Central European Journal of Economic Modelling and Econometrics

Where do Moderation Terms Come from in Binary Choice Models?

Alfredo A. Romero*

Submitted: 10.01.2014, Accepted: 22.03.2014

Abstract

If the most parsimonious behavioral model between an observed behavior, Y, and some factors, X, can be defined as $f(Y|X_1, X_2)$, then f_{x_1} will measure the impact in behavior of a change in factor X_1. Additionally, if $f_{x_1x_2} \neq 0$, then the impact in behavior of a change in factor X_1 is qualified, or moderated by X_2. If this is the case, X_2 is said to be a moderating variable and $f_{x_1x_2}$ is said to be the moderating effect. When Y is modeled via a logistic regression, the moderation effect will exist regardless of whether the index function of the logit specification includes a moderation term or not. Thus, including a moderation terms in the index function will help the researcher more precisely qualify the moderation effect between X_1 and X_2. The question that naturally arises is whether the researcher must include the moderation term or not. In this document, we provide the conditions in which moderation terms will naturally arise in a logistic regression and introduce some modeling guidelines. We do so by introducing a general framework that nests models with no moderation terms in three scenarios for the independent variables, commonly found in applied research.

Keywords: moderation terms, moderation effects, logit models

JEL Classification: C15, C25, C52

*Department of Economics, North Carolina A&T State University, e-mail aaromero@ncat.edu

1 Introduction

In economics, as well as in other social and behavioral sciences, it is often necessary to model the behavior of economic agents reacting to different stimuli. The empirical applications of such studies have a wide range and cross many domains, including modeling labor force participation decisions (Faridi, Malik, and Basit, 2009; Theeuwes, 1981; Molina, Saei, Lombardia, 2007; Hyslop, 1999; Mroz, 1987), predicting opinions on legislation (Theilmann and Wilhite, 1987; Segal, Westerland, and Lindquist, 2010), understanding occupational choices of individuals (Boskin, 1974; Hansen, Wahlberg, and Faisal, 2010), qualifying the effectiveness of methods of teaching (Spector and Mazzeo, 1980), and explaining migration and relocation decisions (Nakosteen and Zimmer, 1980; Tunali, 1986). In the most general framework of predicting or modeling observed economic behavior, the objective is to account for the expected behavior of an observed response, Y, by relating the probability of observing that behavior to a set of explanatory variables, X, through some link function, usually the logit or the probit. It is then of interest to the researcher to understand the factors that determine the choice. In a two-variable case situation, where the most parsimonious behavioral model between an observed behavior, Y, and some sociodemographic characteristics or factors, X, can be defined as $f(Y|X_1, X_2)$, then f_{x_1} will measure the impact in behavior of a change in factor X_1, while f_{x_2} will do the same for the factor X_2. Additionally, if $f_{x_1x_2} \neq 0$, then the impact in behavior of a change in factor X_1 is qualified, or *moderated* by X_2. If this is the case, X_2 is said to be a moderating variable and $f_{x_1x_2}$ is said to be the moderating effect. It is this moderating factor that is of particular interest to researchers in order to qualify the marginal effect of a variable interest when controlling for additional information on the individual.

In logistic regression, the moderation effect will exist regardless of whether the index function of the logit specification includes a moderation term or not. The inclusion of the moderation term allows the researcher to more accurately and reliably assess the marginal effect of X_1 on Y as well as to more precisely qualify the moderation effect between X_1 and X_2. The question that naturally arises is whether the researcher must include the moderation term or not in the logit regression model. In this document, we provide the conditions in which moderation terms will naturally arise in a logistic regression and introduce some modeling guidelines. We do so by presenting a general framework that nests models with no moderation terms in three scenarios for independent variables, commonly found in applied research: continuous and continuous explanatory variables; continuous and dichotomous explanatory variables; and dichotomous and dichotomous explanatory variables variables. We show that in logistic regression, moderation terms appear to be the rule, rather than the exception, as long as the explanatory variables are non-independent. For this reason, we suggest researchers using logistic regression to follow a general-to-specific approach and exclude the moderation term only if warranted by likelihood ratio testing. The rest of this document is organized as follows: section 2 presents the modeling framework for logistic models, section 3 derives three models where the moderation terms naturally arise from the probabilistic structure of the data, section 4 provides the results of simulation analysis to test the existence of moderation terms. Section 5 summarizes this document and provides some modeling guidance.

