Overstating and understating interaction results in international business research

Allison F. Kingsley[*,1], Thomas G. Noordewier, Richard G. Vanden Bergh

Grossman School of Business, University of Vermont, United States

ABSTRACT

Approximately one-third of international business (IB) articles include conditional hypotheses, yet the vast majority risk errors in testing or interpreting the results. Scholars typically restrict their empirical analysis to the coefficient of the interaction term in the regression, exposing themselves to the hazard of overstating or understating results. To mitigate the risk of misstating, we advocate that IB scholars also evaluate the statistical significance of the marginal effect of the primary independent variable over the range of values of the moderating variable. We demonstrate that overstating results can occur when the interaction term coefficient is statistically significant but the marginal effect is not significantly different from zero for some value(s) of the moderating variable. Understating can occur when the interaction term coefficient is not statistically significant, but the marginal effect is statistically different from zero for some value(s) of the moderating variable. In this article, we describe, using simulated data, these two possibilities associated with testing conditional hypotheses, and offer practical guidance for IB scholars.

© 2017 Elsevier Inc. All rights reserved.

1. Introduction

International business (IB) theory often includes conditional hypotheses. A conditional theory reflects scholars' recognition of the need to include a moderating variable in a proposed cause-and-effect relationship. The output of such theorizing takes the form of a hypothesis in which the relationship between a dependent variable and a primary explanatory variable of interest varies across the level or existence of some other moderating variable. To test a conditional hypothesis, researchers typically specify a regression model that includes a multiplicative interaction term. However, despite the growing number of articles containing such terms, IB researchers rarely distinguish – either conceptually or statistically – between two very different questions in their analysis of moderated relationships (Aiken & West, 1991).

Commonly, researchers ask only the following question: is the estimated coefficient on the interaction term in the regression statistically significant? If yes, then they generally conclude that support exists for the conditional hypothesis. However, IB researchers seldom explore a second, equally important question identified in the literature as contributing to a more complete test of the conditional hypothesis (e.g. Brambor, Clarke, & Golder, 2006; Berry, Golder, & Milton, 2012; Spiller, Fitzsimmons, Lynch, & McClelland, 2013). This question asks: is the effect of a change in the primary explanatory variable on the dependent variable (or, more simply, the "marginal effect" or "regression slope"), for any specific value of the moderating variable, statistically different from zero? The answer to the latter question provides vitally important additional information about the support for a conditional hypothesis. Whereas the first question asks whether marginal effects differ from *one another* for any two values of a moderating variable, the second question asks whether a marginal effect *differs from zero* for any specific value of a moderating variable (Aiken & West, 1991).

Differentiating between the two questions is critical. As we demonstrate in this paper, it is entirely possible to find, simultaneously, that the estimated coefficient on an interaction term is statistically insignificant *and* that the effect of a change in the primary explanatory variable (i.e., the marginal effect) is statistically different from zero over some portion of the range of the moderating variable. It is also possible for the researcher to find a statistically significant estimated interaction coefficient *but* that the effects of a change in the explanatory variable are significantly

different from zero for only some value(s) of the moderating variable.

A key implication – and the central point of this paper – is that failure to address both questions exposes IB scholars to the potential hazard of either *understating* or *overstating* empirical support for the conditional hypothesis. Understating may occur if researchers discard a conditional hypothesis based on obtaining a non-significant coefficient on the interaction term (Question 1), because they may miss seeing a statistically significant non-zero marginal effect of the primary explanatory variable for some value (s) of the moderating variable (Question 2). Overstating may occur if researchers go no further than to report a statistically significant interaction coefficient when, simultaneously, the marginal effect of the primary explanatory variable for one or more values of the moderating variable is not different from zero.

The approach that we advocate in this paper, by virtue of the different information content embedded in the different tests, is for the researcher to ask both questions. Correctly testing and interpreting interactions matters to the development and advancement of IB theory and practice. If scholars overstate, the field may be exposed to believing some causal relationships hold across more cases than is true, potentially causing managers to make misinformed decisions. Equally if not more damaging, if scholars understate, the field may be systematically losing important information about the state of the world, and distorting managerial practice. The questions international business scholars ask are important and nuanced, and doing justice to those questions requires not only building theories but employing appropriate statistical tests. Good theory yields better practice.

In the remaining pages, we discuss the importance of addressing these two questions to avoid overstating and understating the evidence of conditional effects. The following section details the statistical terminology and technicalities associated with empirically testing the two questions. We then review empirical research in IB to show that researchers typically do not explore interactions as advocated in this paper. In the next section, using simulated data on 200 firms considering how to best grow their direct investment in a foreign country, we walk step-by-step through several illustrative tests and interpretations of interaction models. We also show that the risk of overstating and understating is potentially quite large. The paper concludes by offering statistically sound and approachable recommendations to help IB scholars draw appropriate inferences in tests of interactions.

2. Defining interactions

In this section, we explore more deeply the technicalities of these two questions, and introduce terminology that we use throughout the remainder of the paper.

Recall the first question: is the estimated regression coefficient on the interaction term statistically significant? This question asks whether there is a statistically discernible difference between the marginal effect of a primary explanatory variable across different values of the moderating variable ("Question 1"). Scholars often refer to Question 1 as testing for an "interaction" effect. Consider the case of a researcher with a dichotomous (or binary) moderating variable. To empirically test whether two marginal effects are statistically distinguishable across "high" versus "low" values of the moderating variable, the researcher may specify an interaction model of the form:

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ \qquad (1)$$

where:

X is a (continuous) explanatory variable of interest,

Z is a dichotomous moderating variable (taking on the value 0 or 1),

Y is a continuous dependent variable.

The marginal effect of X on Y, sometimes referred to as the "simple slope," in Eq. (1) is given by:

$$\frac{\delta y}{\delta x} = \beta_1 + \beta_3 Z (\text{i.e., the derivative of Y with respect to X}) \qquad (2)$$

From Eq. (2), the marginal effect of X on Y is a function of a third variable, Z. Given an interaction model such as Eq. (1), it is not appropriate to speak of a single, unconditional marginal effect of X on Y (Brambor et al., 2006; Spiller et al., 2013; Aguinis, Edwards, & Bradley, 2016). When Z is dichotomous, there are two marginal effects to consider (or, as discussed below, when Z is continuous there is a range of marginal effects corresponding to different values of Z).

So, when $Z = 0 : \frac{\delta y}{\delta x} = \beta_1 + \beta_3(0) = \beta_1$

Similarly, when $Z = 1 : \frac{\delta y}{\delta x} = \beta_1 + \beta_3(1) = \beta_1 + \beta_3$

The difference between the two marginal effects is given by $(\beta_1 + \beta_3) - \beta_1 = \beta_3$. If the estimate of the interaction coefficient β_3 is statistically significant (determined by comparing it to its standard error), one can conclude that the two marginal effects (corresponding to $Z = 0$ and $Z = 1$) are discernibly (statistically) different from each other.

Researchers with conditional hypotheses typically stop their analysis with this test of the statistical significance of β_3. The precise reasons for this are speculative, but it is common practice across many disciplines (Aguinis et al., 2016; Ai & Norton, 2003; Brambor et al., 2006; Spiller et al., 2013). As noted, however, it is important that researchers go beyond simply testing Question 1; otherwise, they run the risk of understating or overstating support for the conditional hypothesis, which can serve to compromise the veracity and impact of the proposed theory.

