Detecting Spurious Moderation Effect in Service Research:

An Information-Theoretic Approach

# Introduction

Marketing researchers typically conducted moderation analysis to assess the presence of a moderator variable defined as the variable that affects the relationship between a predictor and dependent variable. That is, when one is hypothesizing about the presence of a moderator variable $(Z)$ and found it to be statistically significant, one can conclude that the relationship between a predictor variable$(X$) and dependent variable$(Y$) will depend on the level of the moderator variable. If such situation occurs, the research will make a conclusion regarding the presence of a moderation or interaction effect. Moderation analysis are typically conducted to improve either newly discovered relationships or existing theories by establishing boundary conditions of such relationships or theories. The accessibility of publications (e.g., books, academic papers) on its conceptual foundations (e.g., Aiken & West, 1991), and availability of user-friendly statistical computer programs (e.g.,Hayes, 2012), had made the use of moderation analysis continues to proliferate in publications in the service-related journals (e.g., *Journal of Service Research*, *Journal of Services Marketing*). The ubiquitous of moderation analysis can also be found in publications in various domains of social sciences such as psychology, health, tourism and management. For instance, using the keyword either ‘interaction’ or ‘moderation’, of the 7 articles published in the Vol 20, issue 4 of the Journal of service research, 5 of them (>70%) tested moderation effect hypotheses. Despite the extant empirical research, however, hardly any papers asses the possibility that the interaction effect could be confounded by curvilinear effect (Ganzach, 1997; Lubinski & Humphreys, 1990). The present research suggests a method that can help researchers in preventing this potential pitfall in detecting true moderation effect. Thus, this research aims at contributing to the improvement of the procedure and practice of moderation analysis. That is, we propose the use of information theoretic approach (Burnham & Anderson, 2003) as an additional indispensable research tool that should be included in the repertoire of researchers who want to conduct moderation analysis.

# Moderation analysis and information-theoretic approach

*Moderation Analysis*

In testing moderation effect hypothesis, moderation is mathematically captured by an interaction term in the regression equation. One of the challenges in testing the moderation effect is the presence of curvilinear effect and multicollinearity between a predictor and moderator variable (Ganzach, 1997; Lubinski & Humphreys, 1990). That is, when the relationship between a predictor and the dependent variable is curvilinear (e.g., quadratic), researchers can make an erroneous conclusion that a moderation effect exists when, in fact, it does not.

Past papers have recommended that researchers should enter quadratic terms in the moderation regression equation and tested it either via join F-test of t-tests (Cortina, 1993; Ganzach, 1997). If the tests accept the null hypothesis of no quadratic effect, then the significant interaction term is unlikely to be spurious. Unfortunately, marketing researchers have not yet embraced this recommendation. In fact, most of moderation papers in the last issue of the *Journal of Service Research*, never reported this procedure.

The present article introduces the information-theoretic approach in moderated regression analysis. This approach utilizes the Akaike Information Criterion and other information derived from the criterion that can help researchers to assess the quality of their moderated regression model among set of plausible candidate models or also called a multi-model inference. While F-test are useful in testing the significance of quadratic effect, it does not give more information regarding the quality of the model among other plausible competing models, which might not present a serious contradiction to theory being tested.

## Information-theoretic approach

The information-theoretic approach is a relatively new class of approach in inference (Burnham & Anderson, 2003), which does not use the traditional null hypothesis testing by means of p-value. Burnham and Anderson (2003) provides an excellent statistical foundation of this method. This approach uses ‘information loss’ that are possessed by each model being considered as the basis of making valid inference about which model is going to be selected. The information loss refers to the loss of information that occurs when an estimation model is used to approximate the reality. A model with minimum information loss is selected as the best model among sets of candidate model. The heart of the approach lies on the use of information criterion, such as the Akaike information criterion (AIC) and/or its small sample corrected version as a measure of information loss. In practice, researchers compute AIC for each of the candidate models and select the model with the smallest AIC value. Comparing AIC values in model selection is not new in testing moderation effect in marketing where researchers typically compare a structural equation model with and without an interaction term based on AIC values. However, theory-proponent researchers were too restrictive in considering these two models only, neglecting other data-driven possible formulation (e.g., model that includes quadratic terms). In this paper, as the focus is on controlling for the presence of curvilinear effect, we recommended that researchers should consider not only a model with and without interaction term but also several models that are plausible. As the information-theoretic approach dictates, the key decision is indeed, in selecting what models to be included in the plausible set. Fortunately, for moderated regression, the choice of models to be included in the set is constrained by the structure of the model (i.e., by varying the coefficient of the linear additive model) and can be informed by models reported in the domain-related literature. Thus, this makes information-approach is appealing. In the next section, we provide an illustration.