2 Binary choice and the logit model

Consider a stochastic vector $\{Y_i, X_i,\ i = 1, \ldots, N\}$ whose joint density (probability) function takes the form $f(Y_1, \ldots, Y_N, X_1, \ldots, X_N; \varphi)$. For the moment, let $\{Y_i,\ i = 1, \ldots, N\}$ be a random variable distributed Bernoulli, with $E(Y_i) = p$ and $Var(Y_i) = p(1-p)$; and let $\{X_i = (X_{1,i}, \ldots, X_{K,i}),\ i = 1, \ldots, N\}$ be a vector of K random variables of unspecified distribution but with a proper joint density function $f(X_i, \psi_2)$. By construction, (Y_i, X_i) has identical distribution. Imposing independence to this joint distribution allows us to represent the density of (Y_i, X_i) as

$$f\left(Y_i, X_i\right) = \prod_{i=1}^{N} f\left(Y_i, X_i; \varphi\right) \tag{1}$$

Applying Bayes' theorem to (1) yields $f(Y_i|X_i)f(X_i) = f(X_i|Y_i)f(Y_i)$. Using this result, and the fact that Y_i can only take two values, it is possible to establish the following ratio (see Arnold, Castillo, and Sarabia, 1999 and Bergtold et al., 2010):

$$\frac{f\left(X_i|Y_i = 1; \eta_1\right)}{f\left(X_i|Y_i = 0; \eta_1\right)} \frac{f(Y_i = 1; p)}{f\left(Y_i = 0; p\right)} = \frac{f\left(Y_i = 1|X_i; \psi_1\right)}{f\left(Y_i = 0|X_i; \psi_1\right)} \frac{f\left(X_i; \psi_2\right)}{f\left(X_i; \psi_2\right)} \tag{2}$$

Notice that the appropriateness of (2) is conditional upon the existence and compatibility of the conditional density functions $f(Y_i|X_i)$ and $f(X_i|Y_i)$ but not on the existence and compatibility of $f(X_i, \psi_2)$. Since $f(Y_i|X_i; \psi_1)$ is Bernoulli distributed (see Chen and Liu, 1997; Spanos, 1999) with density function $f(Y_i|X_i; \psi_1) = g(X_i; \psi_1)^{Y_i}[1 - g(X_i; \psi_1)]^{1-Y_i}$, substituting this into equation (2) yields,

$$\frac{f(X_i|Y_i = 1; \eta_1)}{f(X_i|Y_i = 0; \eta_1)} \frac{\pi_1}{\pi_0} = \frac{g(X_i; \psi_1)}{1 - g(X_i; \psi_1)}, \tag{3}$$

where $\pi_j = p^j(1-p)^{1-j}$ for $j = 0, 1$. Thus (see Kay and Little, 1987),

$$g(X_i; \psi_1) = \frac{\pi_1 f(X_i|Y_i = 1; \eta_1)}{\pi_0 f(X_i|Y_i = 0; \eta_1) + \pi_1 f(X_i|Y_i = 1; \eta_1)} = \frac{\exp\{h(X_i; \eta_1)\}}{1 + \exp\{h(X_i; \eta_1)\}}, \tag{4}$$

where $h(X_i; \eta_1) = \ln \frac{f(X_i|Y_i=1;\eta_1)}{f(X_i|Y_i=0;\eta_1)} + \kappa$, and $\kappa = \ln \pi_1 - \ln \pi_0$. Thus, a proper statistical model in which the dependent variable is binary and the conditional relationship is Bernoulli naturally establishes the logistic cumulative density function as the transformation function and requires the use of logit specifications to model the statistical dependency of Y_i on X_i. As mentioned above, this does not rule out the possibility of using other transformation functions and obtaining similar results. It is clear from (4) that the functional form of the index function $h(\cdot)$ depends entirely on the conditional distribution of X_i given the two outcomes of Y_i. This probabilistic dependence is the basis for the derivation of logistic regressions with moderation terms.

