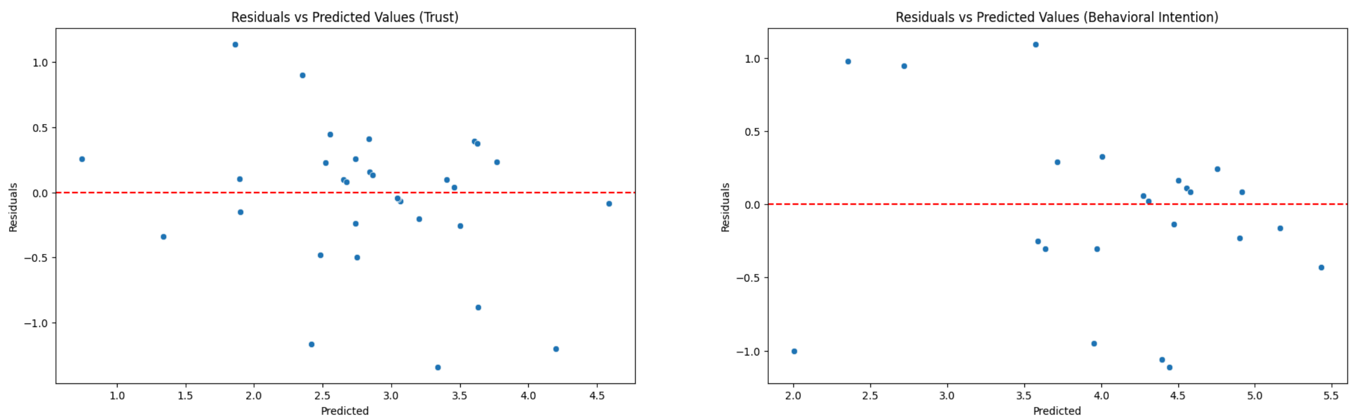
In the following, we address concerns related to model validity for interested readers. Specifically, we test for model overfitting, and unobserved heterogeneity, detailed as follows:

1. Overfitting analysis: Given that high R^2 and f^2 values might raise concerns about potential overfitting (where the model inadvertently learns noise from the training data, compromising its ability to predict unseen data) [1], we tested for model overfitting, adhering to recommended approaches presented in previous studies [2]. The associated code can be found in the supplemental package.

First, we analyzed model residuals to assess the appropriateness of the model fit. Residuals represent the differences between the observed and predicted values for dependent variables and are expected to be randomly scattered around 0 in a well-specified model. This suggests that the errors are random and the model is correctly specified. Figure 1 presents the plots for the residuals against the predicted values for the `trust' and `behavioral intention' constructs. The residuals are indeed randomly scattered, suggesting that the model errors are random.



(a) (b)

Figure 1: *Residuals vs predicted values plot: (a) Trust, (b) Behavioral intention.*

Next, we triangulated the lack of overfitting using train-test split and cross-validation [3]: (1) We performed an 80-20 split, where the model was trained on 80% of the data and tested on the remaining 20%. The performance was assessed using the mean squared error (MSE), where low values indicate a good fit. The MSE values for the trust and intention constructs were 0.317 and 0.381, respectively, both of which fall within the accepted MSE range. (2) We implemented k-fold cross-validation with k set to 5, which is in line with recommended practices in prior studies [2, 3]. The data was divided into 5 subsets, and the model was trained and tested 5 times, each time on a different subset of the data. The performance was again assessed using MSE. The mean MSE for cross-validation were 0.323 and 0.377 for the trust and intention constructs, respectively. These values are close to the test MSE, suggesting that the model does not overfit the data [2,3]. Overall, the analyses suggest that our model is generalizing effectively to unseen data, confirming that model overfitting is not a concern in our study.

2. Unobserved heterogeneity: While control factors capture observed heterogeneity, latent classes of respondents, can also exist, potentially threatening result validity [4]. These latent classes represent groupings based on unmeasured criteria, leading to possible differences in responding across groups. For example, there might be unobserved differences in the model due to heterogeneity between the two organizations where the study was conducted. We used Becker's approach [5] to detect the presence of unobserved heterogeneity in our model. First, dividing the sample size of 238 by the minimum sample size of 95 (see Sec. III-C), yielded a theoretical upper bound of two segments. Next, we ran the FIMIX algorithm [4] in SmartPLS for one (treating the original sample as a single segment) and two segments and compared the results using multiple retention criteria (see Table 1). The optimal number of segments is indicated by the lowest values for each criterion (italicized in Table 1), except for the criterion `EN', where higher values indicate better separation.

Sarstedt et al. [6] recommend starting the fit analysis by jointly considering modified Akaike's Information Criterion with factor 3 (AIC3) and Consistent AIC (CAIC) (Group 1 in Table 1). The result is most appropriate when both criteria suggest the same number of segments.[1] In our case, AIC3 suggests two segments, while CAIC suggests one. Therefore, the next step in the fit analysis is to assess the modified Akaike's Information Criterion with factor 4 (AIC4) and Bayesian Information Criterion (BIC) (Group 2 in Table 1). In our case, both criteria converge, indicating that the one-segment solution is the most appropriate. This confirms the absence of unobserved heterogeneity, meaning there are no subgroups in the sample that are not captured in the model, further confirming the appropriateness of the current solution.

Table 1: FIMIX - Optimal number of segments

Group	Criterion	**1-Segment**	2-Segment
1	AIC3	1483.82	*1480.38*
	CAIC	*1542.85*	1601.91
2	**AIC4**	***1500.81***	1515.38
	BIC	***1525.85***	1566.91
3	EN	0	0.32

**EN measures the degree of separation between segments. A single segment always has an EN value of 0, and thus is not considered in this case.*

REFERENCES

[1] Hawkins, D. M. (2004). The problem of overfitting. Journal of chemical information and computer sciences, 44(1), 1-12.

[2] Russo, D. (2024). Navigating the complexity of generative ai adoption in software engineering. ACM Transactions on Software Engineering and Methodology.

[3] Kohavi, R. (1995, August). A study of cross-validation and bootstrap for accuracy estimation and model selection. In Ijcai (Vol. 14, No. 2, pp. 1137-1145).

[4] Sarstedt, M., Ringle, C. M., & Hair, J. F. (2017). Treating unobserved heterogeneity in PLS-SEM: A multi-method approach. Partial least squares path modeling: Basic concepts, methodological issues and applications, 197-217.

[5] Becker, J. M., Rai, A., Ringle, C. M., & Völckner, F. (2013). Discovering unobserved heterogeneity in structural equation models to avert validity threats. MIS Quarterly, 665-694.

[6] Sarstedt, M., Becker, J. M., Ringle, C. M., & Schwaiger, M. (2011). Uncovering and treating unobserved heterogeneity with FIMIX-PLS: which model selection criterion provides an appropriate number of segments?. Schmalenbach Business Review, 63, 34-62.