Moderating Effects:
The Myth of Mean Centering

by
H. Gatignon*
and
J. Vosgerau**

2005/30/MKT

Moderating Effects: The Myth of Mean Centering

Hubert Gatignon

and

Joachim Vosgerau*

April 5, 2005

* Hubert Gatignon is the Claude Janssen Chaired Professor of Business Administration and Professor of Marketing at INSEAD, Boulevard de Constance, 77305 Fontainebleau Cedex, France. Tel: +33 (0)1 60 71 26 28, Fax: +33 (0)1 60 74 55,
E-mail: Hubert.Gatignon@insead.edu.
Joachim Vosgerau is a Doctoral candidate in Marketing at INSEAD, Boulevard de Constance, 77305 Fontainebleau Cedex, France. Tel: +33 (0)1 60 72 90 49, Fax: +33 (0)1 60 71 26 80, E-mail: Joachim.Vosgerau@insead.edu.

Moderating Effects: The Myth of Mean Centering

Abstract:

Mean centering of variables is often advocated for estimating moderated regressions to reduce the multicollinearity that results from introducing the product term of two variables (x_1x_2) as an independent variable in the regression equation. We establish that, contrary to conventional wisdom, mean centering does not reduce multicollinearity at all, even if the bivariate correlation between x_1 and x_1x_2 is decreased. Using Monte Carlo simulation, we also address the current admonition to systematically include all the product-term components in the moderated regression model; we conclude that such an automatic full-model specification is ill advised due to the structural multicollinearity it introduces. Finally, we propose a varying parameter model (VPM) to test more effectively moderating effects and show under what conditions VPM outperforms OLS estimations.

Moderating Effects: The Myth of Mean Centering

The development of theories in many of the social sciences and in marketing in particular requires the use of contingencies, such that one variable influences (moderates) the manner in which another variable exerts its impact on a criterion (dependent) variable of interest. For example, Bowman and Narayandas (2004) find that B2B customers' satisfaction with a vendor leads to stronger loyalty the smaller the customer is. To test such hypothesized moderating effects, a product term between the focal factor and the moderator variable is created in regression models (Saunders 1956). The issues involved when introducing a term that is the product of two variables have been discussed since the early 1980s (Sharma, Durand and Gur-Arie 1981). A consensus has developed that recommends mean-centering variables in order to reduce multicollinearity (Aiken and West 1991; Cohen and Cohen 1983; Jaccard, Turrisi and Wan 1990; Jaccard, Wan and Turrisi 1990; Smith and Sasaki 1979). This consensus follows Cronbach (1987)'s observation that the correlations between the original mean-centered variables and their product term are much smaller than the correlations among the raw variables and their product term. This method appears to be pervasive in social science research in general (Kromrey and Foster-Johnson 1998) and in Marketing studies in particular: 72% of the articles from 1991 to 2005 in the *Journal of Consumer Research*, the *Journal of Marketing*, the *Journal of Marketing Research*, *Marketing Science*, and *Management Science* that diagnose multicollinearity problems use mean-centering as a solution.

In this paper we show that mean-centering as a remedy of multicollinearity is a myth: as suggested in the econometrics literature, mean centering does not make any difference in terms of multicollinearity (Besley 1984). That is because Cronbach's conclusion that the transformed variables are "almost certain not to be multicollinear (p. 414)" ignores the multicollinearity that remains in the full set of explanatory variables. We then analyze how researchers *can* minimize the consequences of multicollinearity. First, we investigate the behavior of the parameters impacted by the inherent collinearity of the data and the factors that drive these consequences.[1] McClelland and Judd (1993) provide some evidence that, surprisingly, the power to detect interactions *increases* with the collinearity among the predictor variables. However, their analysis is only performed when such interactions do exist and both the product term and its components are specified in the regression equation. Our Monte Carlo simulations show the extent of the consequences of multicollinearity under both conditions where moderating effects exist and where they do not.

We also address the question of model specification with regard to the inclusion of all the components of the product term. This admonition (e.g., Irwin and McClelland 2001) follows from the interpretation of interaction terms in experimental analysis (Cohen 1978) and from the requirement of the nested structure of hierarchical tests (McClelland and Judd 1993). We show there are serious negative consequences (in terms of the reduced significance of the model parameters other than the product term) of specifying a full model, i.e., with the product term (even if insignificant) and all its components, without a theory-based reason to do so.

Finally, we propose that the variable parameter approach to moderator effects provides an easier basis for interpreting the moderated regression coefficients. With this varying-parameter approach, the estimation makes use of the full information available in a moderator effect that exhibits a stochastic component, as opposed to one that imposes a purely systematic and strict restriction on the moderator equation. We show under what conditions this varying parameter approach outperforms typical Ordinary Least Squares estimation in that particular context of moderated regressions.

[1] Although we recognize that other problems may cause difficulty in detecting moderating effects, such as non-linearity of moderating effects (Jaccard, Wan and Turrisi 1990), measurement errors (Jaccard and Wan 1996) or limited residual variance of the interaction term due to constraints on the range of the moderator variable (McClelland and Judd 1993), we focus on the issue of multicollinearity that dominates the literature.

In summary, this paper explains why mean centering is a myth that does not help in better identifying moderator effects. We first show that mean centering does not change the ill-conditioned data that need to be analyzed for estimating moderating effects. We then describe a Monte Carlo simulation design that helps understand the multicollinearity issues involved in moderated regression. In particular, we investigate the antecedents and consequences of multicollinearity, and bring in the notion of stochasticity of the moderator effect. We then present our analysis answering six fundamental questions with important practical implications:

1. What are the chances of inferring the existence of moderating effects when none are present?
2. Does the product term introduce multicollinearity that affects the estimates of the effect of its components?
3. Do the answers to these questions vary with the correlation between the focal and the moderator variable?
4. What is the impact of specifying the moderator variable as a separate constant effect, in addition to being a moderator variable?
5. Are the answers to questions 2 through 4 the same when there are moderating effects in the data?
6. What is the impact of recognizing the existence of a stochastic element in the moderating effect?

In the conclusion, we summarize the implications of our results for conducting empirical tests of moderating effects.

THE MYTH OF THE EFFECT OF MEAN CENTERING ON MULTICOLLINEARITY

The structural relationship between the dependent variable y and the independent variables is typically expressed as a regression model with measures of the variables that present the typical properties of being ratio or interval scale.[2] The typical regression model with two variables is shown in equation 1:

$$y = b_0 + b_1 x_1 + b_2 x_2 + b_3 x_1 x_2 \tag{1}$$

One problem identified with estimating this model concerns the fact that typically the product term $x_1 x_2$ is highly correlated with the individual variables x_1 and x_2. Hence, Cronbach (1987) advised the use of mean-centered variables, based on the fact that the correlation between the product term of the deviations from the mean-transformed variables with these deviations from the means themselves is decreased by a large magnitude. The covariance of x_1 and x_2 can be written as:

$$\mathrm{V}[x_1, x_1 x_2] = \mathrm{V}[x_1 x_2]\mathrm{E}[x_1] + \mathrm{E}[(x_1 - \bar{x}_1)^2 (x_2 - \bar{x}_2)] + \mathrm{E}[x_2]\mathrm{V}[x_1] \tag{2}$$

It becomes clear that when the mean-centered variables $x_1^d = x_1 - \bar{x}_1$ and $x_2^d = x_2 - \bar{x}_2$ are used, the expression above is reduced to:

$$\mathrm{V}[x_1^d, x_1^d x_2^d] = \mathrm{E}[(x_1 - \bar{x}_1)^2 (x_2 - \bar{x}_2)]\,. \tag{3}$$

Because the variances of the raw and the mean-centered variables are equal ($\mathrm{V}[x_1] = \mathrm{E}[(x_1 - \bar{x}_1)^2] = \mathrm{E}[(x_1 - \bar{x}_1 - \bar{x}_1^d)^2] = \mathrm{V}[x_1^d]$), the correlations can easily be compared by comparing the covariances above. The variances being positive by definition, the larger the (typically positive) expected values of the raw variables are (especially the larger the scaling constant of an interval scale variable), the larger will be the correlation. This is reflected in Cronbach's (1987) remark that the product term is unlikely to be predicted by a linear combination of its components.

Based on that observation, many recent studies positing moderator effects have analyzed their data by performing a regression on the mean-centered variables:

$$y = g_0 + g_1 x_1^d + g_2 x_2^d + g_3 x_1^d x_2^d\,. \tag{4}$$

[2] The issues discussed in this paper also apply to cases where the moderator variable is discrete as they are simply a special case of what we present here.

This transformation was recommended by Cronbach based uniquely on the fact that the higher correlations between x_1 and x_1x_2, and/or between x_2 and x_1x_2, are thought to lead to multicollinearity problems, and the model using mean-centered variables "is less vulnerable to computational errors" (Cronbach 1987, p. 416). He does, however, recognize that the residual sum of squares term is exactly the same for the two models (that is, with raw or with mean-centered variables), and that the coefficients of the product terms are identical, as well as their standard errors. It would appear *a priori* logical that if the correlations between the components and the product term of the mean-centered variables are reduced, then the multicollinearity problem found in performing a regression using raw variables would be solved when using mean-centered variables.

However, this is not the case. The product term (interaction) coefficients are identical, and only those coefficients naturally affected by the transformation are different. However, this is not the result of a reduction in multicollinearity as can be seen by using the two-variable example above. Developing Equation (4), the coefficients of the transformed variables can be easily transposed to the coefficients *ß* in Equation (1):

$$y = a_0 + (g_1 - g_3\bar{x}_2)x_1 + (g_2 - g_3\bar{x}_1)x_2 + g_3x_1x_2 \qquad (5)$$

In the Equation above, it is clear that the coefficients of x_1 and x_2 in Equations (1) and (4) are not the same.

Multicollinearity is a particular problem when the parameter estimates are sensitive to small changes in the data. For a typical linear regression of the form $\mathbf{y} = \mathbf{X}b + \mathbf{u}$, the best linear estimator is $\hat{b} = \mathbf{Ay}$ where $\mathbf{A} = [\mathbf{X'X}]^{-1}\ \mathbf{X'}$. The key element of stability is the ability to perform the matrix inversion. From a computational point of view (also taken by Cronbach), the problem occurs when estimating parameters if the determinant of the X'X matrix is close to zero for the derivation of the inverse of **X'X**. It turns out that the explanation of multicollinearity reduction (as suggested by Cronbach) does not follow here because the determinants of the matrix of cross products of the independent variables in the two equations are equal. Because other diagnostic indices of collinearity can be affected by the data transformation as demonstrated by Belsley (1984), such diagnostics (e.g. the variance inflation factor) are meaningless and can be misleading after mean centering. We computed two covariance matrices—for raw-variables and for mean-centered variables—on 57600 data sets in which the correlation between x_1 and x_2 was manipulated on 5 levels (0, 0.2, 0.4, 0.6, and 0.8; see Monte Carlo simulations below), and compared their respective determinants in Equation (6) and (7):

$$|\mathbf{X}'\mathbf{X}| = \begin{vmatrix} T & T\bar{x}_1 & T\bar{x}_2 & T\bar{x}_{12} \\ T\bar{x}_1 & \mathbf{x}_1'\mathbf{x}_1 & \mathbf{x}_1'\mathbf{x}_2 & \mathbf{x}_1'\mathbf{x}_{12} \\ T\bar{x}_2 & \mathbf{x}_2'\mathbf{x}_1 & \mathbf{x}_2'\mathbf{x}_2 & \mathbf{x}_2'\mathbf{x}_{12} \\ T\bar{x}_{12} & \mathbf{x}_{12}'\mathbf{x}_1 & \mathbf{x}_{12}'\mathbf{x}_2 & \mathbf{x}_{12}'\mathbf{x}_{12} \end{vmatrix} \tag{6}$$