Recall the second question: is the effect of a change in the primary explanatory variable on the dependent variable, for any specific value of the moderating variable, statistically different from zero? More precisely, the question asks, are the marginal effects of the primary explanatory variable statistically different from zero, for one, both, or neither level of the moderating variable? ("Question 2") To empirically test this (or any other null hypothesis-specified value), a researcher compares the marginal effect from Eq. (2) to its standard error. As shown above, the marginal effect when $Z = 0$ is simply represented by β_1, and the marginal effect when $Z = 1$ is represented by $\beta_1 + \beta_3$. If a marginal effect is not statistically different from zero, it means that there is no relationship between the explanatory variable of interest and the dependent variable at *that* specific value of the moderating variable.

It is entirely possible for the estimated $\hat{\beta}_3$ coefficient in an interaction model to be statistically significant, and hence for the marginal effects to be different from one another across the levels of the dichotomous moderating variable Z, and *yet* for the marginal effect of X on Y to be statistically distinguishable from zero for both, only one, or even neither of the two values of the moderator.

Or, again in the dichotomous case, it is possible for the $\hat{\beta}_3$ in the interaction model to be statistically insignificant, so that there is no discernible difference in marginal effect across high and low value of the moderator Z, but at the same time for the marginal effect of X on Y to be statistically different from zero for neither, only one, or even both of the two values of the moderator. In the case of a model

where Z is continuous, the estimated $\hat{\beta}_3$ may be statistically significant but with marginal effects different from zero over only some portion of the range of Z. Or, the estimated $\hat{\beta}_3$ coefficient in the model may be statistically insignificant, but reveal marginal effects that are statistically different from zero for at least some values of the moderating variable Z. In short, a statistically significant interaction coefficient $\hat{\beta}_3$ informs the researcher whether marginal effects differ from each other across levels of the moderating variable (Question 1), but not whether the marginal effects are statistically discernible from zero for all values of the moderating variable (Question 2). The two questions address different aspects of moderation.

In this paper, we draw heavily upon prior research that has advanced understanding of how to test and interpret conditional hypotheses, including seminal research across disciplines such as economics, political science, sociology, marketing and management (e.g., Ai & Norton, 2003; Aiken & West, 1991; Allison, 1977; Blalock, 1965; Brambor et al., 2006; Schoonhoven, 1981). As noted at the outset, Aiken and West (1991) identify two questions particularly relevant to conditional hypotheses. Brambor et al. (2006) highlight the importance of what we refer to as Question 2. Building on that insight, Berry et al. (2012) introduce briefly the terms "understating" and "overstating" in their treatment, but with a different objective − namely to show that scholars ignore evidence in symmetric conditional hypotheses.

Other scholars have also examined specific problems by discipline. For instance, in political science, Brambor et al. (2006) find that more than two-thirds of all articles in top tier political science journals misinterpret regression coefficients in interaction models or failed to properly report significance. Similarly, Ai and Norton (2003) find that nearly all economics articles made similar errors in interpreting the interaction coefficient in nonlinear models. In the field of management, Aguinis et al. (2016) report that only one out of 205 articles properly assessed conditional hypotheses, when analyzing research across multiple dimensions (i.e. measurement, range, sample size, statistical power, correlation, interpretation). Spiller et al. (2013) likewise conclude that moderated regression analyses in marketing journals were either sub-optimally performed (e.g. artificial split of continuous variable) or outright incorrect.

Our contribution to the literature on interactions is threefold. First, we connect the two questions into one coherent approach to testing and understanding conditional hypotheses. Second, we define clearly when scholars can experience the understating or overstating problem. Third, we are the first to examine the challenges of testing conditional hypotheses in the international business literature.

3. Interactions in IB

To explore the potential for understating or overstating in IB, we reviewed articles published in the two leading IB journals – *Journal of World Business* (JWB) and *Journal of International Business Studies* (JIBS) – for the years 2013, 2014, and 2015. We searched each journal volume for the following terms: interaction, interaction effect(s), interaction term(s), multiplicative interaction, product term(s), contingency effect(s), moderator variable(s), moderating variable(s), conditional effect(s), and polynomial term(s). JWB and JIBS published 343 articles during 2013, 2014, and 2015; we identified 116 total articles, or approximately one-third of all published articles, with interaction effects or conditional hypotheses. Of the 116 articles with interaction models, 68% involve linear interaction models (e.g. OLS) and 32% employ nonlinear interaction models (e.g. Tobit, probit, logit).

For ease of exposition, we focus our in-depth review on the 79 articles with linear interaction models. Each of those articles was individually and independently reviewed by two of the authors, with a third available to arbitrate any disagreements. We analyzed each article's empirical models and review the author's interpretation of the interaction term and marginal effects. The goal was to assess researchers' approaches with respect to model interpretation, specifically whether they answer the two questions we raise.

We first track if the interaction terms coefficient $\hat{\beta}_3$ is statistically significant in the main regression model. All but one of the 79 articles find at least one statistically significant interaction term. At face value, because scholars seem to be overwhelmingly publishing articles with statistically significant interaction terms, this is evidence that researchers may be at risk of overstating results. On the other hand, only one article does not find a statistically significant coefficient to any tested interaction term, which leads to the concern that researchers may not be submitting articles with a non-significant $\hat{\beta}_3$ or journals may be systematically rejecting those articles. If, in fact, findings of non-significance go unpublished, understating may also be a significant problem. This recognition that the field may be disproportionately publishing overstated results or not publishing other results and thus understating the support for proposed causal theory, suggests that theory development in the field may be distorted.

In terms of interpretation, the majority (58%) of the articles do not report more than the sign and level of statistical significance of the interaction coefficient $\hat{\beta}_3$. These papers, overwhelmingly, only address Question 1. Although roughly a third of articles go beyond simply reporting the sign and level of significance of the $\hat{\beta}_3$ coefficient to investigate the moderating relationship more fully, the majority do no more than plot the interaction. However, a plot of the marginal effects does not address Question 2 (i.e., whether the marginal effects individually are statistically different from zero). Our review finds that less than 10% of articles test and interpret interactions in such a way as to address this second question. The evidence, therefore, is that the majority of IB researchers are exposing themselves to the hazard of either overstating or understating the results. Better testing and interpretation of interaction models is merited.

Our review does nevertheless yield a few IB articles that attempt to avoid both the understating and overstating problem (e.g. Grinstein & Riefler, 2015; Jandhyala & Weiner, 2014; Jandhyala, 2015; Min & Smyth, 2014). For example, we find Jandhyala (2015) noteworthy in that the article illustrates the importance of answering both Question 1 and Question 2. In the analysis, Jandhyala finds that the coefficient on one interaction term is not statistically significant. However, the author avoids the understating problem by examining the marginal effect over the range of that moderating variable, and finds in more than a third of the sample that the marginal effect is, in fact, statistically significant, which provides partial support for the conditional hypothesis. At the same time, a coefficient on a different interaction term is statistically significant, yet Jandhyala avoids the overstating problem by noting that the marginal effect is statistically significant for only two-thirds of the values of the moderating variable. In this case, the conditional hypothesis is supported, but not fully. Had Jandhyala's analysis stopped at just interpreting the interaction coefficient, the author would have understated the importance of the first moderator and overstated the importance of the second. As analyzed, Jandhyala provides the reader with a much more nuanced and thorough understanding of the nature of the conditional hypothesis.

