International Journal of Research in Marketing 33 (2016) 172–182

Contents lists available at ScienceDirect

Intern. J. of Research in Marketing

journal homepage: www.elsevier.com/locate/ijresmar

Full Length Article

Diagnosing harmful collinearity in moderated regressions: A roadmap☆

Pavan Rao Chennamaneni [a,1], Raj Echambadi [b,*], James D. Hess [c,2], Niladri Syam [d,3]

[a] Department of Marketing, University of Wisconsin-Whitewater, Whitewater, WI, 53190-1790, United States
[b] Department of Business Administration, University of Illinois at Urbana-Champaign, Champaign, IL 61820, United States
[c] Department of Marketing and Entrepreneurship, University of Houston, Houston, TX 77204, United States
[d] Department of Marketing, Trulaske College of Business, University of Missouri, Columbia, MO 65211, United States

ARTICLE INFO

Article history:
First received on November 21, 2012 and was under review for 10½ months.
Available online 11 September 2015

Area Editor: Els Gijsbrechts

Keywords:
Collinearity diagnostics
Variance inflation factors
Condition indices
Moderated models
Multiplicative interactions

ABSTRACT

Collinearity is inevitable in moderated regression models. Marketing scholars use a variety of collinearity diagnostics including variance inflation factors (VIFs) and condition indices in order to diagnose the extent of collinearity in moderated models. In this paper, we show that VIFs are likely to misdiagnose the extent of collinearity problems by conflating lack of variability (small variance) or lack of magnitude (small mean) in data for collinearity. Condition indices accurately diagnose collinearity, however, they fail to identify when collinearity is actually harmful to statistical inferences. We propose a new measure, C^2, based on raw data, which diagnoses the extent of collinearity in moderated regression models. More importantly, this C^2 measure, in conjunction with the t-statistic of the non-significant coefficient, can indicate adverse effects of collinearity in terms of distorting statistical inferences and how much collinearity would have to disappear to generate significant results. The efficacy of C^2 over VIFs and condition indices is demonstrated using simulated data and its usefulness in moderated regressions is illustrated in an empirical study of brand extensions.

1. Introduction

Moderated regression models are ideal for testing a contingency hypothesis that the effect of one independent variable, say U, is moderated by a second independent variable, say V, by adding a cross product term, U × V, as an additional explanatory variable (Irwin and McClelland 2001). Moderated regressions, which have become a method of choice for marketing scholars to test multiplicative interactions, have been used in a variety of empirical settings including business-to-business (Fang, Palmatier, Scheer, & Li, 2008), consumer behavior (van Doorn & Verhoef, 2011), international marketing (Leenders & Eliashberg, 2011), product development (Cui & O'Connot, 2012), and retailing (Srinivasan & Moorman, 2005). Owing to the potential for strong linear dependencies among the regressors, U and V, and the interaction term, U × V, researchers fear the existence of high levels of collinearity that may lead to flawed statistical inferences. Indeed, a review of influential marketing journals including Journal of Marketing, Journal of Marketing Research, Marketing Science, and International Journal of Research in Marketing, for the period, 2005–2015, shows that 83 papers that used moderated regressions expressed concerns about collinearity.

The challenges faced by marketing researchers when confronted with collinearity issues can be illustrated using an example of brand extensions. Literature suggests that consumer attitudes toward the newer brand extensions (Attitude) are determined by

☆ The names of the authors are listed alphabetically. This is a fully collaborative work.

* Corresponding author. Tel.: +1 217 244 4189.
URL's:URL: chennamp@uww.edu (P.R. Chennamaneni), rechamba@illinois.edu (R. Echambadi), jhess@uh.edu (J.D. Hess), syamn@missouri.edu (N. Syam).
[1] Tel.: +1 262 472 5473.
[2] Tel.: +1 713 743 4175.
[3] Tel.: +1 573 882 9727.

http://dx.doi.org/10.1016/j.ijresmar.2015.08.004
0167-8116/© 2015 Elsevier B.V. All rights reserved.

the quality perceptions of the parent brand (Quality), perceptions of how easily capabilities can be transferred between the parent and extension product class, (Transfer), and the interaction between Quality and Transfer (c.f. Aaker & Keller, 1990). In order to test whether these relationships are significantly different from zero, the researcher fits the following moderated regression model[4]:

$$\text{Attitude} = \delta_0 + \delta_1\text{Quality} + \delta_2\text{Transfer} + \delta_3\text{Quality} \times \text{Transfer} + \varepsilon. \tag{1}$$

Given that the multiplicative interaction term, Quality × Transfer, is constructed from the regressors, Quality and Transfer, the first question that confronts the researcher is whether the data are indeed plagued by collinearity and if so, the nature and severity of the collinearity. The researcher surveys the extant marketing literature and finds that the two rules of thumb – values of variance inflation factors (VIFs), which are based upon correlations between the independent variables, in excess of 10, and values of condition indices in excess of 30 – are predominantly used to judge the existence and strength of collinearity.