$$|\mathbf{X}^{d\prime}\mathbf{X}^d| = \begin{vmatrix} T & 0 & 0 & T(\bar{x}_{12}-\bar{x}_1\bar{x}_2) \\ 0 & \mathbf{x}_1'\mathbf{x}_1 - T\bar{x}_1 & \mathbf{x}_1'\mathbf{x}_2 - T\bar{x}_1\bar{x}_2 & \mathbf{x}_{12}'\mathbf{x}_1 - \bar{x}_2\mathbf{x}_1'\mathbf{x}_1 + 2T\bar{x}_1^2\bar{x}_2 - 2T\bar{x}_1\bar{x}_{12} \\ 0 & \mathbf{x}_2'\mathbf{x}_1 - T\bar{x}_1\bar{x}_2 & \mathbf{x}_2'\mathbf{x}_2 - T\bar{x}_2 & \mathbf{x}_{12}'\mathbf{x}_2 - \bar{x}_1\mathbf{x}_2'\mathbf{x}_2 + 2T\bar{x}_1\bar{x}_2^2 - 2T\bar{x}_2\bar{x}_{12} \\ T(\bar{x}_{12}-\bar{x}_1\bar{x}_2) & \mathbf{x}_{12}'\mathbf{x}_1 - \bar{x}_2\mathbf{x}_1'\mathbf{x}_1 + 2T\bar{x}_1^2\bar{x}_2 - 2T\bar{x}_1\bar{x}_{12} & \mathbf{x}_{12}'\mathbf{x}_2 - \bar{x}_1\mathbf{x}_2'\mathbf{x}_2 + 2T\bar{x}_1\bar{x}_2^2 - 2T\bar{x}_2\bar{x}_{12} & \mathbf{x}_{12}'\mathbf{x}_{12} + \bar{x}_2^2\mathbf{x}_1'\mathbf{x}_1 + \bar{x}_1^2\mathbf{x}_2'\mathbf{x}_2 - 2\bar{x}_1\mathbf{x}_{12}'\mathbf{x}_2 - 2\bar{x}_2\mathbf{x}_{12}'\mathbf{x}_1 + 4T\bar{x}_1\bar{x}_2\bar{x}_{12} - 3T\bar{x}_1^2\bar{x}_2^2 \end{vmatrix} \tag{7}$$

The determinants of these theoretically derived covariances are identical in all data sets. Consequently, it is inappropriate to conclude that mean centering reduces the multicollinearity problem. In fact, the statistical analyses are completely identical in all aspects. The reason to choose mean-centered variables versus raw variables in moderator regression cannot be found in the statistical properties of two statistically equivalent models. Mean-centering transformations change the nature of the source for multicollinearity from bivariate collinearity to multivariate collinearity without changing its magnitude. This was recognized by Belsley (1984), who advises against mean centering as a way to handle multicollinearity because it "obscures any linear dependence that involves the constant term"

(Judge et al. 1985, p. 903), that is, the inter-variable correlations are reapportioned.[3] Although the constant term is not typically of interest to researchers, it cannot be removed for the estimation and, therefore, it impacts the estimates of all other coefficients. And as the determinant applies equally to the standard errors of all the coefficients being estimated, it does not solve the estimation problems of any particular coefficient. Actually, it would be surprising if mean centering solved the multicollinearity problem when the R^2s, F-values, the statistics for the expected value and the significance of the interaction parameter are exactly identical (Kromrey and Foster-Johnson 1998; Dunlap and Kemery 1987).[4]

In the next section, we describe a Monte Carlo simulation to analyze the role of multicollinearity in making statistical inference in moderated regression models.

STATISTICAL INFERENCE ABOUT MODERATOR EFFECTS: A MONTE CARLO SIMULATION

Now that we have established that mean centering has no impact whatsoever on the statistical inference of models with moderator effects, it is important to analyze the nature of the problems associated with estimating such models.[5]

Before elaborating on some of the details of the simulation design, it is useful to identify a key feature of a moderator effect which has been ignored in the prior literature: the stochastic nature of the moderating equation. In a moderator regression, or Moderated Multiple Regression (MMR), the coefficient representing the effect of a focal variable x_1 is

---

[3] This is clearly shown with the 'Principal Component Regression' procedure proposed by Morris, Sherman, and Mansfield (1986). However, we agree with Cronbach's (1987) evaluation that this procedure is ill-advised as no legitimate test for interactions is possible after removing the information contained in the component associated with the small third eigenvalue.

[4] Even though the coefficients $\hat{g}_1$ and $\hat{g}_2$ of the mean centered variables x_1^d and x_2^d may show reduced standard errors, they are perfectly equivalent to the estimates obtained from the raw variables: the calculated corresponding parameters (i.e., interpretable when x_1 and x_2 are valued at zero) and their standard errors are identical (Cohen 1978).

[5] We use raw variables in the following analysis to facilitate the interpretation of the coefficients. The results are similarly valid for mean-centered variable models.

not constant but varies according to the level of a moderator variable x_2. Let us express this ***response equation*** for a single observation i:

$$y_i = b_0 + b_{1i}x_{1i} + u_i \quad (8)$$

The only distinction of this Equation from a standard regression model is the subscript i associated with coefficient $ß_{1i}$. Indeed, we now introduce a second equation (the ***moderating equation***) expressing the moderating role of x_2 on the effect of x_1:[6]

$$b_{1i} = a_0 + a_1 x_{2i} + e_i \quad (9)$$

Substituting the expression in Equation (9) into Equation (8) leads to:

$$y_i = b_0 + (a_0 + a_1 x_{2i} + e_i) x_{1i} + u_i \quad (10)$$

By developing, the equation becomes

$$y_i = b_0 + a_0 x_{1i} + a_1 x_{1i} x_{2i} + u_i + e_i x_{1i} \quad (11)$$

If there were no random term in equation (9), Equation (11) would reduce to Equation (1), except for the term specifying the effect of x_2 on y, which is usually added to Equation (8) and which leads to issues analyzed later in this study.

The coefficient a_1 in Equation (9) is directly interpretable in terms of a moderating effect of x_2, as it indicates by how much the effect of x_1 (i.e., $ß_1$) changes when x_2 changes by one unit. It should be noted, however that mean centering affects the interpretation of the coefficients of the components x_1 and x_2. In the raw-variable model, the correct interpretation is the value of the coefficient when the moderator variable takes the value zero. In the mean-centered model, the correct interpretation is the value of the coefficient of the focal variable when that transformed variable is set to zero, i.e., at the mean value of x_2. Because of the varying nature of the effect of x_1 on y, we use the term "***constant effect***" for a_0, as it

[6] Although we introduce a single moderator equation affecting a coefficient of the response equation, the approach can easily incorporate multiple moderating factors on multiple coefficients as well as non-linear moderating effects (e.g. Gatignon 1993, 1984).

represents the ***constant portion of the effect*** of that variable; it also denotes the constant value of the effect of x_1 on y when the moderator variable is not significant.

The conceptual role of the random term in the moderator equation

How useful is the random element in Equation (9)? So far, we have discussed the interpretation of the coefficients in the moderator equation. However, the conceptual role of this random element can also be critical. It fundamentally expresses the stochastic nature of the moderating effect. This is appropriate when the theoretical basis for the moderating effect of x_2 is not exclusive of all other possible variables. The disturbance term e_i recognizes that this equation is not fully deterministic. So, what is the difference between a model with and one without an error term in the moderating effect? Conceptually, the model without the stochastic element is a strictly deterministic model of the moderating effect. In contrast, the fully specified model with a random term allows more flexibility in the constraint on the specification of the marginal coefficient of x_1. It is also more compatible with theory, in that theories do not typically argue that a particular variable is the only moderating factor (and, therefore, that the moderating effect is completely deterministic). This would correspond to a pure restriction on the parameter representing the effect of x_1 on y. In fact, this position would differ from classical statistical inference theory where the existence of random departures from a deterministic model is required as a source of unexplained variance. Also, as pointed out by Gatignon and Hanssens (1987), the stochastic element provides information that may discriminate between the moderating role of x_2 on x_1 and that of x_1 on x_2. This can be seen from the fact that Equation (11) would become, in the case of x_1 moderating the effect of x_2:[7]

$$y_i = b_0 + b_1 x_{1i} + d_2 x_{2i} + d_1 x_{1i} x_{2i} + u_i + e_i x_{2i} \qquad (12)$$

[7] The argument developed by Gatignon and Hanssens (1987) is even more powerful in the context of a non-linear response function associated with a linear functional form for the varying parameter function.

However, the specification of a random term only has an implication if that information is used for the estimation of the model parameters, since it only appears in the error term structure. The error term in Equation (11) now shows heteroscedasticity with $\mathrm{V}[u_i + x_{1i}e_i] = s_u^2 + x_{1i}^2 s_e^2$. This implies that a generalized least squares estimator (EGLS) will be asymptotically more efficient than OLS. Nevertheless, the OLS estimator (that ignores the stochastic element) remains unbiased.

In summary, the random element of a moderator function is essential to fully describe the moderating process being studied and should be an intrinsic part of the model that allows statistical inference about its impact. This stochastic element leads to heteroscedasticity which should be taken into account for efficient estimation. But more importantly, as we will show, it can be a significant help in the context of multicollinearity.

The specific Monte Carlo simulation design

In order to investigate whether the introduction of a moderating term in the estimation model influences the likelihood of finding a significant ***constant*** and ***moderating*** effect, data was generated with and without moderating effects. By estimating a model with a moderating term on these data sets, we can compare the likelihood of finding a significant interaction when in fact there is none (i.e., the moderating effect is set to 0) with the likelihood of finding a significant interaction when there is one (i.e., the moderating effect is set to values greater than 0). In addition to varying the strength of the moderating effect, we also manipulated orthogonally the noise in the regression corresponding to the error term in Equation (8), the noise in the moderating equation corresponding to the error term in Equation (9), the sample size, and the correlation between the two independent variables x_1 and x_2. For each generated data set, a model with a moderating term (Equation (1)) was estimated with Ordinary Least Squares, and a varying parameter model using Equations (8) and (9) was estimated with Estimated Generalized Least Squares.

Data was generated with SAS 9.1 for Windows. In order to simulate data sets with two independent variables and two error components simulating noise in the regression (u_i in equation 8) and noise in the moderating term (e_i in equation 9), four random variables were generated with the *CALL VNORMAL* module (the seed was generated separately for each data set generation using the *RANNOR* function). The first two random variables x_1 and x_2 were generated with a mean of 10 and standard deviation of 1. The latter two random variables representing noise in the regression were generated with a mean of 0 and standard deviation 1. The correlation of the two error-variables with all other variables was set to 0. The correlation between the two independent variables x_1 and x_2 was manipulated at five levels: 0, 0.2, 0.4, 0.6, and 0.8. In order to assure that the resulting sample correlation between the independent variables x_1 and x_2 would never be negative, data generation was repeated until the resulting sample correlation was non-negative (this was necessary only for the case when the correlation between x_1 and x_2 was set to 0, and hence the resulting sample correlation between x_1 and x_2 would take on small positive and negative values). The size of the impact of x_1 and x_2 without moderation (a_1 = 0 in Equation (9)) was fixed at 1 (a_0 and $ß_2$), the intercept was fixed at 4 ($ß_0$ in Equation (8)).