The process of theory development, whereby hypotheses are tested and then refined based on robust empirical testing, may best

be achieved by examining the marginal effect and not just the interaction effect. For example, Jandhyala finds that there is a threshold effect with one of the moderating variables, whereby it matters only above a certain level. A finding like this is important for theory and practice. The scholar may want to explore theoretically why the relationship between the explanatory variables and the dependent variable may not be relevant over all values of the moderating variable, or susceptible to a threshold effect. Further, the manager could err presuming the moderating relationship holds for cases below the threshold. Similarly, by analyzing the marginal effect on the second moderating variable, Jandhyala finds that the two ends of the sample behave quite differently. The marginal effect is positive but declining for the left end of the spectrum, yet it is negative and declining for the other end of the spectrum. Should scholars or managers not understand this nuance their theoretical or practical decisions could be in error.

To be clear, this review of IB research does not necessarily conclude that any one article or stream of IB scholarship is flawed. Indeed, it is very possible, if not probable, that many of the extant findings would hold up under greater statistical scrutiny. Our goal here is thus not to criticize prior research but to participate in a pragmatic conversation about best-practice methods for answering the big conditional questions in IB (Doh, Luthans, & Slocum, 2015).

4. Simulated data and analysis

To illustrate the distinction between ascertaining whether marginal effects differ from one another (Question 1) and whether any specific marginal effect (for some value of a moderating variable) is individually different from zero (Question 2), we present the hypothetical case of a researcher interested in modeling the growth of firm-level foreign direct investment (FDI) in a host country. We assume, to simplify discussion, that the researcher has correctly specified – both theoretically and functionally – the model being tested, and that interpretation of variable coefficients is therefore not hindered by biased parameter estimates.

4.1. Scenario 1: dichotomous moderating variable

4.1.1. Hypothesis 1

To begin, consider the case of a researcher with a conditional hypothesis in which the moderating variable is dichotomous (or binary). Suppose, drawing upon IB scholarship (e.g., Brouthers, 2002; Delios & Henisz, 2003; Johanson & Vahlne, 1977; Kogut & Zander, 1993; Andersson, Dasi, Mudambi, & Pedersen, 2016), that this researcher proposes the following conditional hypothesis, in which *FDI Growth* is the dependent variable, *Years Experience* is the independent variable, and *JV Partner* is the moderating variable:

Hypothesis 1. An increase in firm experience positively affects FDI growth, and this effect is greater when the firm invests with a locally based partner in the host country.

We illustrate how to conduct a complete test of this interaction proposition by means of a simulated data set of 200 foreign firms. The primary, or focal, independent variable in Hypothesis H1 is *Years Experience*, representing the number of years that the foreign firm has invested in operations located within the host country. The variable *Years Experience* was created by drawing 200 random observations from a truncated normal distribution: with a mean and standard deviation of 7 and 6, respectively, and with values restricted to the range [1,20].

The moderating variable in the hypothesis is a binary variable, *JV Partner*, representing whether the foreign firm is currently

investing in the host country with a locally based joint venture partner (*JV Partner* = 1) or not (*JV Partner* = 0). Two hundred observations of the variable *JV Partner* were drawn randomly from a uniform distribution, while restricting the draws to have integer values of either 0 or 1.

Table 1a provides the descriptive statistics for the two randomly generated explanatory variables.

The dependent variable in our hypothetical researcher's investigation is *FDI Growth*, measured as the percentage growth in foreign direct investment by a foreign firm in a host country. We simulate observations of this variable using a data generating process that relates *FDI Growth* to *Years Experience* and *JV Partner*, plus a random error term. In the data generating process, the error term is assumed to be normally distributed, with a mean of 0 and a standard deviation of 3 (i.e., $\varepsilon \sim N(0,3)$).

To test Hypothesis H1, we estimate the following interaction model:

$$FDI\ Growth = \beta_0 + \beta_1*Years\ Experience + \beta_2*JV\ Partner + \beta_3*Years\ Experience * JV\ Partner + \varepsilon \qquad (3)$$

To generate a simulated sample of *FDI Growth* we assume, arbitrarily, that the data generating process is as follows:

$$FDI\ Growth = 1.0 + 0.15*Years\ Experience + 0.02*JV\ Partner + 0.05*Years\ Experience * JV\ Partner + \varepsilon \qquad (3a)$$

The idea behind this data generating process is that there is a population relationship affecting a foreign firm's foreign direct investment such that: (a) the growth in foreign direct investment for the foreign firm in the host country increases by 0.15 percentage points for each year of experience the firm has operating in that country, and (b) the firm will grow foreign direct investment 0.02 percentage points if the firm invests with a local partner, and an additional 0.05 percentage points for every year of experience *if* it invests with a local partner. Since we only observe a sample of 200 firms from the total population, there is error in the data ($\varepsilon \sim N(0,3)$) and using ε, we draw 200 random observations.

4.1.2. Regression results

To highlight problems that often arise when interpreting interaction terms, we generate three random samples of *FDI Growth*. Table 1b presents the descriptive statistics for the dependent variable *FDI Growth* for all three samples.

The estimated coefficients obtained from ordinary least squares (OLS) regressions of *FDI Growth* on the explanatory variables are presented in Table 2. Columns (1a), (1b) and (1c) represent the regression results for the three samples of *FDI Growth*.

Including the interaction term in Eq. (3) allows for explicit testing of H1. As discussed previously, and addressing Question 1, a

Table 1a
Descriptive Statistics – Explanatory Variables.

Variable	mean	s.d.	min	max
Years Experience	8.93	4.67	1.00	19.97
JV Partner	0.56	0.50	0.00	1.00

n = 200.

Table 1b
Descriptive Statistics – Dependent Variable – Three random samples.

FDI Growth	mean	s.d.	min	max
Model (1a)	2.58	3.17	−8.53	12.62
Model (1b)	2.99	3.24	−4.40	11.58
Model (1c)	2.16	3.12	−6.87	9.85

n = 200.

Table 2
OLS Estimation Results.