At this point, an illustration of the use of probabilistic dependencies and the derivation of logistic specifications is in order. Consider the following case, let a stochastic vector $\{Y_i, X_i, i = 1, \ldots, N\}$ be such that not only Y is a binary choice random variable but also X is limited to taking only two values, High and Low (This is the case of the Tetrachoric Correlation. For a more detailed illustration see Romero (2011)). In this case, X is also Bernoulli distributed with density function $f(X_i|Y_i = j; \eta_1) = \rho_j^{X_i}(1-\rho_j)^{1-X_i}$, where η_1 is an appropriate set of parameters. Substituting into (4) yields,

$$E(Y_i|X_i = x_i) = \frac{\exp \beta_0 + \beta_1 x_i}{1 + \exp \beta_0 + \beta_1 x_1}, \tag{5}$$

where $\beta_0 = \kappa + \ln \frac{1-\rho_1}{1-\rho_0}$, $\beta_1 = \ln \frac{\rho_1}{\rho_0} - \ln \frac{1-\rho_1}{1-\rho_0}$, and $\kappa = \ln \pi_1 - \ln \pi_0$. Notice then that the index function contains a linear combination of X that is completely derived from the statistical properties of $f(Y_i)$, $f(Y_i|X_i)$, and $f(X_i|Y_i)$.

3 Where do moderation terms come from?

In the least squares tradition, when Y is a continuous random variable that can be modeled using regression analysis, the existence of a moderating effect requires the introduction of moderation terms (also known as interaction terms to econometricians) in the regression equation (Hayes and Matthes, 2009). In this case, for $f_{x_1x_2}$ to be not zero, the regression equation must include the factor or *focal* variable (x_1) and the *moderating* variable (x_2) entering as linear terms into the equation, as well as the cross product of the variables as the moderation factor (x_1x_2). Thus, in least squares modeling, the moderation term and the moderation effect are the same. However, when Y is a dichotomous random variable that can be modeled via a logistic regression, there exists a separation between the moderation term and the moderation effect, since the latter will exist regardless of whether the index function of the logit specification includes an moderation term or not (Mood, 2010). That is, in logistic regression, if Y and X_2 are not independently distributed, $f_{x_1x_2} \neq 0$ even if X_1 and X_2 are uncorrelated. In other words, while the inclusion of moderation terms is not a necessary condition for the existence of moderation effects, the addition of the moderation terms in the index function allows the researcher to more precisely and reliably qualify the moderation effect between X_1 and X_2. Not only that, failure to do so will result in statistically misspecified models and produce biased, and perhaps, inconsistent estimators of the marginal effect of X_1 and X_2 on

Y and the moderation effect between X_1 and X_2 because of the problem of statistical omitted variable bias (Mood, 2010).

It is not surprising that moderation terms are used widely in applied econometrics for their importance in qualifying the marginal effect of variables of interest in the observed response. Unfortunately, most applied researchers misinterpret the coefficient of the moderation term in nonlinear models. Nevertheless, in this document we will not provide a framework for computing interaction effects nor will we illustrate their appropriate interpretation. Excellent research papers on those subjects have been put forth by Ai and Norton (2003); Norton, Wang, and Ai (2004); and Hayes and Matthes (2009). In the logistic regression literature, it is commonplace to establish and assess the existence and magnitude of moderation terms by simply adding a multiplicative factor of two explanatory variables and test its statistical significance via a likelihood ratio test or a similarly appropriate metric. Nevertheless, based on the discussion in section 2, whether a specification of the index function contains or not an moderation term will depend completely on the statistical properties of the joint distribution of $\{Y_i, X_i, i = 1, \ldots, N\}$ and the related marginal and conditional densities. In what follows, we present three general models of moderation terms that nest index functions without moderation terms. Table 1 presents the three cases studied.

Table 1: Modeling Framework

Model	X_1	X_1
(1) Continuous - Discrete	Bernoulli Distributed	Exponential Distributed
(2) Discrete - Discrete	Bernoulli Distributed	Bernoulli Distributed
(3) Continuous - Continuous	Normally Distributed	Normally Distributed

The first model considers non-independent explanatory variables where one variable is continuous, takes non-negative values, and is positively skewed (say education, age, household income, home prices, etc.); whereas the other variable is dichotomous (say gender, employment status, etc.). The second model considers both explanatory variables to be dichotomous. The third model considers a situation where one variable is continuous and normally distributed (say real prices, capital-labor ratios, well-behaved test scores, etc.).