The size of the moderating effect (a_1 in Equation (9)) was manipulated at six levels: 0, 0.2, 0.4, 0.6, 0.8, and 1.0. The first level 0 implies that there is no moderating effect in the generated data. For levels greater than 0 implying a moderating effect, noise within the moderating equation (INTERNOISE) was manipulated at six levels by multiplying the random error variable (e_i in Equation (9)) with the following constants: 0, 0.4, 0.8, 1.2, 1.6, 2.0. This provided moderating equation variances of 0, 0.16, 0.64, 1.44, 2.56, and 4.0 which correspond to explained variance of the moderating effect ranging from 0.99% to 86.2%.

Noise in the regression (REGNOISE) was manipulated at 8 levels by multiplying the random error variable (u_i in Equation (8)) with the following constants: 0.1, 0.7, 1.3, 1.9, 2.5, 3.1, 3.7, and 4.3. This provided error variances of 0.01, 0.49, 1.69, 3.61, 6.25, 9.61, 13.69, and 18.49 corresponding to explained variances in the base case of no moderating effect from 9.76% to 99.5%. Data sets were generated with four differing samples sizes: 50, 175, 300, and 425. At each level of the manipulated factors 10 data sets were generated. Thus, the total number of data sets generated for the 5760 experimental conditions is 57600. For each data set two models with a moderating term were estimated: A model corresponding to Equation (1) was estimated with OLS, and a varying parameter model corresponding to Equations (8) and (9) was estimated with EGLS (details of the estimation method is described in the Appendix). We now report the results according to the six questions mentioned in the introduction.

SIMULATION RESULTS: ANSWERS TO KEY QUESTIONS

We begin with analyzing the results of the simulated data using only the data generated without moderating effects (N=9600).

What are the chances of inferring the existence of moderating effects when none are present?

We focus first on the moderating effect parameter. Given that there is no such effect in the data, the proportion of significant parameters by chance alone would be 5% at the .05 confidence level (Lehmann 2001). Multicollinearity, when it exists in a dataset, normally affects all parameter estimates. However, Ganzach (1997, p. 236) claims that "the coefficient of the product term in the regression may be significant even when there is no true interaction. The reason for this is that when the correlation between X and Z increases, so does the correlation between XZ and X." This would, however, be inconsistent with

McClelland and Judd's (1993) proposition that "given that a change in origin can always be found to ensure a zero covariance between the product and its components and given that such a change of origin does not alter the moderator statistical test, the covariance, if any between the components and their product is in principle irrelevant for detecting moderator effects (p. 378)." So in our case, where no moderating effect exists and correlations between the independent variables x_1 and x_2 range from 0 to 0.8, we should expect 5% of the coefficients to be significant at the .05 level. Figure 1a plots the percentage of estimated moderator effect parameters that are significant at the .05 confidence level. With OLS estimation, the proportion of significant coefficients using a two-tailed test is 5.1%. This percentage is not statistically different from what would be expected by chance (5%).[8] The parameter estimates are unbiased and normally distributed around the zero point (Figure 1.b). Consequently, the model is consistent with the data generating function and the multicollinearity in the data does not affect the likelihood of finding a significant moderator effect when none is in the data.

So, there is no evidence that the product term picks up part of the variance that would be explained by the focal or the constant effect of the moderator variable, which can be the case when two variables are correlated. The inherent correlation between the product term and its components restricts the relationships in the data in a way that the correlation between the product term and its components has little effect on the moderator effect estimate when there is none in the data. Thus, researchers can be reassured that they are unlikely to report moderator effects when there are none.

[8] The critical value at a = .05 is 5.44%.

Does the product term introduce multicollinearity that affects the constant effects?

Is it also the case that the effects of the components of the product term, i.e., the constant effects, are unaffected by multicollinearity introduced by the specification of the product term? Figure 2 provides information with regard to this question. Each bar represents the percentage of the estimated parameters that is significant for a given level of correlation between the focal and the moderator variables. Concentrating on the left side of the graph, the first group of bars from the left represents the intercept $ß_0$ in Equation (1) at each of the five levels of correlation (increasing from left to right starting at 0 until 0.8). The second group next to the right of this first group corresponds to $ß_1$, i.e., the coefficient of x_1. Then comes the group corresponding to $ß_2$, the coefficient of x_2 and finally the fourth group from the left is for the moderator effect $ß_3$ discussed in the question above, shown by subgroup of correlation level rather than aggregated as in Figure 1. The data generating function used for this analysis (represented in these four groups of bars) specified both variables x_1 and x_2 as having constant effects but no moderator effect, while the estimated model specification contains both of these variables as well as their product term. We address separately the question of the impact of the correlation between the focal and the moderator variable. It is noticeable that the significance level of the moderator effect does not vary depending on the correlation between these two variables. However, to answer the question raised above, we concentrate on the second and third group of bars corresponding to the estimates of $ß_1$ and $ß_2$.

Even though the base level is difficult to assess because significance depends on the true size of the effect (set to one for both variables) and on the regression noise (error term variance of the regression), several points can be made. Significance is greater than just by chance at the .05 level, with approximately a 30% chance of being significant. However, this is less than would be generally expected. The same comment can be made about the intercept coefficients, where only 20% of them are significant (while the value was set at 4). The low significance of these effects hints to the existence of a multicollinearity problem that would affect all the coefficients except the moderator effect.

However, this conclusion cannot be made without considering further analyses. In particular, if multicollinearity is the cause of the problem, is it because of the inherent correlation of the product term and its components? This question can be answered by considering the parameter estimates from a constant-only-effect model that does not specify the product term (no moderator effect). The summary statistics of these parameter estimates are provided in Figure 3, where each group of bars from left to right corresponds to the parameters $ß_0$, $ß_1$ and $ß_2$ (the bars within each group corresponds to each level of the correlation between x_1 and x_2). The Figure shows that most of the coefficients are significant (more than 90% of them are significant for both $ß_1$ and $ß_2$ in all but the highest correlation level of 0.8, in which case still more than 80% are significant). Comparing these significance levels with those of the fully specified model, the level is reduced dramatically from 90% to 30% simply by specifying a product term, even when, in fact, there is none and its estimates are insignificant.

Another analysis helps evaluate the poor ability to detect significant constant effects from a model estimated with a moderator effect: how sensitive are the results to changes in the noise level of the regression? Figure 4 provides summary statistics that answer that question. Again, each group of bars from left to right represents the percentage of significant coefficients, but this time, each bar reflects a different level of noise in the regression. Within each group the leftmost bar corresponds to the lowest level of noise in the regression (with the two independent variables explaining 99.5% of the variance of the dependent variable), the second 80%, and the following ones explaining 54.2%, 35.65%, 24.24%, 17.23%, 12.75% and 9.76%, as presented in the data generation section of the paper. As expected, with an almost perfect regression, the parameters are all significant. However, this percentage drops rapidly to the 20% level of significance only for a rather good level of fit compared to what one can expect in practice (slightly above 50% explained variance).

Taken together, these results clearly show that when a product term is introduced for testing for moderator effects, multicollinearity is introduced that impacts all coefficients that are estimated with large variances. The inherent collinearity introduced by the product term has a huge impact on the ability to make any inference about the constant effects. Consequently, it would be prudent to recommend performing the estimation of a model with potential moderating effects in two steps. The first step is to estimate the full model. Then, if the moderator effect is insignificant (which is likely to be the case if there is no such effect, as shown in the first part of our analysis), a second step consists of re-estimating a model without the product term to estimate the constant effects of the two variables. These two steps correspond to the hierarchical test procedure; however, if the coefficient of the product term is significant, coefficients of a model without product term should not be interpreted as they are biased due to model misspecification.

Do the answers to these questions vary with the correlation between the focal and the moderator variable?

Contrary to prior convictions that a high correlation between x_1 and x_2 leads to multicollinearity which "may have an adverse effect on the estimated coefficients" (Morris, Sherman and Mansfield 1986), McClelland and Judd (1993) make the case that the product term correlation with its components has little impact on the ability to detect moderating effects. They demonstrate that, surprisingly, the correlation between the focal variable and the moderator variable *improves* the chances of detecting a moderator effect, due to the properties of the distribution of the product term. They consider, however, only cases when there is a moderator effect in the data. We will treat this case as well, but for the moment, we consider the situation when no such effect exists. We analyze the impact of the correlation between the focal variable and the presumed moderator variable on the significance of the estimated moderator effect parameter. We also analyze the impact that this correlation has on the estimation of the constant effects.

Figure 2 (the leftmost four groups of bars), as described above, presents the percentage of significant coefficients for each level of correlation manipulated from 0 to the left of each group of bars to 0.8 at the right of each group of bars. These graphs show no effect of the correlation on the detection of significant moderator effects that are consistent with the random hypothesis of 5%, regardless of the level of correlation. Again, this result demonstrates the reliability and validity of the tests performed for such moderating effects.

The impact of the correlation on the other parameters, i.e., the constant effects, is however, somewhat positive. As Figure 2 shows, the percentage of significant constant effects increases somewhat with the correlation. Even at a very high correlation level of 0.8, it is not the correlation between the focal and moderator variable that creates multicollinearity and the (consequently) lower level of significance. The effect is small and consistent with the positive impact of the correlation noted by McClelland and Judd (1993) on the moderator effect when such an effect exists. When no such effect exists in the data, the positive benefits of the distribution of the product term do not appear on the moderator effect but on the constant effect that is impacted by the covariance of the product term with its individual variable (including the intercept term).

Interestingly, this phenomenon is not reproduced with the constant-only-effect model, for which the results are given on Figure 3 for each level of correlation. Here, the increase in the correlation reduces the significance level of the constant effects indicating increased multicollinearity. But this effect is relatively small and only marginally significant for the largest discrepancy between the lowest and the highest correlation levels.

In summary, the correlation between the focal and the presumed moderator variable does not impact statistical inference of either the moderating effect or of the constant effects. It is the inherent correlation between the product term and its components that lowers the significance of the constant effect estimates (but not the moderating effect). The problems are therefore inherent to the model specification, but not due to the lack of independence between the focal and the moderator variables. Thus, it is clear that mean centering cannot help at all to remedy this problem, as demonstrated earlier, since it neither affects the multicollinearity in the data nor the multicollinearity inherent to the model specification. Unfortunately, no method exists to remove this multicollinearity inherent to the model specification without losing information and/or without affecting the interpretability of the parameters such as 'Principle Component Regression' (Morris, Sherman and Mansfield 1986) where components with lowest eigen values are removed.

What is the impact of specifying the moderator variable as a constant effect in addition to the moderation effect?

We continue to focus on the impact of multicollinearity in testing moderator effects when there are no such effects in the data. The last question we address in this context concerns the role of specifying a constant effect of the presumed moderator variable in addition to the moderating effect being tested. This has been recommended based on interpretation of interaction term arguments (Cohen 1978; Irwin and McClelland 2001) and on the necessary condition of nested models required for hierarchical testing procedures (McClelland and Judd 1993). However, equations (8) and (9) are clear indications that the coefficients are all interpretable in the absence of constant effect of the moderator variable, and the residuals sum of squares of nested models (with and without a product term) can be compared without any problem (as well as other fir statistics based on the residuals sum of squares). So it is not so much in these terms that one finds justification for model specification but more appropriately in the misspecification-bias argument. If there is a constant effect for x_2 and no such variable is specified in the estimated model, this will invariably introduce a misspecification bias, even in the case where x_1 and x_2 are uncorrelated. This is because the missing variable will be correlated with the product term of the moderator effect. It is, therefore, critical to evaluate the impact of specifying a model with x_2 on the ability to detect significant effects (constant and moderator effects).