Dependent Variable – Growth in Foreign Firm FDI in Host Country				
Variable	Coefficient	(1a)	(1b)	(1c)
Years Experience	β_1	0.143* (0.035)	0.115 (0.113)	−0.075 (0.293)
JV Partner	β_2	−1.736 (0.058)	−0.758 (0.439)	−1.298 (0.180)
Years Experience * JV Partner	β_3	0.192* (0.034)	0.103 (0.289)	0.194* (0.042)
Constant	β_0	1.332 (0.057)	1.882* (0.013)	2.612** (0.000)
Observations		200	200	200
R^2		0.157	0.068	0.028
$F(3,196)$		12.18**	4.76**	1.89

p-values in parentheses.
* $p < 0.05$.
** $p < 0.01$.

statistically significant $\hat{\beta}_3$ indicates that the marginal effects are different from each other. To address Question 2, we differentiate Eq. (3) and obtain the marginal effect of *Years Experience* on *FDI Growth*:

$$\delta FDI\ Growth/\delta Years\ Experience = \beta_1 + \beta_3 {}^* JV\ Partner \quad (4)$$

Eq. (4) explicitly depicts how the marginal effect of interest depends (i.e., is conditional) on the value of the moderating variable *JV Partner*. Since *JV Partner* is coded dichotomously, the marginal effect of *Years Experience* on *FDI Growth* in the absence of a *JV Partner* (i.e., when *JV Partner* = 0) is given by $\beta_1 + \beta_3 {}^* 0$ (i.e., β_1). Conversely, when a *JV Partner* is present (i.e., *JV Partner* = 1), the marginal effect is equal to $\beta_1 + \beta_3 {}^* 1$ (i.e., $\beta_1 + \beta_3$). Using the coefficient estimates in Table 2, the marginal effect values corresponding to the two levels of the moderating variable *JV Partner* (i.e., 0 and 1) are readily calculated (see Table 3) for all three samples (1a), (1b) and (1c).

To assess the statistical significance of each marginal effect in Eq. (4) requires estimation of its standard error. The estimated standard error of the marginal effect, depicted in Eq. (5) below, is calculated as:

$$\hat{\sigma}_{\frac{\delta FDIgrowth}{\delta Experience}} =$$
$$\sqrt{var(\hat{\beta}_1) + (JVPartner)^2 * var(\hat{\beta}_3) + 2 * JVPartner * cov(\hat{\beta}_1, \hat{\beta}_3)}$$
$$(5)$$

The important thing to note about this expression is that in addition to requiring estimates of the variance of $\hat{\beta}_1$, the variance of $\hat{\beta}_3$, and the covariance between the two estimated coefficients, the estimated standard error also depends upon the value of the moderating variable *JV Partner*.

Table 3
Marginal Effect of *Years Experience* on *FDI Growth*.

Moderating Variable	(1a)	(1b)	(1c)
JV Partner = 0	0.143* (0.035)	0.115 (0.113)	−0.075 (0.293)
JV Partner = 1	0.335** (0.000)	0.218** (0.001)	0.11 (0.060)

p-values in parentheses.
* $p < 0.05$.
** $p < 0.01$.

To highlight the challenges and potential pitfalls of testing conditional hypotheses, we briefly discuss each of the three samples of *FDI Growth*, considering in each case the answer to Question 1 of whether the interaction term $\hat{\beta}_3$ is statistically significant, and the answer to Question 2 of whether any marginal effect is statistically discernible from zero. Beginning with the results from sample (1a), it can be seen (Tables 2 and 3), that the interaction coefficient in the model (i.e., Eq. (3)) fitted to this sample of *FDI Growth* is statistically significant, and that the calculated marginal effect of *Years Experience* on *FDI Growth* is statistically significant (i.e., different from zero) for both levels of the contingent local partner variable (i.e., *JV Partner* = 0 and *JV Partner* = 1).[2] Additionally, the marginal effect is greater (amplified) when the firm invests with a locally based partner (i.e., *JV Partner* = 1). These findings lend support to H1, and are consistent with what researchers might expect to find when $\hat{\beta}_3$ is statistically significant. In this case, for sample (1a), the researcher finds support for H1 since both Question 1 and Question 2 are asked and answered in the affirmative.

It is quite possible, on the other hand, for the researcher to find that the interaction coefficient $\hat{\beta}_3$ is not statistically significant and yet, at the same time, for one or both of the marginal effects to be statistically discernible from zero. This scenario may be counter-intuitive to many IB scholars. For example, in sample (1b), despite the insignificant $\hat{\beta}_3$ coefficient, the results show that the marginal effect of *Years Experience* on *FDI Growth* is statistically significant in the case of *JV Partner* = 1, though not when *JV Partner* = 0. This result appears to provide some support for the conditional hypothesis.

The results from sample (1b) exemplify the need to empirically test and analyze both questions. Had the researcher considered only Question 1 by analyzing the statistical significance of the interaction coefficient $\hat{\beta}_3$ to find evidence for H1, s/he would have incorrectly drawn the conclusion of no support for the moderating hypothesis, understating the interaction results. This underscores the fact that the coefficient $\hat{\beta}_3$ (relative to its estimated standard error) in Eq. (3) does not, by itself, provide *all* of the information that may be useful to the researcher to fully assess the marginal effect of interest. It is worth re-iterating that the marginal effect in Eq. (4) involves not just the coefficient $\hat{\beta}_3$, but also the coefficient $\hat{\beta}_1$ and the value of the moderator *JV Partner*. Additionally, assessing the statistical significance of this effect requires more than knowledge only of the variance of $\hat{\beta}_3$, but also information on the variance of $\hat{\beta}_1$, the covariance between $\hat{\beta}_3$ and $\hat{\beta}_1$, and the value of the moderator *JV Partner*.

Sample (1c) illustrates yet another, and perhaps counterintui-

---

[2] A common mistake that IB scholars make when testing a conditional marginal effect using an interaction model (e.g., Eq. (3)) is to attempt to simultaneously interpret the estimated coefficient on the primary independent variable (i.e., β_1) as evidence of an "unconditional" marginal effect (also referred to as a "main effect") (Brambor et al., 2006). As we have noted, β_1 is the marginal effect of *Years Experience* on *FDI Growth* only when the value of *JV Partner* is zero. Accordingly, when a researcher proposes and finds (statistically significant) evidence of a conditional marginal effect, it does not make sense – logically or statistically – to simultaneously attempt an unconditional marginal effect explanation. Stated differently, the conditional marginal effect (i.e., $\beta_1 + \beta_3 {}^* JV\ Partner$) in Eq. (4) *subsumes* any unconditional or main effect interpretation of the relationship between *Years Experience* and *FDI Growth*. The attempt to interpret the coefficient β_1 as an unconditional marginal effect essentially ignores the more nuanced explanation embodied in the conditional marginal effect.

[3] The *F*-test in Sample (1c) indicates that the overall model is not statistically significant. However, we have found, for different underlying population models, examples where the estimated β_3 is significant and neither of the marginal effects are significantly different from zero and, importantly, the *F*-test shows the overall model is statistically significant.

tive, potential hazard in testing and interpreting marginal effects.[3] In this sample of *FDI Growth*, the interaction coefficient $\hat{\beta}_3$ is statistically significant (Table 2), yet the calculated marginal effects are not statistically significant (i.e., different from zero) for either value of the moderating variable, i.e., *JV Partner* = 0 or *JV Partner* = 1. The researcher who interprets a significant interaction coefficient β_3 by itself as evidence in support of H1 makes the potential mistake of overstating the information that this coefficient conveys about the hypothesized marginal effect. A significant $\hat{\beta}_3$, it turns out, does not imply that a statistically significant (i.e., different from zero) marginal effect exists for either value of the dichotomous moderating variable. Again, in light of the discussion above, this is understandable: the estimate and statistical significance of the marginal effect in Eq. (5) cannot be assessed based solely upon a reading of the interaction coefficient (and standard error of) $\hat{\beta}_3$ alone (Question 1), but rather requires knowledge of the information contained in Eq. (5) and the marginal effect's corresponding standard error (Question 2).