3.1 Model 1: Exponential-Bernoulli

Model Let X_1 given $Y = j$ be Bernoulli distributed; X_2 given $Y = j$ be exponentially distributed, and assume X_1 and X_2 are not independent (this assumption will be relaxed below). Although there is no explicit functional form for the joint distribution of $f(X_{1,i}, X_{2,i}|Y_i = j; \eta_1)$, it is still possible to parameterize this distribution through sequential conditioning since:

$$f(X_{1,i}, X_{2,i}|Y_i = j; \eta_1) = f(X_{2,i}|X_{1,i} = l, Y_i = j; \eta_1')f(X_{1,i}|Y_i = j; \eta_1'')$$

; thus, assuming X_2 given $X_1 = l$ and $Y = j$ be exponentially distributed, the joint distribution of (X_1, X_2) given $Y = j$ is (see Bergtold et al., 2010),

$$f(X_{1,i}, X_{2,i}|Y_i = j; \eta_1) = \left\{ \frac{\rho_j}{\theta_{j1}} \exp\left[-\frac{x_{2,i}}{\theta_{j1}} \right] \right\}^{x_{1,i}} \left\{ \frac{(1-\rho_j)}{\theta_{j0}} \exp\left[-\frac{x_{2,i}}{\theta_j} \right] \right\}^{1-x_{1,i}} \quad (6)$$

Substituting this into (4) yields (the derivation of this and the subsequent specifications is available upon request),

$$E(Y_i|X_i = x_i) = \frac{\exp\{\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i} x_{2,i}\}}{1 + \exp \beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i} x_{2,i}}, \quad (7)$$

Clearly, the probabilistic structure of the data requires the addition of moderation terms to the logit-link function. In this situation, failure to add them to the specification will lead to omitted variable bias as long as the explanatory variables are not independently distributed. If they indeed are independently distributed, that is, if $\theta_{j1} = \theta_{j0}$, the resulting specification becomes logit$(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i})$, where the moderation term is no longer needed, and is logit$(\cdot)$ represents the logit function.

3.2 Model 2: Bernoulli-Bernoulli Model

Let X_1 and X_2 given $Y = j$ be bivariate Bernoulli distributed. Furthermore, assume that X_1 and X_2 are not independently distributed (we will relax this assumption below). Using the sequential conditioning procedure illustrated in Model 1, we have

$$f(X_{1,i}, X_{2,i}|Y_i = j; \eta_1) = f(X_{2,i}|X_{1,i} = l, Y_i = j; \eta_1')f(X_{1,i}|Y_i = j; \eta_1''), \quad (8)$$

where X_1 given $Y = j$ is Bernoulli(ρ_j) and X_2 given $X_1 = l$ and $Y = j$ is Bernoulli(ρ_{j1}). In this case, the explicit functional form of this joint density becomes,

$$f(X_{1,i}, X_{2,i}|Y_i = j; \eta_1) = \left\{\rho_{j1}^{x_{2,i}}(1-\rho_{j1})^{1-x_{2,i}}\right\}^{x_{1,i}} \left\{\rho_{j0}^{x_{2,i}}(1-\rho_{j0})^{1-x_{2,i}}\right\}^{1-x_{1,i}} \cdot \cdot \rho_j^{x_{1,i}}(1-\rho_j)^{1-x_{1,i}}, \quad (9)$$

where $j = 0, 1$. Substituting (7) into (4) yields the following logit specification,

$$E(Y_i|X_i = x_i) = \frac{\exp\{\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i} x_{2,i}\}}{1 + \exp\{\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i} x_{2,i}\}}, \quad (10)$$

where, once again, the probabilistic structure of the data requires the addition of moderation terms to the logit-link function. Allowing X_1 and X_2 to be independently distributed implies $\rho_{jl} = \rho_{j0}$ for $j = 0, 1$. If this is the case, the above derivation yields logit$(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i})$, where once again, the moderation term is no longer needed.

