Journal of International Business Studies (2020) 51, 283–298

www.jibs.net

EDITORIAL

Misconceptions about multicollinearity in international business research: Identification, consequences, and remedies

Thomas Lindner[1], Jonas Puck[1] and Alain Verbeke[2,3,4]

[1] Institute for International Business, Vienna University of Economics and Business, Welthandelsplatz 1, 1020 Vienna, Austria; [2] Haskayne School of Business, University of Calgary, 2500 University Drive NW, Calgary T2T1N4, Canada; [3] Henley Business School, University of Reading, Reading, UK; [4] Solvay Business School, Vrije Universiteit Brussel, Brussels, Belgium

Correspondence:
A Verbeke, Haskayne School of Business, University of Calgary, 2500 University Drive NW, Calgary T2T1N4, Canada
e-mail: alain.verbeke@haskayne.ucalgary.ca

Abstract
Collinearity between independent variables is a recurrent problem in quantitative empirical research in International Business (IB). We explore insufficient and inappropriate treatment of collinearity and use simulations to illustrate the potential impact on results. We also show how IB researchers doing quantitative work can avoid collinearity issues that lead to spurious and unstable results. Our six principal insights are the following: first, multicollinearity does not introduce bias. It is not an econometric problem in the sense that it would violate assumptions necessary for regression models to work. Second, variance inflation factors are indicators of standard errors that are too large, not too small. Third, coefficient instability is not a consequence of multicollinearity. Fourth, in the presence of a higher partial correlation between the variables, it can paradoxically become more problematic to omit one of these variables. Fifth, ignoring clusters in data can lead to spurious results. Sixth, accounting for country clusters does not pick up all country-level variation.
Journal of International Business Studies (2020) **51**, 283–298.
https://doi.org/10.1057/s41267-019-00257-1

Keywords: multicollinearity; collinearity; regression analysis; hierarchical modeling; quantitative research methods

INTRODUCTION

International Business (IB) researchers frequently rely on quantitative empirical analyses. Indeed, more than a third of the Journal of International Business Studies (JIBS) articles available online (as of January 2019) include some form of regression analysis. As the unit of analysis is often the international firm, IB researchers are seldom able to collect data through controlled experiments where participants are randomly assigned to treatment and control groups, or through other ways of making sure observations are independent. As a result, IB researchers usually deal with large control models in their regression analyses in order to make their units of analysis comparable. Nielsen and Raswant (2018), among others, highlight that integrating the "right" control variables is a significant challenge for IB researchers. Consequently, IB studies may suffer

Online publication date: 5 August 2019

from omitted variable bias, which affects the robustness and validity of their results.

Extensive control models potentially pose a challenge in quantitative IB research, which in conjunction with dependence on typically small sets of historical data make it almost impossible to base quantitative analysis on unrelated variables. Survivor bias, for example, makes it difficult to investigate how balance sheet positions affect profitability since the firms for which data are available are those with a combination of resources that lead to company success. Hence, a large body of quantitative IB research features empirical analyses that have highly collinear variables. This is reflected in the fact that multicollinearity is acknowledged in about 400 of the approximately 1400 quantitative articles in JIBS available online as of January 2019; in at least 150 of them authors compute variance inflation factors (VIFs) to determine whether a variable introduces "too much" multicollinearity into a regression analysis. Just how much collinearity in terms of VIFs is too much varies from a high of 20 (Greene, 2003, Griffiths, Judge, Hill, Lütkepohl, & Lee, 1985), to over 10 (Wooldridge, 2014), to as low as five (Rogerson, 2001) or even three (Read & Read, 2004). In addition to there being no agreement about just what is too much collinearity, there is a wide range of strategies to address it. One of the most frequent (e.g. Meyer & Sinani, 2009; Muethel & Bond, 2013; Zhao, Park, & Zhou, 2014) is to exclude variables that have high partial correlations with other variables. While doing that can indeed reduce collinearity among variables, it simultaneously increases the risk of omitted variable bias.