In Figure 2, the two groups of bars on the right side display significance levels for estimated regressions on data generated without a constant effect of x_2. The middle group of bars displays significance levels of an estimated regression without a constant effect of x_2, the rightmost group of bars displays significance levels of an estimated regression with a constant effect of x_2.[9] Focusing first on the moderator effect parameter $ß_3$, there is no difference across the two estimations, with and without x_2. If x_2 is not specified as having a constant effect, the constant effect of x_1 (and the intercept as well) is highly likely to be significant (although less so as the correlation between x_1.and x_2 increases, indicative of the increased collinearity of the data). However, when x_2 is specified as having a constant effect as well as a moderator effect, the percentage level of significant constant effects of x_1 is reduced considerable from the 60%-90% range (depending on the correlation) to the 30% level found in prior cases where x_2 was a significant factor.

Therefore, introducing an irrelevant factor in the regression can create significant inference problems regarding the other factors that are relevant. As argued above, there is no reason due to interpretation or due to testing procedures to include an irrelevant factor. Because of the impact this irrelevant factor has on the estimated relevant parameters, model specification with or without x_2 as a constant effect can only be justified based on theoretical considerations. So, contrary to common recommendations (Irwin and McClelland 2001;

[9] We do not consider the case where x_2 is in the data generating function but is not in the estimated model, as this is the typical problem of misspecification bias that has been analyzed in the econometrics literature.

Irwin 2001), it is not wise to automatically specify constant effects in addition to moderator effects if theory only justifies a moderating effect. The absence of significance of a constant effect of x_2 cannot be used as a pre-test to decide whether or not to include x_2 because, as shown earlier, any lack of significance could be due to multicollinearity. Consequently, only theory (measurement and substantive) can drive the decision to include a variable as a constant effect or not. However, it should be recognized that this reasoning is not independent of the scale of the focal variable, as if it is interval scaled instead of ratio scaled, the product term implies that the effect of x_2 is only defined up to a constant; this requires that a constant effect for x_2 is introduced in the estimated equation.

Are the answers to questions 2 through 4 the same when there are moderating effects?

So far, we have considered a number of questions of practical relevance to the researcher with regard to data that did not contain moderating effects. We started with this analysis, which has not been investigated in the prior literature on the subject. We now consider the same questions when such moderator effects *do* exist in the data. While such Monte-Carlo simulation analysis has been performed on such data, we revisit the conclusions in light of our earlier demonstration that multicollinearity is a real issue that cannot be resolved by mean centering or any other transformation due to the nature of the product term.

Consider the question of the difficulty of finding moderating effects, which has motivated a large subset of the literature on the estimation of moderating effects. Figure 1a provides a graph of how the percentage of significant moderator effect parameters evolves as the true moderator effect increases. Here again, the absolute numbers are difficult to interpret because the significance depends on the size of the effect and on the noise in the regression. At minimum, the graph shows that the percentage increases as one would expect with the size of the true moderator effect.

Figure 5 graphs not only the percentage of significant moderator effects but also those of the other parameters. The graphs for the moderator effects are reproduced on this Figure (although they are the same as in Figure 1a) to allow comparison with the percentages obtained for the other parameters. While the size of the true moderator effect is smaller than those of the constant effects (except for the highest value of 1 where they are equal), the likelihood of finding a significant constant effect is much lower than the likelihood of concluding that there is a significant moderating effect. In fact, these likelihoods are extremely low, around the 10% level for any level of moderator effect when one exists (true moderator effect>0), which is not far from the 5% that would be obtained just by chance. Therefore, our earlier conclusion that multicollinearity has little impact on the moderator effect but has severe impact on constant effects can be generalized to cases where moderator effects exist: the impact of multicollinearity on the ability to detect a moderator effect is less severe than its impact on the ability to detect constant effects.

These graphs also confirm the extreme sensitivity of the estimated parameters to the inclusion of an irrelevant constant effect variable. When a moderator effect exists without a constant effect for that variable, the moderator effect is not difficult to detect. The percentages are close to 100%, as can be seen from the middle group of bars in Figure 5. However, the multicollinearity problems remain if the estimated equation contains such a constant effect (right-side group of graphs in Figure 5). The constant effect of x_1 is also better detected (60% level) if the irrelevant x_2 variable is omitted. This indicates that the model specification is critical but not in the traditional sense of an omitted variable that creates a bias. Instead, the issue in the context of estimating moderator effects is that including irrelevant variables for constant effects at the same time as moderator factors is an important source of multicollinearity problems.

It is interesting to contrast the graphs in Figure 6 with those of Figure 2. In both Figures, the graphs show the percentage of significant coefficients as a function of the correlation between x_1 and x_2. But in one case (Figure 2), there is no moderator effect in the data and in the other (Figure 6), such effects exist (the plots are aggregated across other conditions). The bars that correspond to the moderator effects are around the 5% level in Figure 2 (as discussed earlier) while somewhere around the 30% line in Figure 6, except when x_2 is not involved (the middle group of bars), where it reaches the 90% level. Even in the case of extremely high correlation between x_1 and x_2 (i.e., 0.8), the percentage of significant moderator effects is above 80%. The other effects would appear generally similar, leading to the same conclusions as presented earlier. Introducing a constant effect at the same time as a moderator effect reduces significantly the likelihood of finding any significant constant effects.

One aspect of the graphs in Figure 6 is unexpected. As noted by McClelland and Judd (1993), "increasing correlation between X and Z [here x_1 and x_2], with all else equal, improves the chances of detecting moderator effects," (p. 380). This is indeed reflected in the results graphed in Figure 6 for the model with a constant effect of x_2 being estimated, whether the data contain such a constant effect (group of bars on the left side) or not (group of bars on the right side). The effect, however, is not very large, as it requires high correlations for no more than an average 5% difference. Nevertheless, it does correspond to the statistical explanations provided by McClelland and Judd (1993). What is intriguing is that this effect is reversed if x_2 is not specified as a constant effect to be estimated (middle group of bars in Figure 6). Furthermore, this reversal applies to all the coefficients (consistent with what happens in Figure 2 where there is no moderating effect in the data). The pattern exhibited in the middle bars of Figure 6 is typical of the increased multicollinearity induced by the correlation between x_1 and x_2, which is reflected in a smaller percentage of significant coefficients as the correlation increases. The overall multicollinearity problems are even more dominant when x_2 appears both as a moderator effect and as a constant effect, as evidenced by a significantly lower percentage of significant coefficients. Therefore, it appears that the general multicollinearity problem interacts with the particular specific form of correlation implicit between the product term and its components. This does not affect the moderator (McClelland and Judd 1993) but does impact the constant effects.

In summary, the conclusions obtained from simpler data involving no moderator effect can be extended to data including moderating effects. It is the inherent correlation between the product term and its components that lowers the significance of constant effects, even though it does not affect the significance of moderator effects. The correlation between the focal variable and the moderator variable increases somewhat the significance of the moderator effect when this variable is also specified as having a constant effect (McClelland and Judd 1993). However, this result does not hold independently of the model specification, as it decreases slightly the likelihood of finding significant moderator effects when x_2 is only specified as a moderating factor.

What is the impact of recognizing the existence of a stochastic element in the moderating effect?

Given the major problems of multicollinearity demonstrated above (introduced by both the product term and by the specification of moderator variables as constant effects), it is critical to investigate solutions to this problem. Multicollinearity is a problem of statistical power in the sense that it does not introduce biases but makes it difficult to establish significant relationships. The mean centering approach proposed earlier was shown to be incapable of resolving the problems at hand. Therefore, it is necessary to investigate new avenues for solving this statistical efficiency problem. We do so below.

Additional information is the general solution to multicollinearity (Leamer 1978). We propose a method that takes into account the full information about the specification of a moderator effect, i.e., not only the deterministic relationship that explains the varying nature of a moderator effect but also its stochastic component as depicted in the model specified by equations (8) and (9). An asymptotically efficient estimation of such a model is the Generalized Least Squares estimator. The Appendix provides the details of such an estimation method. The improved efficiency is due to the information contained in the residuals and the particular form of heteroscedasticity implied by the moderator effect equation. However, in the case of no moderator effect, the variance of the error term should be independent of the focal variable and the only gain in efficiency to be expected would be due to the stochastic nature of the constant effect of x_1.

The datasets generated and discussed above were re-analyzed with the varying parameter model estimated with the Hildreth and Houck (1968) method (described in the Appendix). This method provides parameter estimates that are asymptotically efficient while OLS estimates are not. Our criterion for measuring efficiency is the Generalized Mean Squared Errors defined as $\mathrm{MSE}(\hat{b}) = (\hat{b} - b)(\hat{b} - b)' + \mathrm{V}[\hat{b}]$ (Judge et al. 1985). Even though the OLS parameter estimates are unbiased, they are unreliable due to multicollinearity and this is reflected by the first component of the MSE measure. We compare the gains in efficiency by taking the difference in the trace of the Generalized Mean Squared Errors obtained from OLS and those obtained from VPM estimations. This difference provides an overall measure of the improvements across all the model parameters. Figure 7a plots the gains in efficiency obtained from VPM estimation relative to OLS as a function of the size of the noise in the moderator equation. Indeed, the nature of the heteroscedasticity specified by Equation (12), $\mathrm{V}[u_i + x_{1i} e_i] = s_u^2 + x_{1i}^2 s_e^2$ explains why the benefits of the approach are not unconditional. Efficiency gains depend on the ability to estimate the components of the variance, i.e., s^2_u and s^2_e. This ability is a function of the relative size of the three components, s^2_u , $s^2_{e\cdot}$, and the variance exhibited by x_{1i}. If the first component s^2_u is large compared to the second part of the sum there is little gain in efficiency left since the variance is basically homoscedastic. Also, if x_{1i} exhibits only a small variance relative to the noise of the response equation or of the moderator equation, the auxiliary regression corresponding to the estimation of Equation (33) in the Appendix cannot provide sufficient information. Because the variance of x_{1i} was fixed to one in our Monte-Carlo simulation, we only need to consider how the efficiency varies as a function of the noise in the moderator equation, i.e., Figure 7a. As expected, the gain in efficiency increases with the noise in the moderator equation. The extent of the gains in efficiency can also be observed as the sample size decreases (Figure 7b). Therefore, varying parameter estimation appears particularly useful when the sample size is small and the moderating effect is noisy. These are rather typical contexts in which empirical tests are performed. Two additional observations must be made, however: (1) the gain in efficiency is not identical across the estimated parameters and (2) a gain in efficiency does not necessarily imply a greater proportion of statistically significant parameter estimates.

Considering separately the gains in efficiency of the intercept and the gains on the other parameters, the pattern remains the same as in Figures 9a and 9b; however, the magnitude is much larger for the intercept term (50 times larger for the intercept than for the other coefficients). Given that in most cases the intercept is not of strong theoretical interest, the benefits of using a more efficient estimation method seem to be limited. However, this conclusion is based on average effects, and in any particular case analysis may benefit from this more efficient estimation.

The second point concerns the statistical significance of the individual parameters. The impact of using a more efficient estimation method on the statistical significance does not appear very strong (Figures 10a and 10b show small differences with the results shown in Figures 7 and 8). This is due to the fact that, in spite of smaller variances of the parameter estimates, the varying parameter model estimates are closer to the true parameter which are typically smaller than the ones obtained from OLS.[10] Taking the ratio of the smaller parameter estimate with its smaller standard error does not necessarily lead to stronger t statistics. This is why the differences are small in Figure 1a between the results of the varying parameter model with those obtained from OLS estimation. However, our Monte-Carlo study shows that varying parameter model estimates tend to reduce the effect of multicollinearity in providing estimates that are more reliable (closer to their true value) and with a smaller standard deviation.