To summarize, if scholars studying Samples (1a), (1b), and (1c) stop with Question 1, they might conclude that the significant interaction coefficient $\hat{\beta}_3$ in Samples (1a) and (1c) provides full support for Hypothesis 1, while researchers using Sample (1b) might conclude that the hypothesis lacks empirical support. However, by addressing Question 2, scholars find that there are significant marginal effects in Samples (1a) and (1b) but not in Sample (1c). This suggests, within Sample (1a), that Hypothesis 1 has full support, but that concluding no support for Sample (1b) might have been premature, or, on the other hand, that there is only partial empirical support in the case of Sample (1c).

4.2. Scenario 2: continuous moderating variable

4.2.1. Hypothesis

Frequently, a conditional hypothesis involves a continuous moderating variable rather than a dichotomous one. Consider, for example, a new scenario in which a researcher investigates the relationship between firm experience and FDI growth, and expects that this association is moderated by the level of host country innovativeness. Hypothetical scenarios 1 and 2 are not related to each other. For ease of exposition we use the same sample of *Years Experience* and same dependent variable to motivate the hypotheses. We also use the same error structure for the data generating process. However, the data generating processes for scenarios 1 and 2 are independent of each other.

This researcher's conditional hypothesis might take the following form:

Hypothesis 2. An increase in firm experience positively affects FDI growth, and this effect is greater with higher levels of host country innovativeness.

To simplify exposition, we use the same definitions and data generating process for the primary independent variable (*Years Experience*) and dependent variable (*FDI Growth*) as employed in Scenario 1. However, in Scenario 2 the new moderating variable, *Innovation*, is continuous by construction. *Innovation* captures the general environment in the host country affecting innovation. By country, the measure calculates the number of patent applications filed under the United States Patent and Trademark Office (USPTO)

Table 4a
Descriptive Statistics – Explanatory Variables.

Variable	mean	s.d.	min	max
Years Experience	8.93	4.67	1.00	19.97
Innovation	38.24	10.76	15.60	65.36

n = 200.

system, per 10,000 of the country's population. Larger values represent a more favorable environment for innovation. The variable *Innovation* is generated by randomly drawing observations from a truncated normal distribution between 15 and 69 with a mean and standard deviation of 37 and 12, respectively. We arbitrarily create this variable for illustrative purposes only.

Table 4a below provides the descriptive statistics for the two randomly generated explanatory variables.

To test Hypothesis H2, we estimate the following interaction model:

$$FDI\ Growth = \beta_0 + \beta_1*Years\ Experience + \beta_2*Innovation + \beta_3*Years\ Experience * Innovation + \varepsilon \quad (6)$$

To generate a simulated *FDI Growth* we assume that the data generating process is as follows:

$$FDI\ Growth = 1.0 + 0.15*Years\ Experience + 0.005*Innovation + 0.0025*Years\ Experience * Innovation + \varepsilon \quad (6a)$$

The idea behind the data generating process for the model in Eq. (6a) is similar to that for the model in Eq. (3a). The population relationship affecting the growth in a foreign firm's foreign direct investment is assumed to increase 0.005 percentage points for each level of innovativeness for the host country, and an additional 0.0025 percentage points for the years of experience interacting with the level of host-country innovativeness. As in Scenario 1, we draw 200 random observations of *FDI Growth* and assume sample error of $\varepsilon \sim N(0,3)$.

4.3. Regression results

Following the strategy used in Scenario 1, we generate two random samples of *FDI Growth*. Table 4b presents the descriptive statistics for the dependent variable *FDI Growth* for both samples.

The estimated coefficients obtained from ordinary least squares (OLS) regressions of *FDI Growth* on the explanatory variables are presented in Table 5. Columns (2a) and (2b) represent the regression results for the two samples of *FDI Growth*.

As before, by differentiating Eq. (6) we obtain the marginal effect of *Years Experience* on *FDI Growth*:

$$\delta FDI\ Growth/\delta Years\ Experience = \beta_1 + \beta_3*Innovation \quad (7)$$

Eq. (7) is, of course, analogous to Eq. (4) except that any marginal effect of interest is conditional on the values of the continuous moderating variable *Innovation*. As in the dichotomous case, to answer Question 1, a statistically significant $\hat{\beta}_3$ indicates that any two marginal effects are discernibly different from one another (see Aiken & West, 1991 for a simple proof). To address Question 2, the researcher could, at this stage, use Eq. (7) to create a table of marginal effects and corresponding standard errors similar to Table 3, employing selected values of the moderating variable

Table 4b
Descriptive Statistics – Dependent Variable – two random samples.

FDI Growth	mean	s.d.	min	max
Model (2a)	3.52	3.10	−4.02	12.97
Model (2b)	3.35	3.09	−6.75	12.86

n = 200.

---

[3] The *F*-test in Sample (1c) indicates that the overall model is not statistically significant. However, we have found, for different underlying population models, examples where the estimated β_3 is significant and neither of the marginal effects are significantly different from zero and, importantly, the *F*-test shows the overall model is statistically significant.

Table 5
OLS Estimation Results.

Dependent Variable – Growth in Foreign Firm FDI in Host Country			
Variable	Coefficient	(2a)	(2b)
Years Experience	β_1	−0.000 (0.999)	−0.194 (0.245)
Innovation	β_2	−0.022 (0.591)	−0.077 (0.079)
Years Experience* Innovation	β_3	0.008 (0.056)	0.011* (0.010)
Constant	β_0	1.764 (0.283)	4.312* (0.013)
Observations		200	200
R^2		0.236	0.149
$F(3,196)$		20.23**	11.41**

p-values in parentheses.
* $p < 0.05$.
** $p < 0.01$.

Innovation. Alternatively, s/he could create a graph of the marginal effect line, with 95% confidence intervals around the line showing if a marginal effect is statistically significant at each value of the moderating variable.[4]

Models (2a) and (2b) in Table 5 provide examples of the problem in attempting to interpret the estimated coefficient for *Years Experience* (β_1) as an unconditional marginal effect. For both generated samples, the value of *Innovation* is between 15.6 and 65.4. The conditional marginal effect of *Years Experience* on *FDI Growth* (i.e., $\beta_1 + \beta_3$*Innovation*) can never equal β_1, because the value of *Innovation* is never equal to 0 in the sample.

Consider, as an example, the OLS results from model (2b) in Table 5. Fig. 1 below plots the marginal effect of *Years Experience* on *FDI Growth* along with the 95% confidence bands over the relevant values of the moderating variable *Innovation*. Fig. 1 also plots the frequency distribution of *Innovation*, where each bar of the histogram represents a count of the number of observations of *Innovation* in that range of values.

As is apparent from inspection of Fig. 1, *Years Experience* has a statistically significant positive effect on *FDI Growth* over most of the sample values of *Innovation* (from 29 to 65.4). For example, when *Innovation* is equal to 65, the marginal effect of *Years Experience* on *FDI Growth* is equal to approximately 0.52 percentage points: using Table 5 values, this is calculated from model (2b) as $0.52 = -0.194 + (0.011*65)$.