Unlike in IB, multicollinearity among explanatory variables has received a lot of attention in econometric theory and in econometric texts (e.g., Goldberger, 1991; Greene, 2003; Wooldridge, 2014). This is not to say that the topic has been neglected in IB. Editorials (e.g., Cuervo-Cazurra, Andersson, Brannen, Nielsen, & Reuber, 2016) have helped considerably in raising awareness and have also offered potential solutions. Nonetheless, there is some misunderstanding about how findings from econometric theory can be applied to empirical research in IB. We identify in this paper three important gaps between econometric theory and the practice of applied IB research. First, there are almost no illustrations in econometric textbooks of how problems arising from potential collinearity between explanatory variables can be detected. Second, there is a paucity of simulations that illustrate how not treating multicollinearity, or treating it insufficiently, may lead to severe bias, spurious findings, and flawed conclusions. Third, econometric textbooks tend to illustrate the effect of the violation of an assumption on point estimates or their variance with "clean" examples, i.e., simple cases that deal with one problem at a time. Yet, IB research usually deals with complex relationships between many interrelated variables. To the best of our knowledge, there has been no investigation into how specific econometric and data problems affect regression outcomes in so-called "messy" cases where, among other issues, multicollinearity is present.

We address all three gaps. First, we use a simulation to show how researchers can identify multicollinearity issues in regression results. There has been little discussion of what happens when researchers drop or compound collinear variables to reduce multicollinearity – and hence VIFs. We provide evidence of the consequences of such omissions. Second, we show how instability in results (Enright, 2009) is related to multicollinearity between explanatory variables, and explain why that is. We discuss how instability arises from inappropriate treatment of multicollinearity, and respond to calls for explanations of how this could be remedied. For instance, Kalnins (2018) investigates how collinearity affects coefficient instability under measurement error. We first investigate collinearity between relevant independent variables, and then between relevant and irrelevant independent variables. We combine the findings to develop a decision matrix that researchers can utilize to determine which situation they may want to avoid, and when using it, outline strategies to detect and respond to potential multicollinearity issues. Third, we present a complex simulation example that mimics the data analysis process in a typical quantitative IB research project. We use the example to show how multicollinearity may affect research outcomes if treated inappropriately. Finally, we summarize our findings and suggestions and based on them, we develop a list of recommendations.

This paper contributes to the IB literature by highlighting misconceptions about collinearity and how empirical challenges associated with it can be addressed. We argue that multicollinearity does not introduce bias into regression results and that result instability is not a consequence of it. We do show which effects researchers can expect if multicollinearity is addressed inappropriately or insufficiently. As the true relationship between dependent and independent constructs usually is unknown, the kinds of simulations and illustrations we present can help researchers in identifying which econometric problems they might be facing and how they can expect their results to change under different specifications. In doing so, we contribute to the body of methodological literature by reducing the gap between "clean" methodological results in econometric theory and the application of such theory to real-world, "messy" IB research phenomena.

MULTICOLLINEARITY

Multicollinearity is typically one of the most discussed issues in the empirics section of IB manuscripts. In JIBS, approximately 400 of the articles available online (as of January 2019) explicitly address the topic. Much of the discussion centers around VIFs and their detrimental effect on regression results. Our review of papers revealed two main sub-issues: variance inflation and coefficient stability given high collinearity.

It is acknowledged in the econometric literature that "[…] the 'problem' of multicollinearity is not really well defined" (Wooldridge, 2014, p. 83), but there is nonetheless consistency in how specific consequences of collinearity can (or cannot) be treated. Specifically, econometric textbooks argue that multicollinearity is caused by small sample size, i.e., "micronumerosity" (e.g., Goldberger, 1991), and that its solution lies in collecting more data. This also means that multicollinearity is a sample characteristic: when two variables exhibit collinearity in one sample, it does not follow that they will be collinear in all other samples as well. Consequently, the problems discussed most in IB research – coefficient stability and variance inflation – have relatively straightforward answers when drawing on econometric theory.