3.3 Model 3: Normal-Normal model

Fortunately, explicit functional forms for the multivariate normal distribution exist. This allows us to model the joint conditional distribution of X_1 and X_2 given $Y = j$ directly. Let (X_1, X_2) be jointly normally distributed with means $\mu_{x_{1j}}$ and $\mu_{x_{2j}}$ dependent upon Y. Let the variance-covariance matrix also be heterogeneous with respect to Y. Assuming non-zero correlation between X_1 and X_2 (and assumption that will be relaxed below), and substituting the heterogeneous bivariate normal density into (4) yields the following conditional mean equation for the logit model:

$$E(Y_i|X_i = x_i) = \frac{\exp\left\{\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i}x_{2,i} + \beta_4 x_{1,i}^2 + \beta_5 x_{2,i}^2\right\}}{1+\exp\left\{\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i}x_{2,i} + \beta_4 x_{1,i}^2 + \beta_5 x_{2,i}^2\right\}}. \quad (11)$$

From this general specification, specific functional forms, commonly found in the literature, can be derived. For instance, a standard model with no squared terms in the X's but cross products can be obtained if $\rho_0 = \sqrt{\sigma_{x_{11}}^2(\rho_1^2 - 1) + \sigma_{x_{10}}^2}$ and $\sigma_{x_{21}} = \sigma_{x_{20}}^2$. In this case, the logit-link reduces to $\text{logit}(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i} + \beta_3 x_{1,i}x_{2,i})$, which is the standard moderation term specification.
Given the above three models, one might be inclined to think that moderation terms should always be added in logistic models since their existence is warranted by dependence between the explanatory variables, and non-independence between the regressors is likely; especially if the data is not generated through an experimental framework. This is, however, not necessarily so for it is possible to have a logit-link specification that is linear with no moderation terms and yet have non-independent regressors. To show this, and using again model 3, if the variance-covariance matrix is not heterogeneous with respect to Y, but X_1 and X_2 are correlated (i.e., $\sigma_{x_{11}}^2 = \sigma_{x_{10}}^2$, $\sigma_{x_{21}}^2 = \sigma_{x_{20}}^2$, and $\rho_1 = \rho_0 \neq 0$), then the logit-link function reduces to

$\text{logit}(\beta_0 + \beta_1 x_{1,i} + \beta_2 x_{2,i})$, with neither squared terms nor moderation terms.

4 Simulation analysis

We conducted Monte Carlo simulation analysis to illustrate how the specific probabilistic dependence between the variables in the model determines whether the logit-link function will include an moderation term or not. For Models 1 and 2, we conducted two simulation scenarios: one where the resulting model includes an moderation term and one where it does not. Alongside the usual descriptive statistics related to the simulations, we also provide the result of testing the existence of the moderation term by calculating the percentage of times we will reject a model without and moderation term, when indeed the model requires one, via a likelihood ratio test. It is our modeling recommendation that, since the existence of an moderation term seems to be the rule, rather than the exception, especially when the covariates are non-independent, that the researcher follows a general to specific approach when modeling them and then drop the aforementioned term if the likelihood ratio test warrants it. For Model 3, we exploit the flexibility of the normal specification and provide three different simulation scenarios: a model with moderation terms and squares of the explanatory variables; a model with moderation terms and linear terms of the explanatory variables; and a model where the explanatory variables enter linearly in the logit-link function and neither cross-products nor quadratic terms exist, although the covariates are not independently distributed. All the simulations where conducted using Matlab software with a sample size of $T = 200$ and $R = 10,000$ replicas. Additional sample sizes ($T = 100{,}500{,}1000$) were also estimated to verify convergence of the distribution of the estimated to normality, and are available upon request.

4.1 Model 1: Exponential-Bernoulli model

Moderation Terms: The stochastic process $\{(Y_i, X_i), i = 1, \ldots, N\}$ was generated letting $\pi_1 = .75$; $\rho_1 = .5$; $\rho_0 = .6$; $\rho_{01} = .6$; $\rho_{00} = .5$; $\rho_{11} = .4$; and $\rho_{10} = .75$. The simulation results are summarized in Table 2. The table includes the true parameters warranted by these probabilistic structure as well as the sample statistics of the estimated parameters and the percentage of times we will reject a model without and moderation term, when indeed the model requires one.
No Moderation Terms: The stochastic process $\{(Y_i, X_i), i = 1, \ldots, N\}$ was generated letting $\pi_1 = .7$; $\rho_1 = .5$; $\rho_0 = .6$; $\rho_{01} = .6$; $\rho_{00} = .6$; $\rho_{11} = .4$; and $\rho_{10} = .4$. The simulation results are also summarized in Table 2. Notice that this specification imposes independence between the regressors and thus requires no moderation term, as shown by the true parameters of this model. In this situation, we would expected the percentage of rejections to be higher than in a model that does require the moderation terms, as it is indeed the case.