Critically, since the confidence interval bands do not cross 0 for values of *Innovation* greater than or equal to 29, we can conclude that the marginal effects are statistically different from zero (at the 95% level) over the range of *Innovation* from 29 to 65.4. A closer look at the histogram reveals that approximately 80% of the observations in the sample have values of *Innovation* greater than or equal to 29. Since the estimate to the β_3 coefficient is statistically significant, this example demonstrates the danger of overstating support for the conditional hypothesis. Researchers often imply that a statistically significant interaction coefficient alone (Question 1) shows the moderating variable exerting a conditional influence over the entire range of the contingent variable (Question 2). If, however, as in the case of sample (2b), the conditional marginal effect is statistically significant over only a portion of the range of the moderating variable, then support for the hypothesis (e.g., H2 in this case) must be regarded as only partial.

The case of sample (2a) highlights the potential hazard of understating the statistical evidence (i.e., incorrectly assuming that no conditional marginal effect exists when, in fact, there is such evidence of an effect). In our review of the IB literature, we observed that virtually every time the $\hat{\beta}_3$ coefficient was not statistically significant (as in our Model (2a)), the authors concluded that there was no support for the conditional hypothesis. In these cases, the authors may have missed nuanced moderation effects. Our conjecture is that most of the scholars publishing these articles do not recognize the potential importance of answering Question 2 by examining the standard errors of the marginal effect of the independent variable – appropriately calculated – over the entire range of substantively relevant values of the moderating variable. The importance of this is highlighted in Fig. 2 below, in which we show the graph of the marginal effect of *Years Experience* on *FDI Growth* over the range of the moderating variable *Innovation* for model (2a). Crucially, while the coefficient $\hat{\beta}_3$ in this model is not statistically significant, the plot of Fig. 2 shows that there is nevertheless a range of values for *Innovation*, from 21 to 65.4 (94% of the observations in the sample), over which the hypothesized marginal effect is statistically significant. A finding such as this shows support (at least partial) for the conditional hypothesis, and would likely be of great interest to most researchers.

To summarize, if scholars studying Samples (2a) and (2b) stop with Question 1, they might miss out on nuanced moderation related to conditional Hypothesis 2. In the case of Sample (2a), where the interaction coefficient $\hat{\beta}_3$ is statistically insignificant, there is in fact evidence of moderation. For instance, the marginal effect when *Innovation* is equal to 18 is not statistically different from zero, and yet the marginal effect when *Innovation* is 24 is statistically significant.[5] In the case of Sample (2b), where the interaction coefficient $\hat{\beta}_3$ is statistically significant, addressing Question 2 reveals marginal effects over some portion of the range of the moderating variable that are not statistically different from zero, for instance when *Innovation* is between the values of 15 and 29. This suggests that, over this range of values, there is no moderation.

4.4. Assessing likelihood of understating or overstating results

To calculate the likelihood that a researcher may understate or overstate support for the conditional hypothesis, we conducted a Monte Carlo simulation using the population model with *JV Partner* as the moderating variable. While we kept the error structure constant ($N \sim (0,3)$), we ran the Monte Carlo simulation for the following three population models:

FDI Growth $= 1.0 + 0.15$*Years Experience* $+ 0.02$*JV Partner* $+ 0.05$*Years Experience * JV Partner* $+ \varepsilon$ (8)

[4] The commands are extensions of earlier code made available by Matt Golder and straightforward to implement in Stata 13. All Stata code to replicate the data, tables, and graphs presented in this paper is available from the authors' website.
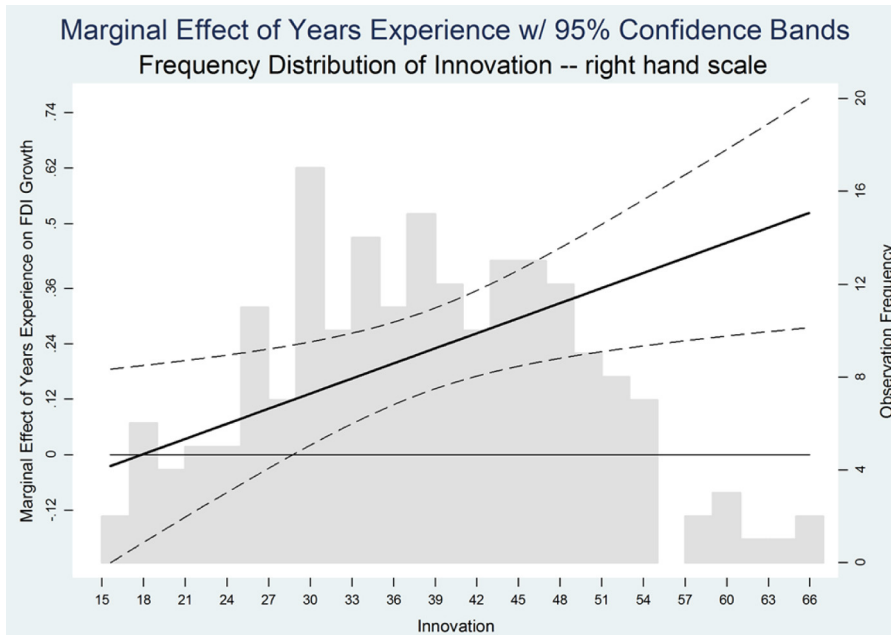
[5] An interesting case (though not part of our simulations) is that in which the coefficient $\hat{\beta}_3$ is statistically insignificant and, at the same time, the marginal effect of *Years Experience* on *FDI Growth* is statistically non-zero over the entire range of *Innovation*. Here, evidence of moderation is lacking because an insignificant $\hat{\beta}_3$ means, by definition, that the marginal effects for any two values of *Innovation* do not significantly differ from one another. Thus, while the marginal effect of *Years Experience* on *FDI Growth* differs from zero across the entire range of *Innovation*, different values of *Innovation* do not significantly (statistically) differentiate calculated marginal effects. The net effect of answering both Question 1 and Question 2, in this case, is that the scholar finds only evidence of a main effect of *Years Experience* on *FDI Growth*.

Fig. 1. Marginal Effect of Years Experience on FDI Growth based upon Model (2b) in Table 5.

$FDI\ Growth = 1.0 + 0.15 * Years\ Experience + 0.02 * JV\ Partner - 0.05 * Years\ Experience * JV\ Partner + \varepsilon$ \hfill (9)

$FDI\ Growth = 1.0 + 0.25 * Years\ Experience + 0.02 * JV\ Partner - 0.25 * Years\ Experience * JV\ Partner + \varepsilon$ \hfill (10)

The first population model (8) is Eq. (3a) above. For the second model (9) we only change the sign of the parameter on the interaction term. For the third model (10) we keep the signs the same as the second model but change the magnitude of the parameters on *Years Experience* and the interaction term *Years*

Experience * JV Partner. We emphasize that these three models are arbitrarily chosen from among the infinite set of population models. The results of this analysis are not meant to demonstrate a general likelihood of understating or overstating but rather to illustrate that the likelihood depends upon the unobservable population parameters.