Nonetheless, there is still controversy in IB over how to successfully address collinearity. We think this is due partly to a lack of examples using simulation, and partly also to the available examples being too "clean" to improve our understanding, when in fact IB research contexts tend to be "messy", i.e., there are many non-independent variables and sample sizes tend to be relatively small. Consequently, we will address stability and variance inflation using simulations and a more complex example than usually found in the literature.

In this paper, "collinearity" and "multicollinearity" refer to a relationship between two or more distinct constructs, not two or more measures of the same construct. Before moving to possible solutions, we will, using a minimum of econometric theory and several illustrations from simulation, investigate the effect that multicollinearity has on point estimators and estimator variance when multicollinearity between explanatory variables is present. First, we will look into a case where both collinear variables are part of the true relationship between independent and dependent constructs. We will then investigate the effects of multicollinearity between independent constructs when one of these constructs is not part of the proposed relationship between independent and dependent constructs. In both cases, the hypothetical research context for simulation is that of the determinants of a firm's foreign direct investment (FDI).

For this simulation, we assume in very generic terms, and as a stylized but realistic example, that the amount firms invest abroad could be a function of three variables, typical for relatively young MNEs. First, we acknowledge the role of prior international experience (*intexp*), reflecting the firm's – or its founders' – demonstrated ability to work across borders and the learning associated with this prior work, thereby indirectly measuring the non-location boundedness of its firm-specific advantages (FSAs), see Verbeke, Zargarzadeh, and Osiyevskyy (2014). Second, firm size (*size*) reflects the firm's proven ability to grow domestically, itself the result of underlying location-bound FSAs in terms of, e.g., innovative knowledge or entrepreneurial effectiveness (whereby the question of course arises, whether these domestically relevant FSAs can be deployed effectively across borders). Third, there is organizational slack (*slack*) in the sense of a quantity of unused resources that from a Penrosean perspective can be deployed to expand into foreign markets (Verbeke & Yuan, 2013). Firm size and international experience thus reflect in generic terms the quality of the firm's resource base, whereas slack represents the quantity of resources available for expansion. Consequently, in the following subsections we investigate variations of the relationship:

$$\mathrm{FDI} = b_0 + b_1 \cdot intexp + b_2 \cdot size + b_3 \cdot slack + N(0, 0.5).$$

Wang, Hong, Kafouros, and Wright (2012) investigate a similar relationship using a much more complex econometric model, which, nevertheless, contains the three variables presented above. For the sake of simplicity, we assume the above relationship to hold. Unless indicated otherwise, we ran 100 models with different correlations between international experience (*intexp*) and *size*, keeping *slack* independent. The observations for *intexp* and *slack* will be drawn from a covariate distribution with correlations varying from just above -1 to just below 1. This means that multicollinearity will vary across the models. We will investigate the effects of collinearity on coefficients, errors, and other model parameters. Each of the 100 models will have 1000 simulated observations. The R code and data for all simulations presented in this paper are available from the authors upon request.

Collinearity between Relevant Independent Variables

In this subsection, we assume that FDI is a function of international experience, firm size, and slack, following the specification

$$\mathrm{FDI} = 2 \cdot intexp + 4 \cdot size + 3 \cdot slack + N(0, 0.5).$$

Variance inflation

In the first analysis, we are interested in how the correlation between *intexp* and *size* affects the standard errors of the coefficients in the regression. The model estimated is

$$\widehat{\mathrm{FDI}} = \beta_0 + \beta_1 \cdot intexp + \beta_2 \cdot size + \beta_3 \cdot slack + \varepsilon.$$

If the estimation works appropriately, we should see the true coefficients in our regression output, i.e., β_0 should be 0, β_1 should be 2, β_2 should be 4, and β_3 should be 3. The left panel of Figure 1 shows the standard errors for the variables in the regression depending on the correlation between the observations for *intexp* and *size*. It is evident that increasing the correlation (in absolute terms) between *intexp* and *size* leads to an inflation of the respective standard errors, but importantly, the standard error of the coefficient for *slack* does not change with this collinearity. Thus, there is no evidence for "smearing" (Greene, 2003, p. 265) of the error inflation to other independent variables. In this context, "smearing" would mean that the standard errors of the other coefficients are affected by collinearity between international experience and slack.