Low correlations or low values of VIFs (less than 10) are considered to be indicative that collinearity problems are negligible or non-existent (c.f. Marquardt, 1970). However, VIF is constructed from squared correlations, VIF ≡ $1/(1-R^2)$, and because correlations are not faithful indicators of collinearity, using traditional rules of thumb for VIF values may lead to misdiagnosis of collinearity problems.[5] Unlike VIFs, a high condition index (>30) does indicate the presence of collinearity. However, the condition index by itself does not shed light on the root causes, i.e. the offending variables, of the underlying linear dependencies. In such cases wherein collinearity is diagnosed as high by the condition index, it is always good practice to examine the variance decomposition proportions (values greater than 0.50 in any row indicates linear dependencies) to identify the specific variables that contributed to the collinearity present in the data (Belsley, Kuh, & Welsch, 1980).

Reverting back to the brand extension example, let us suppose that the t-statistic for the estimate of the interaction coefficient δ_3 is found to be 1.60, considerably below the critical value 1.96. In such a situation, the researcher faces the second critical question: did collinearity adversely affect the interaction coefficient in terms of statistical significance? While variance decomposition metrics help identify the specific variables underlying the potential near-linear dependencies, they do not inform whether collinearity adversely affected the significances of the variables. In other words, none of the current collinearity metrics including VIFs, condition indices, and variance decomposition proportions shed any light into assessing whether collinearity is the culprit that caused the non-significance of the interaction coefficient.

If there was a way to confirm collinearity as the culprit behind the interaction variable's non-significance, the researcher is confronted with a third question: if collinearity in the data could be reduced in some meaningful way through collection of new or additional data, would the measured t-statistic increase enough to be statistically significant? Alternatively, suppose that the data on Quality and Transfer were constructed from a well-balanced experimental design, rather than from a survey, would this experiment lead to a reduction of collinearity sufficient enough to move the interaction effect to statistical significance? An answer to this question would enable the researcher to truly ascertain whether new data collection is needed or whether the researcher needs to focus her efforts elsewhere to identify the reasons for insignificant results. Unfortunately, existing metrics for collinearity diagnosis including correlations, VIF, condition index, or variance decomposition proportions do not provide any insight into this issue.

In this paper, we propose a new measure of collinearity, C^2 that reflects the quality of data to remedy the above mentioned problems. C^2 not only helps diagnose the existence of collinearity but also indicates whether collinearity was the reason behind non-significant effects. More importantly, C^2 can also indicate whether a non-significant effect would have become significant if the collinearity in the data could be reduced, and if so, how much collinearity must be reduced to achieve this significant result.

2. Collinearity in moderated regression

Consider a moderated variable regression.

$$Y = \alpha_0\mathbf{1} + \alpha_1 U + \alpha_2 V + \alpha_3 U \times V + \nu, \tag{2}$$

where U and V are ratio scaled explanatory variables in N-dimensional data vectors.[6] Correlations refer to linear co-variability of two variables around their means. Computationally, correlation is built from the inner product of mean-centered variables. Collinearity, on the other hand, refers to the presence of linear relationships between variables (Silvey, 1969). As noted by Belsley (1984), collinearity diagnostics should be based on the analysis of linear dependencies between raw or uncentered values.

By construction, the term U × V carries some of the same information as U and V and could create a collinearity problem, so let us look at how unique the values of the interaction term U × V are compared to **1**, V and V. The source of collinearity is clear but

[4] Unlike the brand extension literature (c.f. Aaker & Keller, 1990; Bottomley & Holden, 2001) that used a comprehensive model with multiple independent variables and interaction terms, we use a restricted model with only two independent variables and one interaction term for expository simplicity.

[5] While high correlations indicate high collinearity, low correlations do not necessarily imply low collinearity. It is easily possible to have zero correlation between variables but have severe collinearity (Belsley, 1991; Ofir & Khuri, 1986).

[6] Bold $\mathbf{1} = (1,1,\ldots,1)'$. This illustrative moderated regression model is generalized to more than two variables in Table 1, where it will be confirmed that as the number of variables increases, the number of ways in which collinearity can occur increases (Belsley, 1991).