CONCLUSION

Theories in the social sciences including marketing are increasingly involving sophisticated explanatory mechanisms that generate contingent predictions. In many cases, data are available in the quantity and quality that is necessary to test such predictions. Yet, the sophistication of econometric methods commonly employed to treat these data has not always been well matched to the task. What has been missing is an appropriate means of estimating moderating effects. Specifically, particular attention must be given to the multicollinearity resulting from introducing a product term in a regression.

There has been a premature and unwarranted consensus that mean centering, followed by OLS estimation of models with product terms, is the solution. We have documented that this is a myth. Mean centering has no effect at all on multicollinearity. Furthermore, multicollinearity problems in moderated regressions arise not due to ill-conditioning of the data (especially the lack of independence between the focal and the moderator variables), but

[10] Based on the information detailed in Table 1, the median reliability of the coefficients is improved by 10.8%, 8.5%, 9.8%, and 7.7% for the intercept, the constant effects x_1 and x_2, and the moderator effect, respectively.

due to multicollinearity inherent to a model specification with a product term. Contrary to common persuasion, this model-inherent multicollinearity has very little effect on moderating or interaction estimates. However, it has very strong effects on the ability to detect constant effects. Therefore, no simple rule (i.e., always include constant *and* product terms for all variables) can be used for model specification. *Only theory can justify the specification of both moderator effects and constant effects*, especially when including a variable as a constant effect *and* as a moderator variable. Exploratory searches for moderator effects are especially dangerous, as they are likely to dampen the significance of *all* parameter estimates, due to the collinearity structure of data built from product terms.

We propose that Varying Parameter Models (VPM) can contribute to solve the model-inherent multicollinearity problem. In contrast to OLS, VPM makes use of the information about the error term structure, as implied by a conceptually appealing theory of moderating variables. Incorporating this information leads to more efficient estimation, which is badly needed for handling structurally introduced multicollinearity. Varying Parameter Models have the promise of detecting genuine moderating and constant effects, especially in data sets of modest size. As such, VPM is well matched to the increasingly sophisticated theorizing that underlies contemporary social science models.

Table 1: Generalized Mean Squared Errors of OLS and VPM Coefficients as a Function of the Noise in the Moderator Equation

Bias (median)

Noise in Moderator Equation		$ß_0$	$ß_1$	$ß_2$	$ß_3$
0.4	OLS	423.81	4.413	4.362	0.0437
	VPM	*405.98*	*4.285*	*4.153*	*0.0427*
0.8	OLS	1347.79	14.058	13.904	0.1388
	VPM	*1247.73*	*13.238*	*13.043*	*0.1308*
1.2	OLS	2904.03	31.268	29.830	0.3124
	VPM	*2704.78*	*28.710*	*27.331*	*0.2850*
1.6	OLS	5089.42	52.356	52.737	0.5327
	VPM	*4777.03*	*49.489*	*49.156*	*0.4973*
2	OLS	7989.31	83.938	82.159	0.8433
	VPM	*7333.18*	*78.473*	*74.628*	*0.7680*

Variance (median)

		$ß_0$	$ß_1$	$ß_2$	$ß_3$
0.4	OLS	848.20	8.571	8.551	0.0842
	VPM	*792.50*	*8.219*	*8.082*	*0.0817*
0.8	OLS	2627.22	26.390	26.464	0.2604
	VPM	*2435.47*	*25.197*	*24.805*	*0.2492*
1.2	OLS	5677.77	56.968	56.948	0.5617
	VPM	*5225.17*	*54.246*	*53.267*	*0.5398*
1.6	OLS	9867.41	99.257	99.316	0.9767
	VPM	*9066.89*	*94.245*	*92.476*	*0.9337*
2	OLS	15221.97	153.327	153.781	1.5051
	VPM	*13827.96*	*144.385*	*141.826*	*1.4344*

Generalized Mean Squared Error (median)

		$ß_0$	$ß_1$	$ß_2$	$ß_3$
0.4	OLS	1502.67	15.265	15.349	0.1518
	VPM	*1434.47*	*14.805*	*14.709*	*0.1477*
0.8	OLS	4739.46	48.312	48.590	0.4776
	VPM	*4393.04*	*45.784*	*45.092*	*0.4542*
1.2	OLS	10250.19	105.267	103.836	1.0362
	VPM	*9375.44*	*98.048*	*95.168*	*0.9705*
1.6	OLS	17454.35	179.321	176.909	1.7561
	VPM	*16187.78*	*169.078*	*165.470*	*1.6712*
2	OLS	27429.31	280.594	277.936	2.7651
	VPM	*25021.34*	*263.154*	*253.855*	*2.6036*

Figure 1a: **Percentage of Significant Estimated Moderator Coefficients as a Function of the True Moderator Effect**

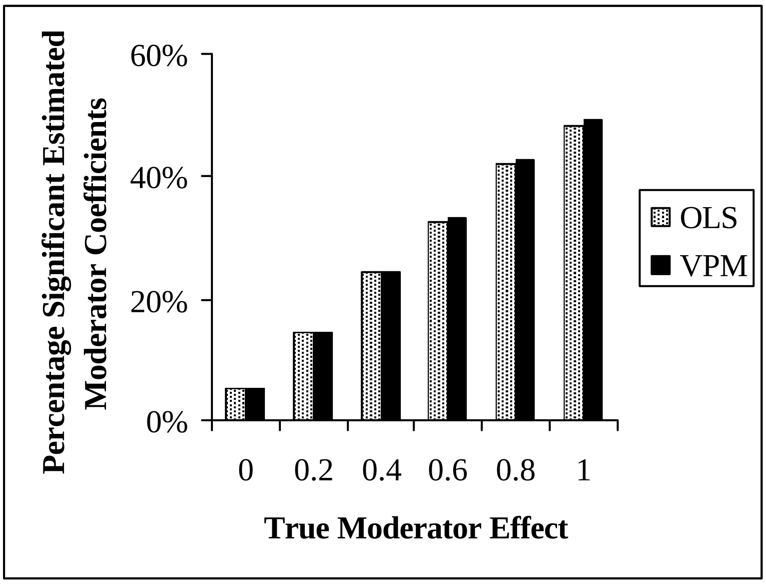
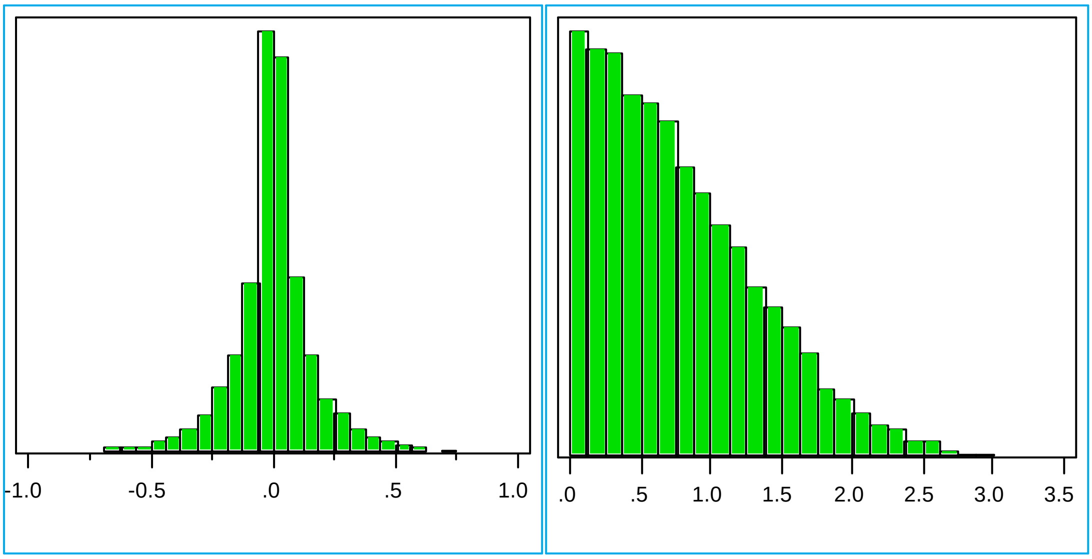


Figure 1b: **Distribution of Estimated Moderator Coefficients (OLS) and of t-Statistics when no Moderator Effect is Present in the Data Generating Function**



Estimate of moderator coefficient T-statistics of moderator coefficient

Figure 2: **Percentage of Significant Coefficients (OLS) as a Function of the Correlation between x_1 and x_2 when no Moderator Effect is Present**

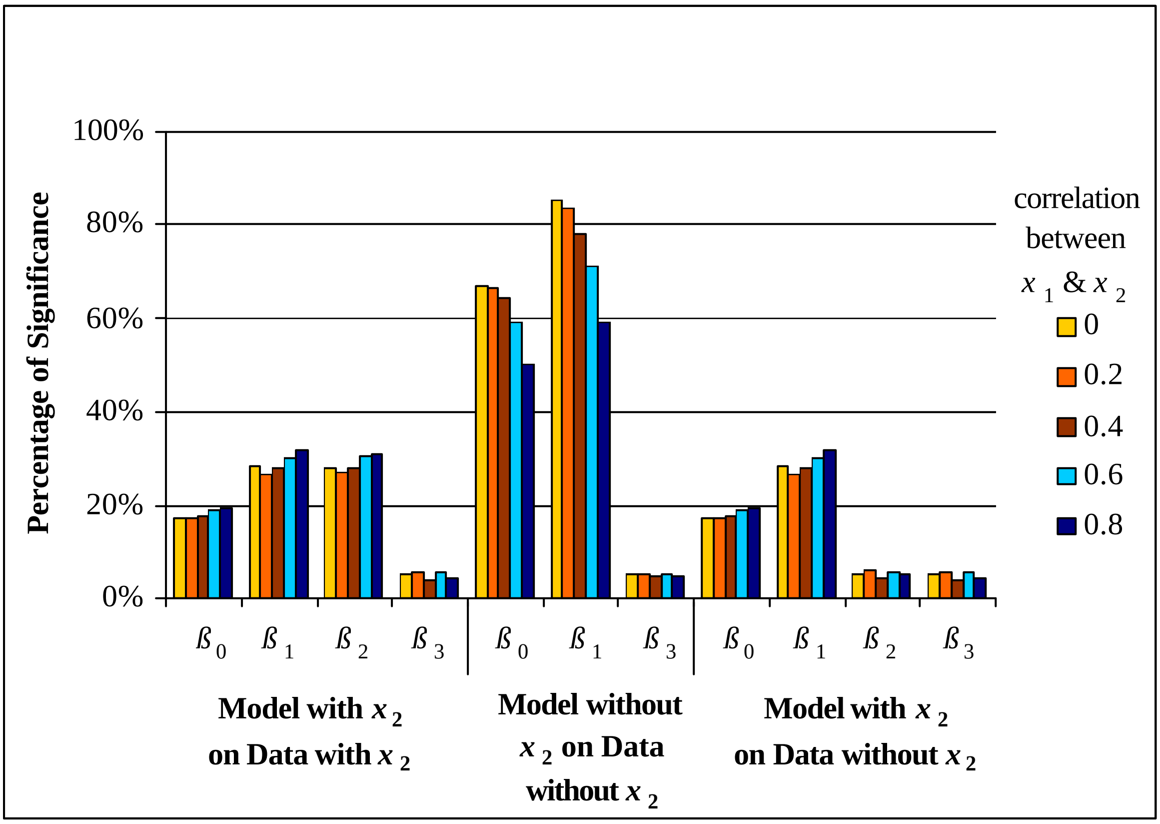
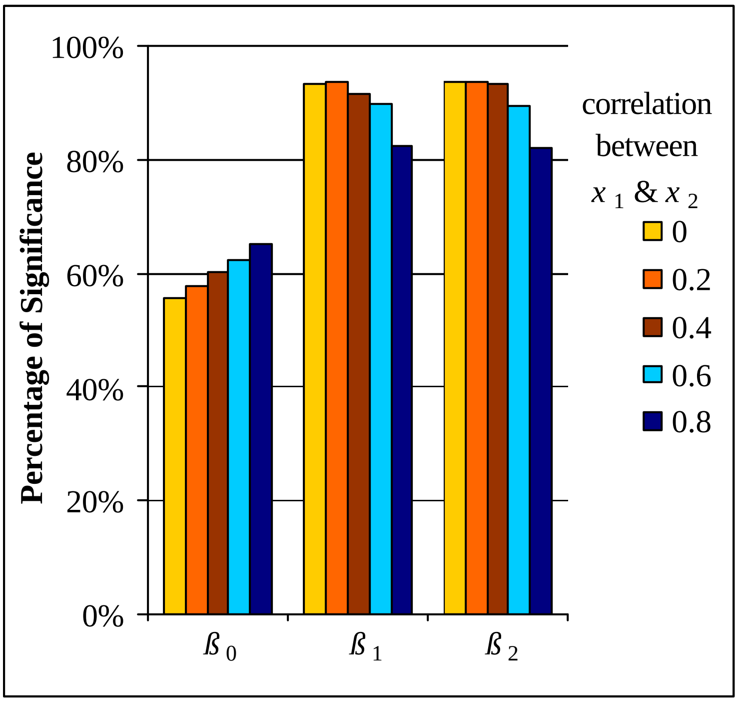