The right panel of Figure 1 shows the square root of VIFs against the correlation between *intexp* and *size*. The VIFs for *intexp* and *size* are almost equal, therefore only one of the two lines is clearly visible. Comparing the left and right panels, we see that there is a strong relationship between the VIFs and error inflation. The VIF of *slack*, which is quite constant at 1 throughout the bandwidth of correlations, again indicates that collinearity between *intexp* and *size* does not "smear" over to other independent variables. The strong relationship

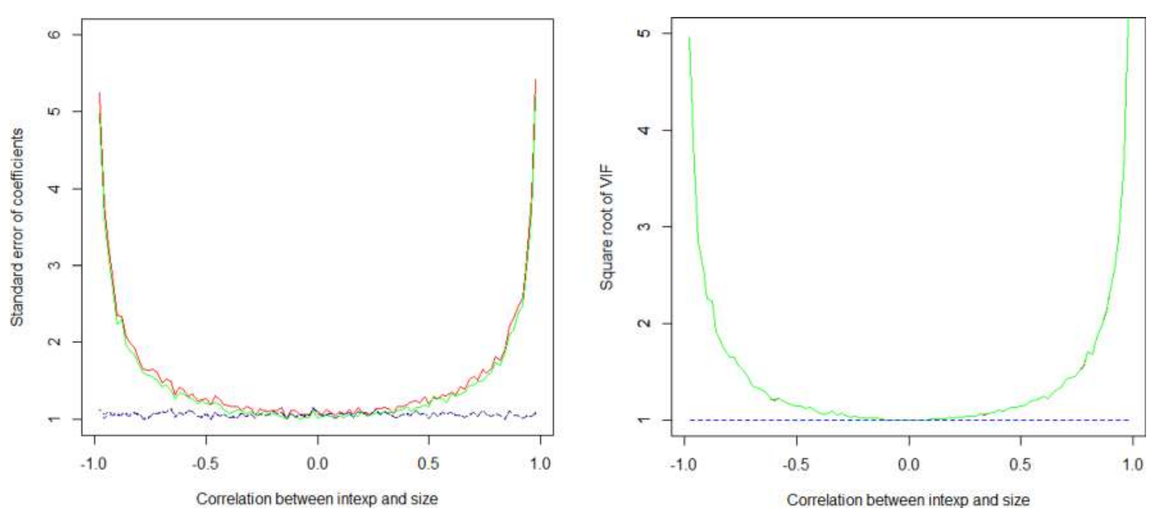

Figure 1 The effect of correlations on standard errors and VIFs. *Left panel*: standard error of coefficient for *intexp* (*red*), *size* (*green*), *slack* (*blue*, *dashed*), and constant (*black*, *dotted*). *Right panel*: variance inflation factors of *intexp* (*red*), *size* (*green*), and *slack* (*blue*, *dashed*).