For each of the three population models we draw 25,000 samples. For each draw we estimate regression model (3). We then analyze the results from each regression to determine if there is an understating or overstating problem. Consistent with Question 1 and Question 2, we define a potential understating problem as occurring when the coefficient on $\hat{\beta}_3$ is not statistically significant
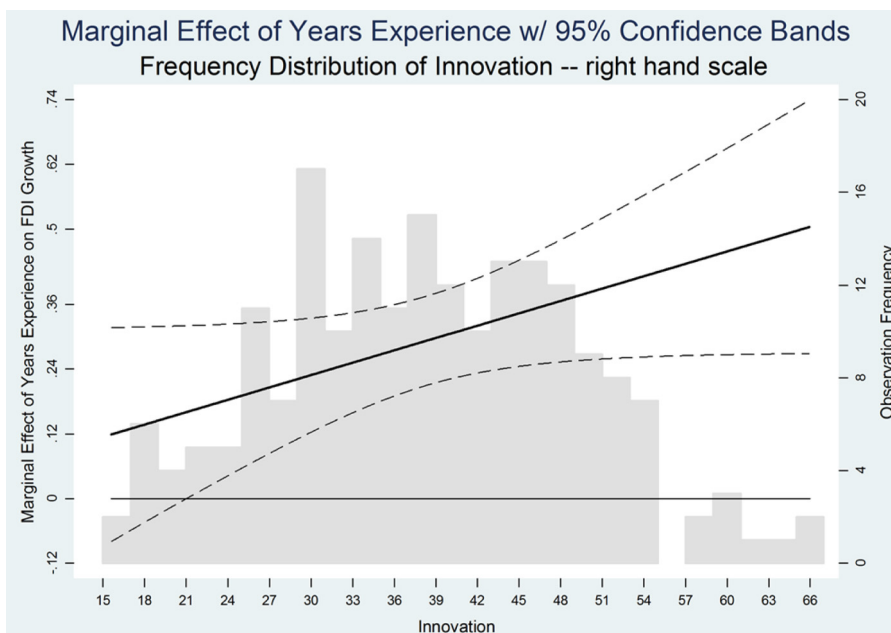

Fig. 2. Marginal Effect of Years Experience on FDI Growth based upon Model (2a) in Table 5.

Table 6
Monte Carlo Simulation Results.

Population Model	Share Potential Understating	Share Potential Overstating
1 (Eq. (8))	35.8%	8.0%
2 (Eq. (9))	43.5%	7.8%
3 (Eq. (10))	16.8%	73.8%

(p-value greater than 0.05) but only one of the estimated marginal effects, $\hat{\beta}_1$ or $\hat{\beta}_1 + \hat{\beta}_3$, is statistically different from zero (p-value less than 0.05), for one or both marginal effect(s). Similarly potential overstating problems arise when the coefficient on $\hat{\beta}_3$ is statistically significant but at least one of the marginal effect, $\hat{\beta}_1$ and/or $\hat{\beta}_1 + \hat{\beta}_3$, is not statistically different from zero, for one or both marginal effect(s). From the 25,000 different regressions for each model we calculate the share of regression estimates that show a potential understating and overstating problem. Table 6 below presents the results for the three population model simulations.

The table shows that the magnitude of the problem varies depending upon the underlying population parameters. For the first population model there is a 35.8% chance that a researcher would draw a sample that resulted in a $\hat{\beta}_3$ coefficient that is not statistically significant (Question 1), even though only one of the marginal effects is statistically different from zero (Question 2), suggesting some degree of moderation exists. This highlights the problem we find in the literature, where the majority of researchers stop their analysis upon finding that $\hat{\beta}_3$ is not statistically significant. Similarly, the simulation exercise shows that under certain population parameters (model 3 in Table 6) there can be a high likelihood of a researcher claiming support for a conditional hypothesis simply because the $\hat{\beta}_3$ coefficient is statistically significant, even though the lack of at least one statistically significant non-zero marginal effect suggests partial or minimal support.

This Monte Carlo simulation showcases the potential hazards of answering Question 1 to the exclusion of Question 2. The exercise reinforces the understating and overstating risks illustrated through the regression examples. Sample (1a) yielded conventional results, but samples (1b), (1c), and (2a) and, to some extent, (2b), showed how Question 1 and Question 2 address different facets of moderation, with a complete empirical test and description of the data requiring asking and answering both.

5. Discussion & recommendations

In this article, we aim to contribute to a broad, constructive conversation on conditional hypotheses in the IB literature and, specifically, how such models are commonly tested using regression models containing multiplicative interaction terms. We have analyzed the most common issues in testing and interpreting two-way interaction models. While we do not explicitly address issues related to nonlinearity (e.g. Zelner, 2009), nested or multi-level data (e.g. Hox, 2002; Kreft & de Leeuw, 1998; Luke, 2004; Peterson, Arregle, & Martin, 2012), three-way interactions (e.g. Andersson, Cuervo-Zazurra, & Nielson, 2014; Dawson, 2014), or nonmonotonicity (Schoonhoven, 1981), we note that our discussion, and guidance offered, is broadly applicable to these more complex cases. Additionally, we refer scholars to discussions about levels of measurement (Allison, 1977), symmetric interactions (Berry et al., 2012), false positive rates (Esarey &

Sumner, 2016), and flexible estimation strategies (Hainmueller, Mummolo, & Xu, 2016).

The key point we make is that researchers run the risk of overstating or understating results if they rely solely on interpreting the significance of the interaction coefficient to assess moderation – a risk shown by our review of the extant IB literature, simulated regressions, and a Monte Carlo exercise. Overstating occurs when the researcher proposes a conditional hypothesis and then assumes, based upon finding a statistically significant interaction coefficient, that the marginal effect for all the values of the moderating variable are statistically different from zero. This assumption, it turns out, is not always true. A second hazard – understating – occurs when a researcher with a conditional hypothesis misses evidence of one or more nonzero marginal effects based solely upon the finding of a statistically insignificant interaction coefficient.

We have advised that scholars seeking to exhaust their understanding of a conditional hypothesis test for both a statistically significant interaction coefficient (Question 1) and non-zero marginal effects (Question 2). As discussed, these are two very distinct questions. Should a researcher posing a conditional hypothesis believe that it is sufficient to address only the first question (i.e., examine only the statistical significance of the interaction term), it would seem to us incumbent upon that scholar to provide theoretical justification why evidence pertaining to the second question is (partially, largely, or entirely) irrelevant. As a general matter, we find it hard to envisage a case in which a researcher would care only about whether marginal effects differ from one another but not be concerned about knowing the specific values of the moderating variable at which marginal effects might be non-zero. Such a (hypothetical) situation would suggest that the statistical significance (or lack thereof) of the interaction coefficient is a sufficient test of the conditional hypothesis. Alternatively, there is the (hypothetical) case in which the researcher posits a conditional proposition containing at least one non-zero marginal effect, but where theory does not predict marginal effects that differ from one another. In this (hypothetical) situation, the statistical significance of the interaction coefficient would not be considered a necessary condition for evidence of moderation. Ideally, of course, the researcher specifies à priori which test(s) should be performed and what sort of statistical evidence would support the conditional hypothesis. However, although this is not a paper on theory development, as a practical matter we find that conditional hypotheses are frequently, if not most often, stated too generally to distinguish between Question 1 and Question 2. The result is that it is all-too-often unclear exactly what the researcher is either theorizing or statistically testing.