Table 2: Summary Results of Monte Carlo Simulation for Model 1

		Sample Statistics				
Parameter	True Parameters	Mean	St. Dev.	Min.	Max.	Percentage of Rejections
		Model With Moderation Terms				
β_0	**0.9163**	0.9029	0.3774	-0.5301	2.6624	0.3151
β_1	**0.4055**	0.4517	0.5068	-1.8926	2.4806	
β_2	**0.6667**	0.7858	0.5806	-1.0176	5.9852	
β_3	**-1.5**	-1.6505	0.7719	-7.0428	1.4742	
		Model Without Moderation Terms				
β_0	**1.7272**	1.762	0.4006	0.3814	3.884	0.9413
β_1	**-0.4055**	-0.4094	0.5261	-2.6322	1.8707	
β_2	**-0.8333**	-0.8418	0.616	-3.8591	3.4295	
β_3	**0**	-0.025	0.8022	-4.0798	3.022	

4.2 Model 2: Bernoulli-Bernoulli Model

Moderation Terms: The stochastic process $\{(Y_i, X_i), i = 1, \ldots, N\}$ was generated letting $\pi_1 = .7$; $\rho_1 = .7$; $\rho_0 = .4$; $\rho_{01} = .6$; $\rho_{00} = .5$; $\rho_{11} = .4$; and $\rho_{10} = .75$. The simulation results are summarized in Table 3. The table includes the true parameters warranted by these probabilistic structure as well as the sample statistics of the estimated parameters and the percentage of times we will reject a model without and moderation term, when indeed the model requires one.

No Moderation Terms: The stochastic process $\{(Y_i, X_i), i = 1, \ldots, N\}$ was generated letting $\pi_1 = .7$; $\rho_1 = .7$; $\rho_0 = .4$; $\rho_{01} = .6$; $\rho_{00} = .6$; $\rho_{11} = .4$; and $\rho_{10} = .4$. The simulation results are also summarized in Table 3. Once again, notice that this second set of parameters imposes independence between the regressors and thus requires no moderation term. This can be seen on column 2 of the second panel of Table 3. Again, just as expected, the percentage of rejections for this reduced model is higher.

Table 3: Summary Results of Monte Carlo Simulation for Model 2

		Sample Statistics				
Parameter	True Parameters	Mean	St. Dev.	Min.	Max.	Percentage of Rejections
		Model With Moderation Terms				
β_0	**-0.539**	-0.5627	0.4159	-2.4849	1.0296	
β_1	**2.3514**	2.4207	0.5574	0.5013	5.5835	0.182
β_2	**1.0986**	1.1326	0.5155	-0.8023	3.541	
β_3	**-1.9095**	-1.9671	0.7118	-5.0251	0.9247	
		Model Without Moderation Terms				
β_0	**0.5596**	0.5827	0.3539	-0.9555	2.8622	
β_1	**1.2528**	1.2807	0.5201	-1.1024	4.0775	0.9486
β_2	**-0.8109**	-0.8473	0.4876	-2.9469	0.9555	
β_3	**0**	0.0117	0.6911	-3.2021	2.8556	

4.3 Model 3: Normal-Normal Model

Moderation and Squared Terms: The stochastic process $\{(Y_i, X_i), i = 1, \ldots, N\}$ was generated letting $\pi_1 = .75$. For $Y = 1$: $\mu_{x_1} = 4$; $\mu_{x_2} = 2$; $\sigma_{x_1} = 1$; $\sigma_{x_2} = 3$; and $\rho = -.75$. For $Y = 0$: $\mu_{x_1} = 3$; $\mu_{x_2} =$ ‘; $\sigma_{x_1} = 2$; $\sigma_{x_2} = 4$; and $\rho = -.75$. The simulation results are summarized in Table 4. The table includes the true parameters warranted by these probabilistic structure as well as the sample statistics of the estimated parameters and the percentage of times we will reject a model without cross-products and squared terms, when indeed the model requires them.