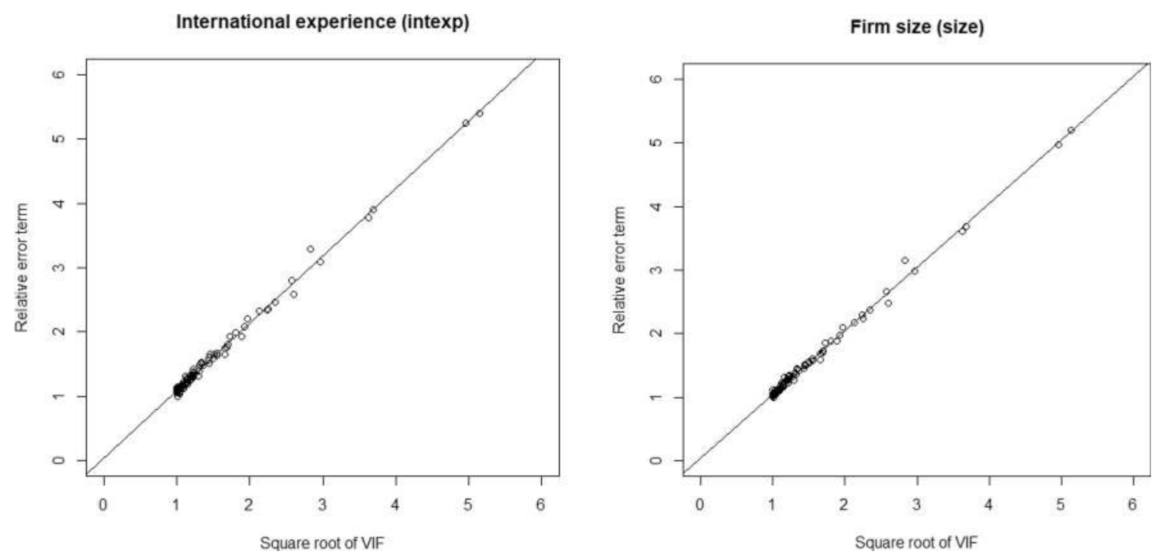

Figure 2 The relationship of standard errors with VIFs. Relationship between relative error terms and square root of VIF for international experience (*left*) and firm size (*right*).

between the square root of the VIFs for *intexp* and *size* call for an illustration of the direct relationship between the two. This is given in Figure 2. It is clear that there is an almost perfect model fit between them. The regression lines for international experience in the left panel of Figure 2 and firm size in the right one almost perfectly capture the variation of the two.

This illustration shows that variance inflation factors are representations of how much error terms are inflated if multicollinearity is present in a linear regression model. The square root of the VIFs is the factor by which error terms are inflated compared to when there is no multicollinearity. It is important to note, however, that inflated variance does not lead to spurious results. In fact, the opposite is true: The higher the collinearity, i.e., the higher the VIFs, the harder it is to find statistical evidence of a relationship. If collinearity between two variables increases, the standard errors in a regression are inflated correspondingly (hence the name for VIFs "variance inflation factors"). If we compare two scenarios, whereby the effects of two independent variables on a dependent construct are the same, but the two independent variables are correlated to different degrees, we will obtain higher standard errors (and consequently lower t values, and higher p values) the higher the correlation (in absolute terms) between the two independent variables. This means that the statistical support for a relationship as suggested by a regression analysis becomes weaker if the independent construct is highly correlated with another independent (or control) variable. In the following section, we explore further how the statistical evidence for a relationship between independent and dependent variables changes with collinearity under different scenarios.

Spurious coefficients

Given that VIFs are fully explained by the relative error terms of the collinear variables, the next step is to investigate the effect of collinearity on the coefficients in a regression. Figure 3 shows that there is no relationship between the correlation of *intexp* and *size* and any of the coefficients in the regression. The red line for β_1, the coefficient of *intexp*, is stationary at its real value of 2, the green line indicating β_2, the coefficient of *size*, is stationary at 4, the dashed blue line indicating β_3, the coefficient of *slack*, is stationary at 3, and the dotted black line, the intercept, is stationary around 0. The differences in the degree to which the lines in Figure 3 fluctuate are the result of how much variance the respective variables are assumed to have when we create them in the simulation. The fact that no trend can be observed for any of the variables with increasing correlations shows that collinearity does not affect the coefficients in a regression.