Our simulation and analysis of extant IB articles suggests there is indeed a risk that researchers are understating or overstating their interaction results. With IB articles, in particular, we find that the majority focus their empirical analysis and interpretation solely on the coefficient on the interaction term in the regression analysis. However, as our discussion has highlighted, restricting analysis to Question 1, and not answering Question 2, can result in researchers misstating the evidence in support of their conditional hypothesis. This materially diminishes the development and impact of IB theory.

To conclude, therefore, we offer the following recommendations for testing and interpreting interaction results to avoid overstating and understating the empirical support for the conditional hypothesis. Our recommendations aim to provide practical guidance to authors and reviewers in international business.

1. State the conditional hypothesis so as to clearly articulate whether marginal effects (i.e., relationships between the primary explanatory variable and the dependent variable) differ from one

another for any two values of the moderating variable (Question 1) and/or whether a marginal effect differs from zero for any specific value of the moderating variable (Question 2).

2. Properly specify the regression model by including the primary explanatory variable X and moderating variable Z, the interaction term XZ, and all relevant control variables (e.g., Eq. (1)).

$$Y = \beta_0 + \beta_1 X + \beta_2 Z + \beta_3 XZ + controls$$

3. Answer Question 1 by evaluating the statistical significance of $\hat{\beta}_3$ to ascertain whether the marginal effects are different from each other.

[a] Recognize that a statistically significant $\hat{\beta}_3$ provides evidence that marginal effects are discernibly different from one another.

[b] Recognize that a statistically non-significant $\hat{\beta}_3$ provides evidence that marginal effects are not discernibly different from one another.

4. Answer Question 2 by evaluating whether the marginal effects ($\frac{\delta Y}{\delta X} = \beta_1 + \beta_3 Z$) are statistically different from zero.

[a] For a dichotomous moderating variable Z, create a table of marginal effects similar to Table 3. (*Use the code and references available on the authors website.*)

[b] For a continuous moderating variable Z, create a figure similar to Fig. 1 or Fig. 2, in which a confidence interval is constructed around the marginal effect line over the entire range of the moderating variable. (*Use the code and references available on the authors website.*)

5. Report both the statistical significance of the coefficient estimate for $\hat{\beta}_3$ (Question 1) and the range of values of the moderating variable Z and percentage of observations for which marginal effects attain statistical significance (Question 2).

6. Discuss the level of support (e.g. full, partial, none) for the conditional hypothesis given the statistical results of both questions.

References

Aguinis, H., Edwards, J., & Bradley, K. (2016). Improving our understanding of moderation and mediation in strategic management research. *Organizational Research Methods*. http://dx.doi.org/10.1177/1094428115627498.

Ai, E., & Norton, E. (2003). Interaction terms in logit and probit models. *Economics Letters*, *80*, 123–129.

Aiken, L., & West, S. (1991). *Multiple R|regression: Testing and interpreting interactions*. London: Sage Publishing.

Allison, P. (1977). Testing for interaction in multiple regression. *American Journal of Sociology*, *83*(1), 144–153.

Andersson, U., Cuervo-Zazurra, A., & Nielson, B. (2014). Explaining interaction effects within and across levels of analysis. *Journal of International Business*, *45*, 1063–1071.

Andersson, U., Dasí, À., Mudambi, R., & Pedersen, T. (2016). Technology, innovation and knowledge: The importance of ideas and international connectivity. *Journal of World Business*, *51*(1), 153–162.

Berry, W., Golder, M., & Milton, D. (2012). Improving tests of theories positing interaction. *The Journal of Politics*, *0*, 1–19.

Blalock, H. M. (1965). Theory building and the statistical concept of interaction. *American Sociological Review*, *30*(June), 374–380.

Brambor, T., Clark, W. R., & Golder, M. (2006). Understanding interaction models: Improving empirical analyses. *Political Analysis*, *14*, 63–82.

Brouthers, K. D. (2002). Institutional, cultural, and transaction cost influences on entry mode choice and performance. *Journal of International Business Studies*, *33*(2), 203–221.

Dawson, J. (2014). Moderation in management research: What, why, when, and how. *Journal of Business and Psychology*, *29*, 1–19.

Delios, A., & Henisz, W. (2003). Policy uncertainty and the sequence of entry by japanese firms, 1980–1998. *Journal of International Business Studies*, *34*(3), 227–241.

Doh, J. P., Luthans, F., & Slocum, J. (2015). The world of global business 1965-2015: Perspectives on the 50th anniversary issue of the Journal of World Business: Introduction to the special issue. *Journal of World Business*, *51*(1), 1–5.

Esarey, J., & Sumner, J. (2016). Marginal effects in interaction models: Determining and controlling the false positive rate. *Manuscript*.

Grinstein, A., & Riefler, P. (2015). Citizens of the (green) world? Cosmopolitan orientation and sustainability. *Journal of International Business Studies*, *46*(6), 694–714.

Hainmueller, J., Mummolo, J., & Xu, Y. (2016). How much should we trust estimates from multiplicative interaction models? Simple tools to improve empirical practice. *Manuscript*.

Hox, J. (2002). *Multilevel analysis: Techniques and applications*. Mahwah, NJ: Lawrence Erlbaum Associates Publishers.

Jandhyala, S., & Weiner, R. (2014). Institutions *sans* frontiers: International agreements and foreign investment. *Journal of International Business Studies*, *45*(6), 649–669.

Jandhyala, S. (2015). International and domestic dynamics of intellectual property protection. *Journal of World Business*, *50*(2), 284–293.

Johanson, J., & Vahlne, J. E. (1977). The internationalization process of the firm – A model of knowledge development and increasing foreign market commitment. *Journal of International Business Studies*, *1*(8), 23–32.

Kogut, B., & Zander, U. (1993). Knowledge of the firm and the evolutionary theory of the multinational corporation. *Journal of International Business Studies*, *24*(4), 625–645.

Kreft, I., & de Leeuw, J. (1998). *Introducing multilevel modeling*. Thousand Oaks, CA: Sage Publications.

Luke, D. A. (2004). *Multilevel modeling*. Thousand Oaks, CA: Sage Publications.

Min, B., & Smyth, R. (2014). Corporate governance, globalization and firm productivity. *Journal of World Business*, *49*, 372–385.

Peterson, M. F., Arregle, J.-L., & Martin, X. (2012). Multilevel models in international business research. *Journal of International Business Studies*, *43*(5), 451–457.

Schoonhoven, C. B. (1981). Problems with contingency theory: Testing assumptions hidden within the language of contingency theory. *Administrative Science Quarterly*, *26*(3), 349–377.

Spiller, S., Fitzsimons, G., Lynch, J. R., & McClelland, J. (2013). Spotlights, floodlights, and the magic number zero: Simple effects tests in moderated regression. *Journal of Marketing Research*277–288 L (April).

Zelner, B. (2009). Using simulation to interpret results from logit, probit: And other nonlinear models. *Strategic Management Journal*, *30*, 1335–1348.