Figure 3: **Percentage of Significant Coefficients for Constant-Only-Effect Model as a Function of the Correlation between x_1 and x_2 when no Moderator Effect is Present**



<u>Figure 4:</u> Percentage of Significant Coefficients (OLS) as a Function of Regression Noise when no Moderator Effect is Present

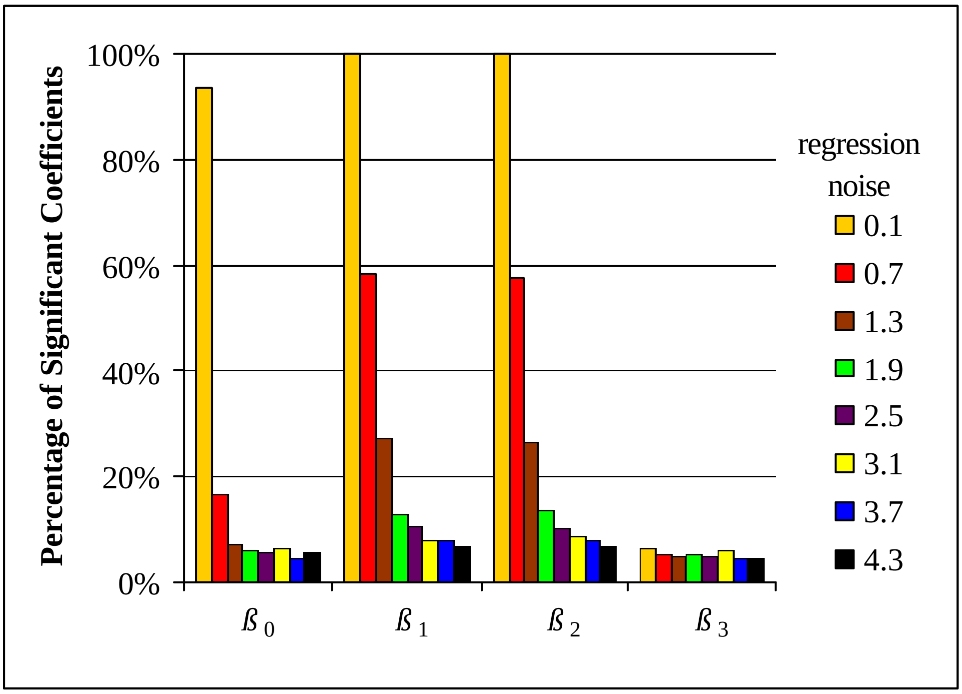
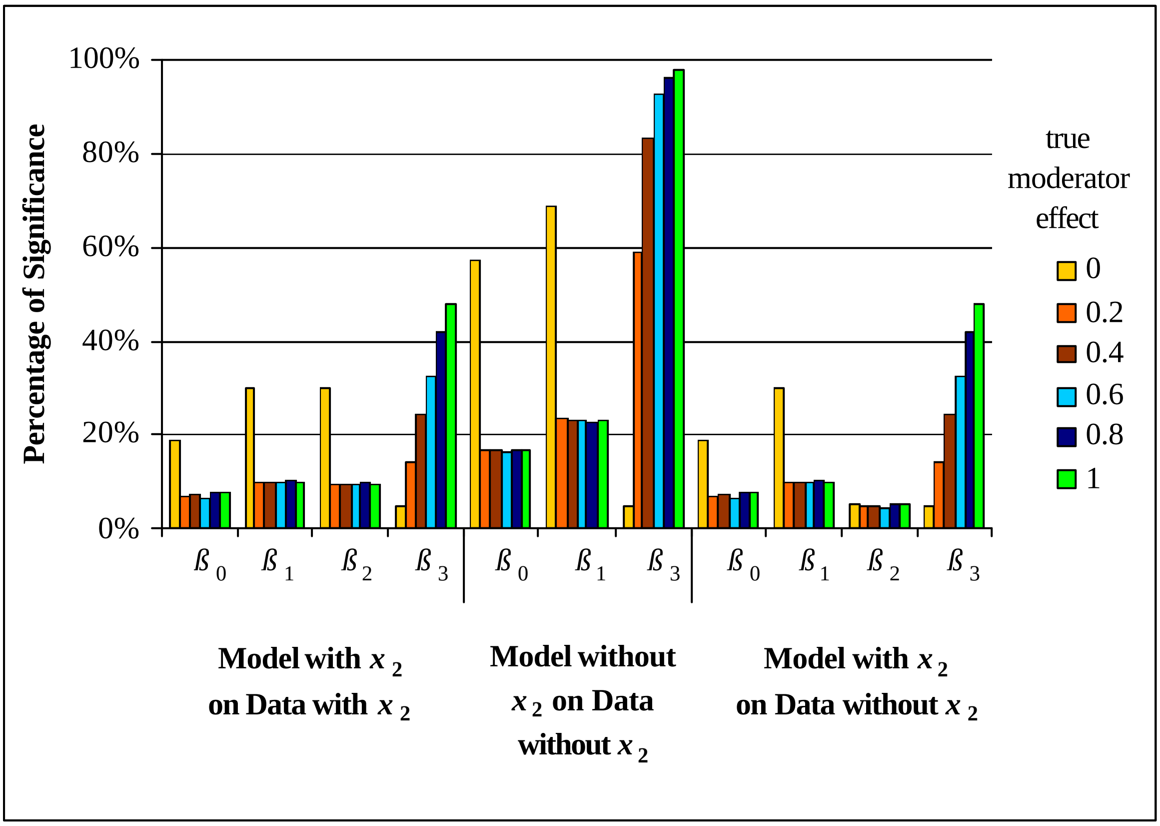
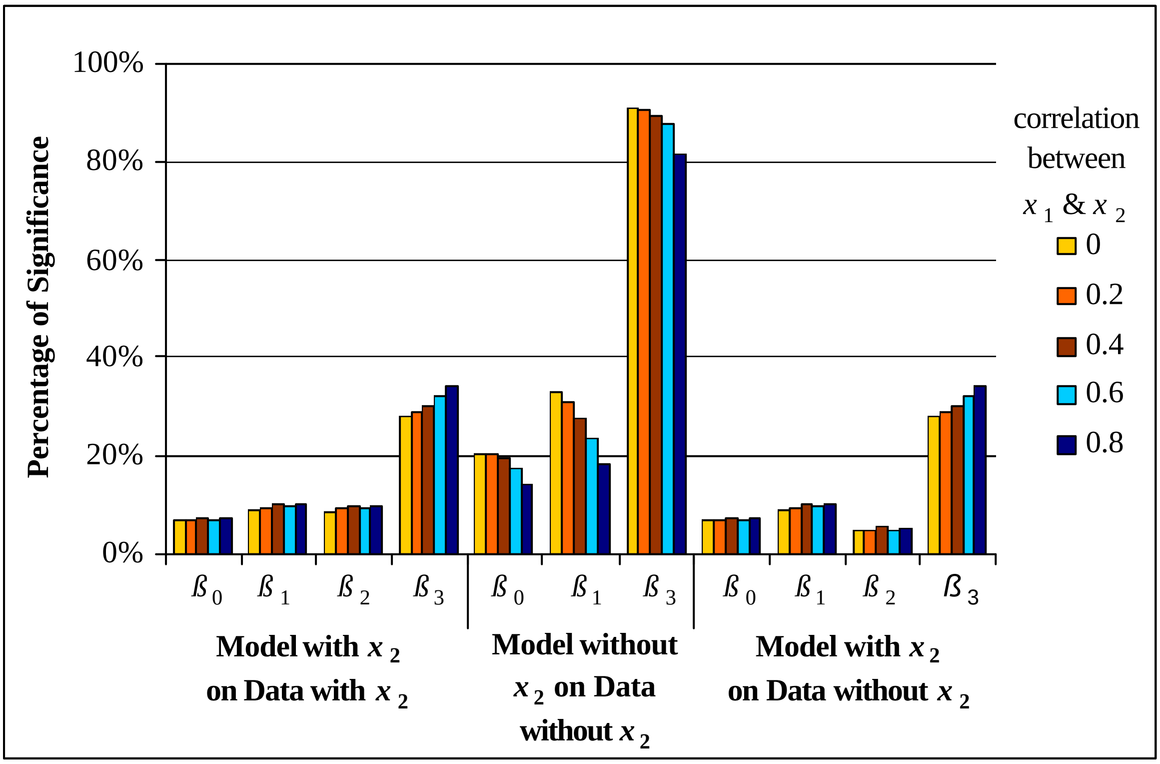


Figure 5: **Percentage of Significant Coefficients (OLS) as a Function of the True Moderator Effect**

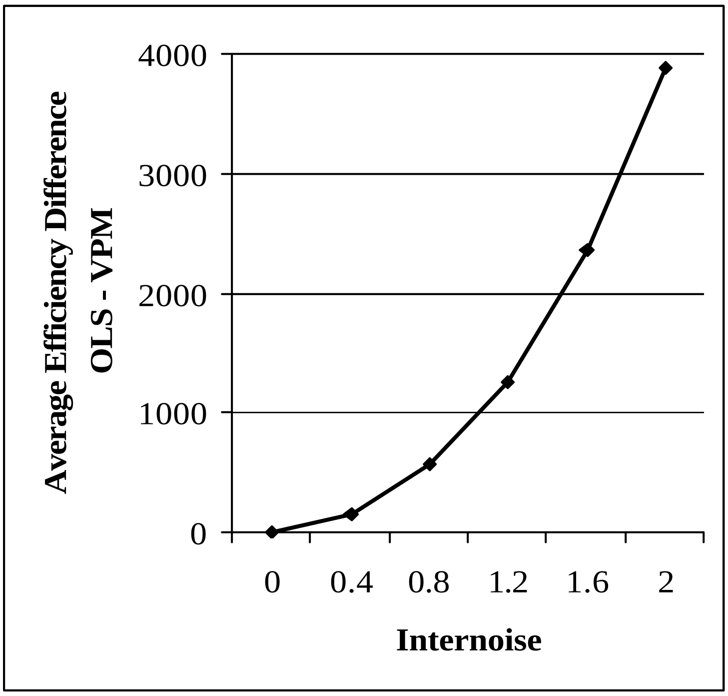


<u>Figure 6:</u> Percentage of Significant Coefficients (OLS) as a Function of the Correlation between x_1 and x_2 in the Presence of Moderator Effects