In reality, it is quite rare in IB research that the "true" underlying relationship is known. Consequently, researchers frequently discuss collinearity

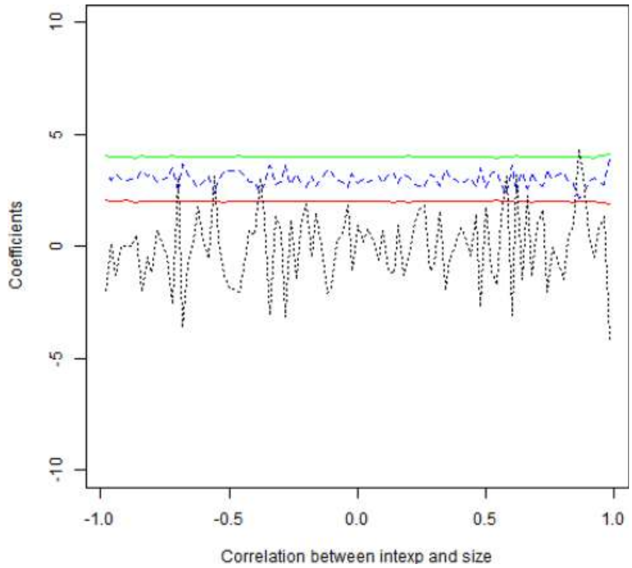

Figure 3 The effect of correlations on regression coefficients. Relationship between correlation of *intexp* and *size* and the coefficients of *intexp* (*red*), *size* (*green*), *slack* (*blue*, *dashed*), and the intercept (*black*, *dotted*) in the regression.

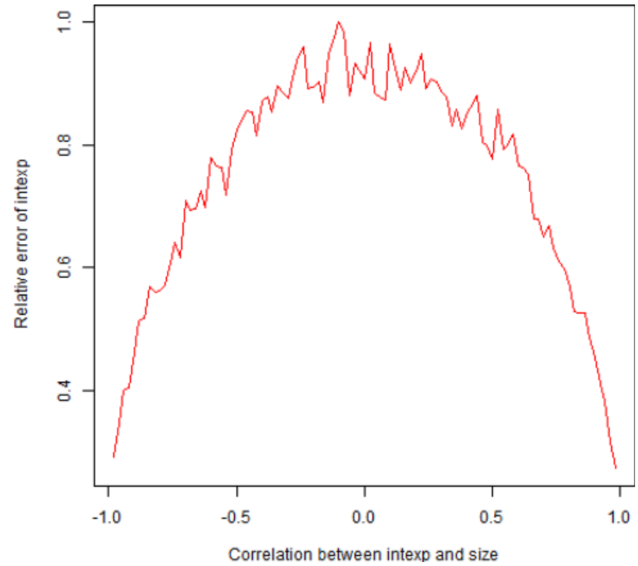

Figure 4 The effect of variable omission on error deflation. Relationship between correlation of *intexp* and *size* and the standard error of β_1', the coefficient for international experience.

when the size or the significance of a coefficient changes as a collinear variable is added to a model. We illustrate this by estimating the equation

$$\widehat{\mathrm{FDI}}' = \beta_0' + \beta_1' \cdot intexp + \beta_3' \cdot slack + \varepsilon',$$

while the true dependency of FDI on experience, size, and slack still looks the same as before:

$$\mathrm{FDI} = 2 \cdot intexp + 4 \cdot size + 3 \cdot slack + N(0, 0.5).$$

When a collinear variable is omitted, VIFs are low, and there is no indication that a relevant variable has been omitted. In our case, the standard error of the coefficient of β_1' for international experience (*intexp*) is significantly reduced if a highly collinear variable (here firm size) is omitted from the regression, as is clear from Figure 4. Thus, omitting a relevant but collinear variable is problematic in a regression because it deflates standard errors and may lead to spurious findings. Of course this is not what researchers intend when they drop one of the two variables to "remedy" high partial collinearity. As we show in this simulation, this can create spurious significance by deflating the standard error of the included variable. If in doubt, a researcher would be well advised to keep the variables in the regression model. Although this may inflate standard errors, it will not create spurious results. As Wooldridge (p. 83) put it when writing on the same topic, "[…] multicollinearity violates none of our assumptions."