**Illustration and discussion**

We illustrate, by use of simulation, how information-theoretic approach can be used as additional tool in moderation analysis. We generate data (N=300) according to this equation, $Y=X+X^{2}+ε, ε\~N\left(0,1\right)$, where $X$ and moderator $Z$ are sampled from multivariate normal distribution having 0.2 correlation. Estimating this relationship with linear and interaction term only will result in a significant interaction effect. As we use an information-theoretic approach, in the next step, we estimate eight candidate plausible models that include a linear model, purely quadratic models, model with linear and interaction term, quadratic models with interaction term and two models with linear, quadratic, interaction and. These formulation were developed on the basis of the literature review of the commonly reported models in published papers and derivation from the linear additive model (i.e. model 1). Among the eight candidate models, model 7 and 8 were derived from varying coefficient of the regression of the linear additive model (i.e., the coefficient as a linear and quadratic function of the independent variable).

We presented the results of our simulation in Table 1.

 Table 1 provides information regarding the quality of the 8 models, which includes the information about the number of parameters in each model, the residual sum of squares, AIC values, distance and weight. These values are common statistics reported in information-theoretic study. Using the D < 4 criterion ((Burnham & Anderson, 2003), model 2, 5, 6, 7 and 8 should be considered as models that are mostly plausible, which fit the data well compared to the other candidate models under consideration. From Table 1, it is obvious that the model with linear and interaction term, which is typically conducted and recommended in moderation regression model in marketing and service research is among the least plausible model, thus it should not be selected for further investigation. Given these findings and researchers can continue to testing the significance of the interaction terms that appear in model 6, 7 and 8 (e.g., using F or t-test). Subsequent analysis show that the interaction term in those models are not significant. Thus, given that the true relationship in the true model is curvilinear, using information- theoretic approach saves researchers from making a type-I error, while the traditional moderated regression model using parametric approach will lead to significant moderation effect. The statistics reported in Table 1 can be reproduced by using an SPSS macro called ModLR, which can be downloaded freely from Internet.

**Table 1. Model comparisons**

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Model\*** | **K** | **RSS** | **AIC** | **D** | **Weight** |
| $$1. y=x+z$$ | 4 | 971.439 | 1212.740 | 418.996 | 0.000 |
| $$2. y=x+x^{2}$$ | 4 | 242.823 | 796.807 | 3.063 | 0.092 |
| $$3. y=z+z^{2}$$ | 4 | 1118.743 | 1255.095 | 461.351 | 0.000 |
| $$4. y=x+z+xz$$ | 5 | 913.361 | 1196.246 | 402.502 | 0.000 |
| $$5. y=x^{2}+z^{2}+x+z$$ | 6 | 237.173 | 793.744 | 0.000 | 0.424 |
| $$6. y=x^{2}+z^{2}+x+z+xz$$ | 7 | 236.668 | 795.104 | 1.360 | 0.215 |
| $$7. y=x^{2}+z^{2}+x+z+xz+zx^{2}$$ | 8 | 235.584 | 795.727 | 1.984 | 0.157 |
| $$8. y=x^{2}+z^{2}+x+z+xz+xz^{2}$$ | 8 | 236.127 | 796.417 | 2.673 | 0.112 |
| \*Note: for simplicity, we did not include the parameters of the model and an error term in each equation. Because of the relatively small sample size, AIC correction was used. |

**Conclusion**

This paper focuses on curvilinear effect that can influence the significance of the interaction term in moderated regression. This paper seeks to contribute to the improvement of researcher’s methodological repertoire in conducting moderation analysis in marketing studies in general and services in particular. We recommended a two-step approach as illustrated in the previous section. We point out that the use of multi-model inference using the information-theoretic approach can help researchers to evaluate the quality of their moderated regression model in lieu of alternative plausible models, while at the same time preventing them to make an erroneously conclusion regarding interaction due to the spurious effect that stems from the presence of curvilinearity effect. The multi-model inference based on a simple moderation model has been integrated into a freely user-friendly SPSS macro, ModLR, which can assist researchers in conducting moderation analysis. We hope that this paper will increase researchers’ awareness on the potential shortcoming of curvilinear effect in testing moderation hypotheses and stimulate researchers’ to consider the use of information-theoretic approach and include this approach in their research toolbox.

# References

Aiken, L. S., & West, S. G. (1991). *Multiple regression: Testing and interpreting interactions*: Sage.

Burnham, K. P., & Anderson, D. (2003). *Model selection and multi-model inference*. USA: Springer.

Cortina, J. M. (1993). Interaction, Nonlinearity, and Multicollinearity: implications for Multiple Regression. *Journal of Management, 19*(4), 915-922. doi:10.1177/014920639301900411

Ganzach, Y. (1997). Misleading interaction and curvilinear terms. *Psychological Methods, 2*(3), 235-247. doi:10.1037/1082-989X.2.3.235

Hayes, A. (2012). Process SPSS macro. *Computer Software and Manual]. Available from* [*http://www*](http://www)*. afhayes. com/public/process. pdf*.

Lubinski, D., & Humphreys, L. G. (1990). Assessing spurious" moderator effects": Illustrated substantively with the hypothesized (" synergistic") relation between spatial and mathematical ability. *Psychological bulletin, 107*(3), 385.