<u>Figure 7a:</u> **Efficiency Difference between OLS and VPM Coefficients as a Function of Noise in Moderator Equation**

(Positive values indicate the extent of increased efficiency of VPM vs. OLS)



<u>Figure 7b:</u> **Efficiency Difference between OLS and VPM Coefficients as a Function of Sample Size**

(Positive values indicate the extent of increased efficiency of VPM vs. OLS)

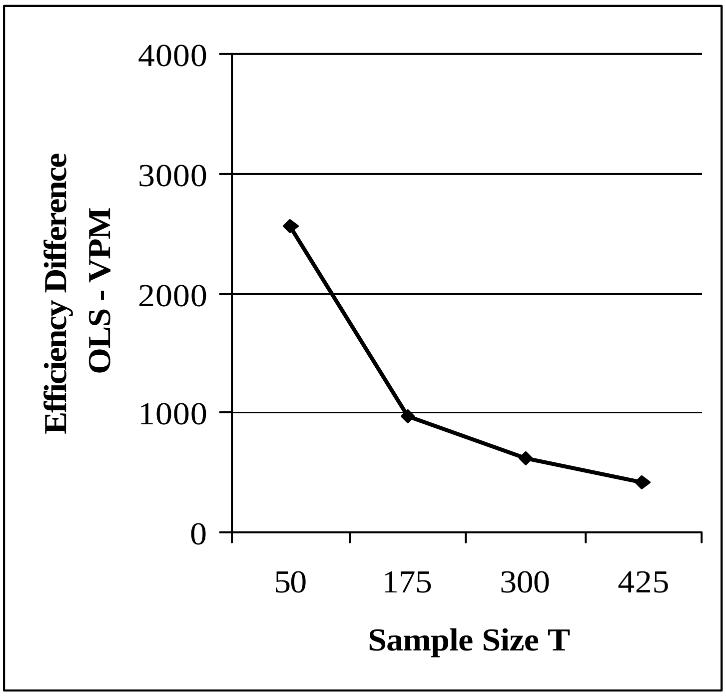


Figure 8a: **Percentage of Significant Coefficients (VPM) as a Function of the True Moderator Effect**

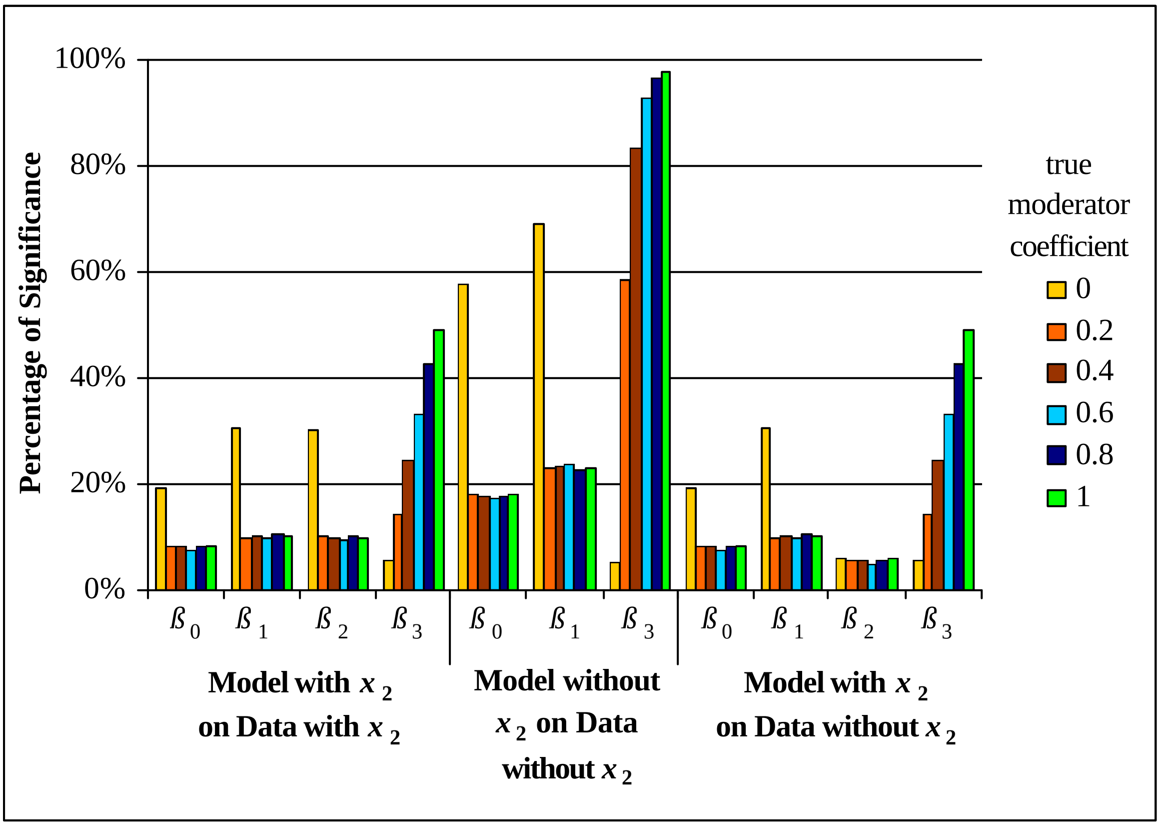
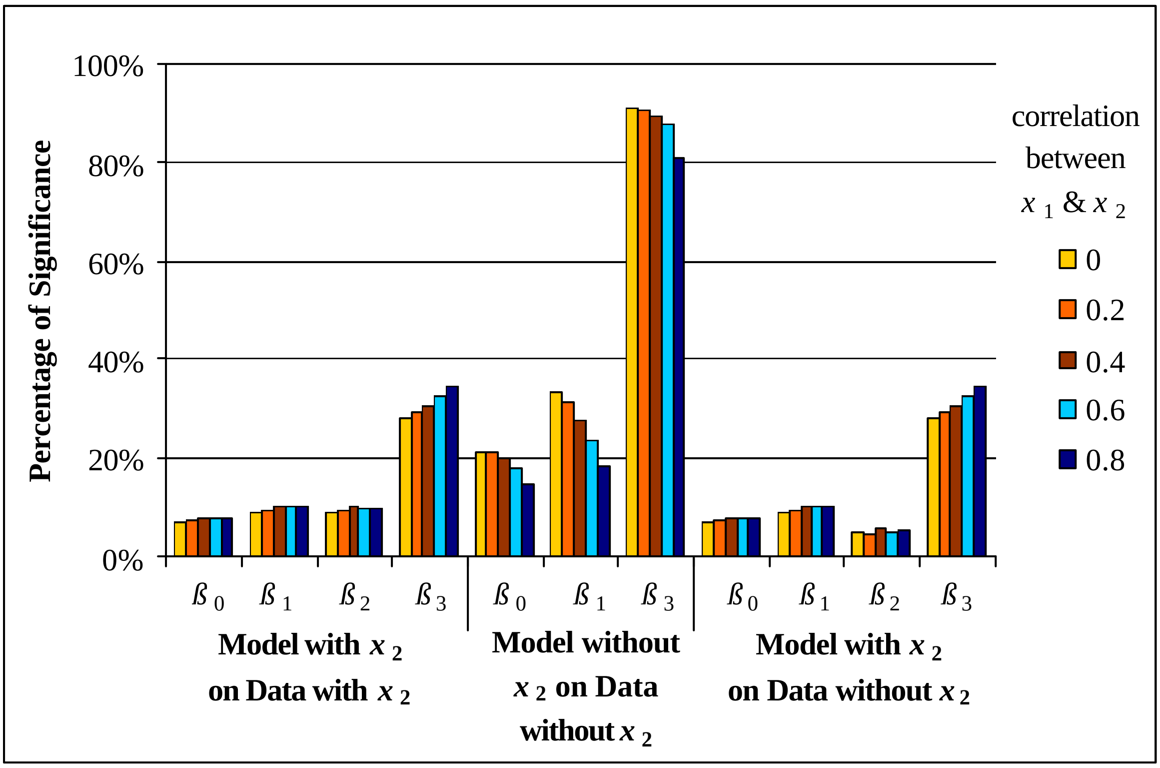


Figure 8b: **Percentage of Significant Coefficients (VPM) as a Function of the Correlation Between x_1 and x_2 in the Presence of Moderator Effects**



References

Aiken, Leona S., and Stephen G. West (1991), *Multiple Regression: Testing and Interpreting Interactions.* Thousand Oaks, CA, US: Sage Publications

Belsley, David A. (1984), "Demeaning Conditioning Diagnostics Through Centering," *The American Statistician*, 38, 2, 73-77.

Bowman, Douglas, and Das Narayandas (2004), "Linking Customer Management Effort to Customer Profitability in Business Markets," *Journal of Marketing Research*, 41 (4), 433-447.

Cohen, Jacob (1978), "Partialed Products are Interactions; Partialed Powers are Curve Components," *Psychological Bulletin*, 85, 4, 858-66

---- and Patricia Cohen (1983), *Applied Multiple Regression/Correlation Analysis for the Behavioral Sciences.* (2nd ed.), NJ, Hillsdale: Lawrence Earlbaum Associates

Cronbach, Lee J. (1987), "Statistical Tests for Moderator Variables: Flaws in Analyses Recently Proposed," *Psychological Bulletin*, 102 (3), 414-17.

Dunlap, William P. and Edward R. Kemery (1987), "Failure to Detect Moderating Effects: Is Multicollinearity the Problem?," *Psychological Bulletin*, 102 (3), 418-20.

Ganzach (1997), "Misleading Intercation and Curvilinear Terms." *Psychological Methods*, 2 (3), 235-47.

Gatignon, Hubert (1984), "Competition as a Moderator of the Effect of Advertising on Sales," *Journal of Marketing Research*, 21, 4 (November), 387-98.

---- (1993), "Marketing Mix Models," in Jehoshua Eliashberg and Gary L. Lilien, Eds., *Handbooks in Operations Research and Management Science*, Amsterdam, The Netherlands: Elsevier Science Publishers B.V.

---- and Dominique M. Hanssens (1987), "Modeling Marketing Interactions with Application to Salesforce Effectiveness," *Journal of Marketing Research*, 24 ,3 (August), 247-57.

Hildreth, Clifford and James P. Houck (1968), “Some Estimators for a Linear Model with Random Coefficients,” *Journal of the American Statistical Association*, 63, 584-95.

Irwin, Julie R. (2001), “Methodological and Statistical Concerns of the Experimental Behavioral Researcher,” *Journal of Consumer Psychology*, 10 (1&2), 97-98.

---- and Gary H. McClelland (2001), “Misleading Heuristics and Moderated Multiple Regression Models,” *Journal of Marketing Research*, 38 (1), 100.

Jaccard, James, Robert Turrisi, and Choi K. Wan (1990), *Interaction Effects in Multiple Regression*. Newbury Park, CA: SAGE Publications, Inc.

---- and Choi K. Wan (1996), *LISREL Approaches to Interaction Effects in Multiple Regression*. Thousand Oaks, CA: Sage Publications, Inc.

