department of Marketing at the BI Norwegian Business School (Norway). His research interests and expertise focus on the digital transformation of marketing strategy and consumer behavior as well as marketing analytics, behavioral research methods, causal inference, machine learning, and computational statistics. His research has been published in several premier academic journals, including Journal of the Academy of Marketing Science (JAMS), International Journal of Research in Marketing (IJRM), Information Systems Research, MIS Quarterly, Psychometrika, Nature Human Behavior, Multivariate Behavioral Research, Journal of Business Research, and Marketing Letters. He is a co-developer and co-founder of the statistical software SmartPLS (www.smartpls.com). More information: https://www.bi.edu/about-bi/employees/department-of-marketing/jan-michael-becker/.

Christian M. Ringle is a Chaired Professor and the Director of the Institute of Management and Decision Sciences at the Hamburg University of Technology (Germany), an Adjunct Professor at the James Cook University (Australia), a Visiting Guest Researcher at the University of California, Berkeley (USA), and an RHB-UKM Endowment Fund Distinguished Chairholder at the National University of Malaysia (UKM). His research focuses on management, marketing, technology and innovation management, method development, business analytics, machine learning, artificial intelligence, and the application of business research methods to decision making. His contributions have been published in journals such as Decision Sciences, European Journal of Marketing, International Journal of Research in Marketing, Information Systems Research, Journal of Business Research, Journal of Service Research, Journal of the Academy of Marketing Science, Long Range Planning, MIS Quarterly, and Organizational Research Methods. Since 2018, Christian has been included in the Clarivate Analytics' list of Highly Cited Researchers (HCR). He is a co-developer and co-founder of the statistical software SmartPLS (https://www-smartpls.com). More info: https://www.tuhh.de/mds/team/prof-dr-c-m-ringle.