Moderation Terms: The stochastic process $\{(Y_i, X_i), i = 1, \ldots, N\}$ was generated letting $\pi_1 = .75$. For $Y = 1$: $\mu_{x_1} = 4$; $\mu_{x_2} = 2$; $\sigma_{x_1} = \sqrt{2}$; $\sigma_{x_2} = 2$; and $\rho = 0.8$. For $Y = 0$: $\mu_{x_1} = 3$; $\mu_{x_2} = 1$; $\sigma_{x_1} = 1$; $\sigma_{x_2} = \sqrt{2}$; and $\rho = 0.5292$. The simulation results are also summarized in the middle panel of Table 4. Notice that this model does not require the quadratic terms of the general specification. The reason is not related to any independence between the variables but rather to a judicious selection of the probabilistic structure of the data.

No Moderation Terms: The stochastic process $\{(Y_i, X_i), i = 1, \ldots, N\}$ was generated letting $\pi_1 = .75$. For $Y = 1$: $\mu_{x_1} = 4$; $\mu_{x_2} = 2$; $\sigma_{x_1} = 1$; $\sigma_{x_2} = 2$; and $\rho = 0.6$. For $Y = 0$: $\mu_{x_1} = 3$; $\mu_{x_2} = 1$; $\sigma_{x_1} = 1$; $\sigma_{x_2} = 2$; and $\rho = 0.6$. The simulation results are summarized in the lower panel of Table 4. Notice that homogeneity of the variance-covariance matrix is a necessary condition for the quadratic and moderation terms to disappear from the specification. Nevertheless, independence of the regressors is not a sufficient condition, as it was for the previous two models.

Table 4: Summary Results of Monte Carlo Simulation for Model 3

		Sample Statistics				
Parameter	True Parameters	Mean	St. Dev.	Min.	Max.	Percentage of Rejections
		Model With Moderation and Squared Terms				
β_0	**-19.2245**	-20.89	4.977	-46.09	-5.9805	0.0035
β_1	**9.2024**	9.955	2.294	2.5951	22.01	
β_2	**2.4603**	2.6509	0.7123	0.4885	6.5304	
β_3	**-0.4881**	-0.5259	0.1554	-1.3869	0.0055	
β_4	**-0.9762**	-1.056	0.2582	-2.4683	-0.1711	
β_5	**-0.0853**	-0.0901	0.0321	-0.2782	0.0486	
		Model With Moderation Terms and No Squared Terms				
β_0	**-2.6123**	-2.5905	2.3607	-13.14	13.4779	0.9999
β_1	**1.0938**	0.9678	1.5711	-10.67	7.4045	
β_2	**-0.0781**	-0.0193	0.6048	-2.219	4.8785	
β_3	**0**	-0.0292	0.2105	-1.6713	0.7824	
β_4	**0**	0.038	0.2619	-0.9626	2.0878	
β_5	**0**	0.0123	0.0624	-0.2302	0.3456	
		Model Without Moderation Terms				
β_0	**-2.6123**	-2.5905	2.3607	-13.14	13.4779	0.9999
β_1	**1.0938**	0.9678	1.5711	-10.67	7.4045	
β_2	**-0.0781**	-0.0193	0.6048	-2.219	4.8785	
β_3	**0**	-0.0292	0.2105	-1.6713	0.7824	
β_4	**0**	0.038	0.2619	-0.9626	2.0878	
β_5	**0**	0.0123	0.0624	-0.2302	0.3456	

5 Conclusion and guidelines

Logistic regression with moderation terms are of particular interest to social scientists modeling behavior of economic agents for they allow researchers to more precisely qualify the marginal response of the explanatory variables and the moderating relationships between them. Since the moderation effect exists whether the index function of the logit specification includes a moderation term or not, it is crucial that the researcher accounts for their existence to avoid statistical omitted variable bias of the estimates and unreliability of inferences. In this document we demonstrated that whether the logistic regression includes a moderation term or not depends entirely on the probabilistic structure of the data, especially when the covariates are not independently distributed. However, given that non-independence is not a sufficient condition of their existence, we recommend researchers follow a general to specific approach when using the logistic regression, only eliminating the moderation terms if warranted by likelihood-ratio testing.