----, Choi K. Wan, and Robert Turrisi (1990), “The Detection and Interpretation of Interaction Effects Between Continuous Variables in Multiple Regression,” *Multivariate Behavioral Research*, 25, 4, 467-78.

Judge, George G., W. E. Griffiths, R. Carter Hill, Helmut Lutkepohl, and Tsoung-Chao Lee (1985), *The Theory and Practice of Econometrics*. New York, NY: John Wiley and Sons.

Kromrey, Jeffrey D. and Lynn Foster-Johnson (1998), “Mean Centering in Moderated Multiple Regression: Much Ado About Nothing,” *Educational and Psychological Measurement*, 58, 1, 42-67.

Leamer, Edward E. (1978), *Specification Searches*. New York: Wiley.

Lehmann, Donald R (2001), “Methodological and Statistical Concerns of the Experimental Behavior Researcher,” *Journal of Consumer Psychology*, 10 (1&2), 90-91.

McClelland, Gary H. and Charles M. Judd (1993), “Statistical Difficulties of Detecting Interactions and Moderator Effects,” *Psychological Bulletin*, 114 (2), 376-90.

Morris, James H., J. Daniel Sherman, and Edward R. Mansfield (1986), "Failures to Detect Moderating Effects with Ordinary Least Squares-Moderated Multiple Regression: Some Reasons and Remedy," *Psychological Bulletin*, 99 (2), 282-88.

Saunders, David R. (1956), "Moderator Variables in Prediction," *Educational and Psychological Measurement*, 16, 209-22.

Sharma, Subhash, Richard M. Durand, and Oded Gur-Arie (1981), "Identification and Analysis of Moderator Variables," *Journal of Marketing Research*, 18 (August), 291-300.

Smith, Kent W. and M. S. Sasaki (1979), "Decreasing Multicollinearity: a Method for Models with Multiplicative Functions," *Sociological Methods and Research*, 8 (1), 35-56

APPENDIX: ESTIMATION OF VARYING PARAMETER MODEL

Varying parameter models have been introduced in the econometrics literature with the random coefficient model of Hildreth and Houck (1968). Judge et al. (1985) show that varying parameter models belong to the class of heteroscedastic error models, where the variance of y_i is a linear function of a set of variables and compares several estimation methods. The case of moderator variables is identical to these models. Consequently, the following estimation method is proposed.

Generalizing the model expressed in Equations (8) and (9), the main equation is linear (possibly after transformation of the original variables) and for a single observation i is:

$$\underset{1x1}{y_i} = \underset{1xK}{\mathbf{x}_i'} \underset{Kx1}{b_i} + \underset{1x1}{u_i} \tag{13}$$

where, for each observation i:

$$\mathrm{E}\left[u_i\right] = 0$$

$$\mathrm{E}\left[u_i^2\right] = s_u^2$$

The coefficient for each variable is also indexed by the observation index i to represent that it varies over observations: b_{ki}. The coefficient of each explanatory variable x_k may now be expressed in terms of moderator variables $\mathbf{r}_i$. The vector $\mathbf{r}_i$ represents the values of G variables for observation i.

$$\underset{1x1}{b_{ki}} = \underset{1xG}{\mathbf{r}_i} \underset{Gx1}{a_k} + \underset{1x1}{e_{ki}} \tag{14}$$

The vector a_k contains the coefficients reflecting the extent to which the variables r_{gi} affect (i.e., moderate) the coefficient b_{ki}. A random term e also reflects the stochasticity of b. Equation (14) can be extended to include all the K coefficients composing the vector b_i (and a fortiori, only a subset of these parameters):

$$b_i = \begin{bmatrix} b_{1i} \\ b_{2i} \\ \vdots \\ b_{Ki} \end{bmatrix} = \begin{bmatrix} \mathbf{r}_i & 0 & \cdots & 0 \\ 0 & \mathbf{r}_i & & \vdots \\ \vdots & & \ddots & \\ 0 & \cdots & & \mathbf{r}_i \end{bmatrix} \begin{bmatrix} a_1 \\ a_2 \\ \vdots \\ a_K \end{bmatrix} + \begin{bmatrix} e_{1i} \\ e_{2i} \\ \vdots \\ e_{Ki} \end{bmatrix} \tag{15}$$

This can be written more compactly as:

$$\underset{Kx1}{b_i} = \underset{(KxKG)}{\mathbf{R}_i} \underset{(KGx1)}{a} + \underset{Kx1}{e_i} \tag{16}$$

This moderator equation is referred to more generally as the *process function* because it explains the way the variables in the original equation (13) predict the dependent variable. The error term is distributed with the following characteristics:

$$\mathrm{E}\left[e_i\right] = \mathbf{0}$$

$$\mathrm{E}\left[e_i e_i^{'}\right] = \underset{\mathrm{KxK}}{\Omega} = diag\left\{s_{w_k}^{2}\right\}$$

Therefore, the error terms in the process function are independent but the variances are different for different variable x_k it tries to explain.

By replacing Equation (16) into Equation (13), we obtain

$$y_i = \underset{1\mathrm{xK}}{\mathbf{x}'_i}\left(\underset{\mathrm{KxKG}}{\mathbf{R}_i}\ \underset{\mathrm{KGx1}}{a} + e_i\right) + u_i \tag{17}$$

$$= \mathbf{x}'_i\,\mathbf{R}_i a + \mathbf{x}'_i e_i + u_i \tag{18}$$

Bringing together the components of the error term, this can be written as:

$$y_i = \underset{1\mathrm{xS}}{\mathbf{p}'_i} a + e_i \tag{19}$$

where:

S= KG

$$\mathbf{p}'_i = \mathbf{x}'_i \mathbf{R}_i$$

$$e_i = \mathbf{x}'_i e_i + u_i$$

Consequently, the error term is normally distributed with the following mean and variance:

$$\mathrm{E}\left[e_i\right] = 0$$

$$\mathrm{E}\left[e_i^2\right] = s_i^2 \qquad = \mathrm{E}\left[\left(\mathbf{x}'_i e_i + u_i\right)\left(\mathbf{x}'_i e_i + u_i\right)'\right]$$

$$= \mathbf{x}'_i\,\mathrm{E}\left[e_i e'_i\right]\mathbf{x}_i + \mathrm{E}\left[u_i u'_i\right] = \mathbf{x}'_i\,\Omega\,\mathbf{x}_i + s_u^2 \tag{20}$$

$$= \left(\sum_k x_{ki}^2 s_{w_k}^2\right) + s_u^2 \tag{21}$$

Let $\mathbf{z}_i = \begin{bmatrix} 1 \\ x_{1i}^2 \\ x_{2i}^2 \\ \vdots \\ x_{\mathrm{K}i}^2 \end{bmatrix} \qquad g = \begin{bmatrix} s_u^2 \\ s_{w_1}^2 \\ s_{w_2}^2 \\ \vdots \\ s_{w_\mathrm{K}}^2 \end{bmatrix}$

Then:

$$\mathrm{E}\left[e_i^2\right] = s_i^2 = \underset{1\mathrm{xQ}}{\mathbf{z}_i'}\ \underset{\mathrm{Qx1}}{g} \tag{22}$$

where Q=K+1

$$\forall i \neq s: \quad \mathrm{E}\left[e_t e_s\right] = 0$$

Therefore the covariance matrix for the full set of observations is:

$$\mathrm{E}\left[ee'\right] = f = diag\left\{\mathbf{z}'_i g\right\} \tag{23}$$

The parameter vector a in Equation (19) can be estimated using the Generalized Least Square (GLS) estimator:

$$\underset{\mathrm{Sx1}}{\hat{a}} = \left[\underset{\mathrm{SxT}}{\mathbf{P}'}\ \underset{\mathrm{TxT}}{f^{-1}}\ \underset{\mathrm{TxS}}{\mathbf{P}}\right]^{-1} \underset{\mathrm{SxT}}{\mathbf{P}'}\ \underset{\mathrm{TxT}}{f^{-1}}\ \underset{\mathrm{Tx1}}{\mathbf{y}} \tag{24}$$

Because f is diagonal, the GLS estimator is therefore a weighted least squares estimator:

$$\hat{a} = \left[\sum_{i=1}^{\mathrm{N}} \left(\underset{1\mathrm{x}1}{\mathbf{z}'_i g}\right)^{-1} \underset{\mathrm{SxS}}{\mathbf{P}_i \mathbf{P}'_i}\right]^{-1} \sum_{i=1}^{\mathrm{N}} \left(\mathbf{z}'_i g\right)^{-1} \underset{\mathrm{Sx11x1}}{\mathbf{P}_i \mathbf{y}_i} \tag{25}$$

The parameter g is not known but if we replace it with a consistent estimator, we can obtain the estimated generalized least squares estimator:

$$\hat{a}_{EGLS} = \left[\sum_{i=1}^{\mathrm{N}} \left(\mathbf{z}'_i \hat{g}\right)^{-1} \mathbf{P}_i \mathbf{P}'_i\right]^{-1} \sum_{i=1}^{\mathrm{N}} \left(\mathbf{z}'_i \hat{g}\right)^{-1} \mathbf{P}_i \mathbf{y}_i \tag{26}$$

It remains to obtain a consistent estimator $\hat{g}$ for g:

$$\mathrm{E}\left[e_i^2\right] = \mathbf{z}'_i g \tag{27}$$

$$\hat{e}_i^2 = \mathbf{z}'_i g + v_i \Rightarrow v_i = \hat{e}_i^2 - \mathrm{E}\left[e_i^2\right] \tag{28}$$

$$\hat{e}_i = \mathbf{y}_i - \mathbf{P}'_i \hat{a}_{OLS}$$

$$\text{where } \hat{a}_{OLS} = \left(\mathbf{P}'\mathbf{P}\right)^{-1} \mathbf{P}'\mathbf{y} \tag{29}$$

However, $\mathrm{E}\left[v_i\right] \neq 0$ (because $\hat{e}_i^2 > 0$)

Let $\underset{\mathrm{Nx1}}{\dot{\hat{e}}} = \left\{\hat{e}_i^2\right\}$ and $\underset{\mathrm{NxQ}}{\mathbf{Z}} = \left\{\begin{matrix} \mathbf{z}_1 \\ \mathbf{z}_2 \\ \vdots \\ \mathbf{z}_{\mathrm{N}} \end{matrix}\right\}$

Then, it can be shown that

$$\mathbf{E}\left[\dot{\hat{e}}\right] = \dot{\mathbf{M}}\mathbf{Z}g \tag{30}$$

where $\mathbf{M} = \underset{\mathrm{N}x\mathrm{N}}{\mathbf{I}} - \mathbf{P}\left(\mathbf{P}'\mathbf{P}\right)^{-1}\mathbf{P}'$ (31)

Consequently:

$$\dot{\hat{e}} = \dot{\mathbf{M}}\mathbf{Z}g + \mathbf{W} \quad (32)$$

The ordinary least square for g can then be obtained:

$$\hat{g}_{OLS} = \left(\mathbf{Z}'\dot{\mathbf{M}}\dot{\mathbf{M}}\mathbf{Z}\right)^{-1}\mathbf{Z}'\dot{\mathbf{M}}\dot{\hat{e}} \quad (33)$$

When replaced into Equation (26), the EGLS estimator of a is:

$$\hat{a}_{EGLS} = \left[\sum_{i=1}^{\mathrm{N}}\left(\mathbf{z}'_i\hat{g}_{OLS}\right)^{-1}\mathbf{P}_i\mathbf{P}'_i\right]^{-1}\sum_{i=1}^{\mathrm{N}}\left(\mathbf{z}'_i\hat{g}_{OLS}\right)^{-1}\mathbf{P}_i\mathbf{y}_i \quad (34)$$