References

[1] Ai, Chunrong and Edward C. Norton (2003). Interaction terms in logit and probit models, *Economics Letters*, 80, 123–129.

[2] Arnold, Barry C.; Enrique Castillo; Jose M. Sarabia (1999). *Conditional Specification of Statistical Models*, Lecture Notes in Statistics, Vol. 73. Springer Series in Statistics.

[3] Bergtold, Jason S.; Aris Spanos; Ebere Onukwugha (2010). Bernoulli Regression Models: Revisiting the Specification of Statistical Models with Binary Dependent Variables. *Journal of Choice Modeling*, 3(2), 1–28.

[4] Boskin, Michael J. (1974). A Conditional Logit Model of Occupational Choice, *The Journal of Political Economy*, 82(2-1), pp. 389–398.

[5] Chen, Sean X. and Jun S. Liu (1997). Statistical Applications of the Poisson-Binomial and Conditional Bernoulli Distributions, *Statistica Sinica*, 7, 875–892.

[6] Faridi, Muhammad Zahir; Shahnawaz Malik; A.B. Basit (2009). Impact of Education on Female Labour Force Participation in Pakistan: Empirical Evidence from Primary Data Analysis, *Pakistan Journal of Social Sciences*, 29(1), 127–140.

[7] Hansen, Jörgen; Roger Wahlberg; Sharif Faisal (2010). Wages and immigrant occupational composition in Sweden, *IZA Discussion Papers*, No. 4823.

[8] Hayes, Andrew F. and Jörg Matthes (2009). Computational procedures for probing interactions in OLS and logistic regression: SPSS and SAS implementations, *Behavior Research Methods*, 41, 924–936.

[9] Hyslop, Dan R. (1999). State Dependence, Serial Correlation and Heterogeneity in Intertemporal Labor Force Participation of Married Women, *Econometrica*, 67(6).

[10] Kay, R. and S. Little (1987). Transformations of the explanatory variables in the logistic regression model for binary data, *Biometrika*, 74(3), 495–501.

[11] Molina, Isabel; Ayoub Saei; M. Jose Lombardia (2007). Small area estimates of labour force participation under a multinomial logit mixed model, *Journal for the Royal Statistical Society*, 170(4), 975–1000.

[12] Mood, Carina (2010). Logistic Regression: Why We Cannot Do What We Think We Can Do, and What We Can Do About It, *European Sociological Review*, 26(1), 67–82.

[13] Mroz, Thomas A. (1987). The Sensitivity of an Empirical Model of Married Women's Hours of Work to Economic and Statistical Assumptions, *Econometrica*, 55(4), 765–99.

[14] Nakosteen, Robert and Michael Zimmer (1980). Migration and Income: The Question of Self-Selection, *Southern Economic Journal*, 46, 840–851.

[15] Norton Edward C.; Hua Wang; Chunrong Ai (2004). Computing interaction effects and standard errors in logit and Probit models, *The Stata Journal*, 4(2), 154–767.

[16] Romero, Alfredo A. (2011). Tetrachoric Correlation and Data Anomalies, *Working Paper, North Carolina A&T State University.*

[17] Segal, Jeffrey A.; Chad Westerland; Stefanie A. Lindquist (2010). Congress, the Supreme Court, and Judicial Review: Testing a Constitutional Separation of Powers Model, *American Journal of Political Science*, 55(1), 89–104.

[18] Spanos, Aris (1999). *Probability Theory and Statistical Inference: Econometric Modeling with Observational Data.* Cambridge, UK: Cambridge University Press.

[19] Spector, Lee C. and Michael Mazzeo (1980). Probit Analysis and Economic Education, *Journal of Economic Education*, 11(2), 37–44.

[20] Theeuwes, J. (1981). Family labour force participation: Multinomial logit estimates, *Applied Economics*, 13(4).

[21] Theilmann, John and Allen Wilhite (1987). A Southern Strategy for Labor, A Pac Connection? *Southeastern Political Review*, 15(1), 69–87.

[22] Tunali, Insan (1986). A General Structure for Models of Double-Selection and an Application to a Joint Migration/Earnings Process with Re-Migration, [in:] Ronald G. Ehrenberg (ed.), *Research in Labor Economics*, 8(B), JAI Press, 235–84.