While it is undeniably problematic to omit a collinear variable from a regression as it may lead to spuriously significant coefficients, it is much worse if it introduces bias into the coefficient of the remaining variables themselves. To investigate the extent to which this is the case, we again simulate the incomplete regression model (with coefficients β_i') shown above. We also run the complete regression model:

$$\widehat{\mathrm{FDI}} = \beta_0 + \beta_1 \cdot intexp + \beta_2 \cdot size + \beta_3 \cdot slack + \varepsilon$$

using the same data as in the incomplete one. We compare the coefficients β_i' to β_i. Figure 5 shows the results of the simulation for the coefficients β_1' and β_1 of international experience (*intexp*) if the collinear variable firm size (*size*) is omitted (red, dashed) or included (red, solid).

Figure 5 shows that omitting a relevant, but collinear, variable may not only lead to deflated error terms and hence spuriously significant coefficients, but also to biased ones. The degree of bias will depend on the degree of correlation between the variable included and the one omitted. Hence, the omission of relevant but collinear variables may create spuriously significant results, it may severely bias coefficients, it may even do both at the same time, yielding spuriously significant biased coefficients. In our case, a researcher who omits the

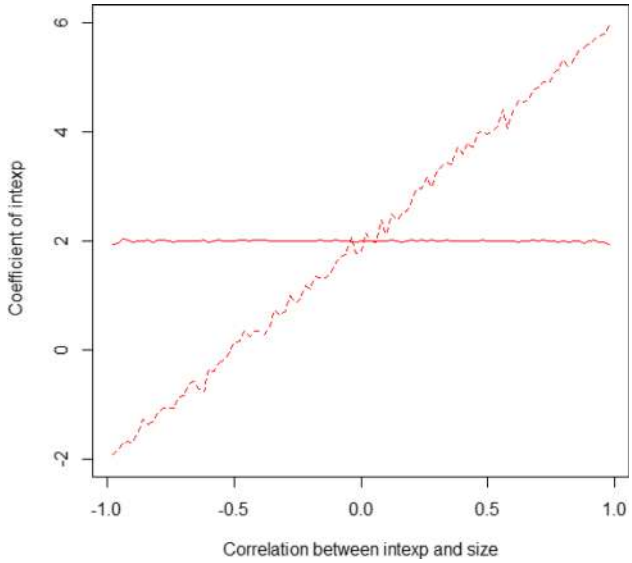

Figure 5 The effect of variable omission on coefficient estimates. Relationship between correlation of *intexp* and *size* and the coefficients β_1 (red, solid) and β_1' (red, dashed).

Table 1 The effect of variable omission on bias in regression results

	(1)	(2)
Intercept	30.248	− 0.212
	(7.867)	(1.576)
International experience	− 1.205	2.036
	(0.075)	(0.026)
Firm size		3.989
		(0.026)
Slack	2.545	3.011
	(1.572)	(0.312)
Observations	1000	1000
Adjusted R^2	0.206	0.969

Regression results for incomplete (1) and complete (2) models. Coefficients β_1' (model 1) and β_1 (model 2) for international experience differ significantly. Correlation between *intexp* and *size* equals − 0.80. True coefficients are 0 for the intercept, 2 for international experience, 4 for firm size, and 3 for slack. Standard errors are in parentheses.

variable firm size from a regression model when there is a negative correlation between international experience and firm size, might conclude that international experience has a negative effect on a firm's FDI. We can also see from Figure 5 that the higher the collinearity (in absolute terms) between two variables, *the more severe* the consequences of an omission. At the same time, including collinear variables in the regression only inflates standard errors and hence makes regression results, if they turn out to be significant, more conservative. This is because collinearity, when one has independent variables in a correctly specified model, makes the estimation more efficient and increases standard errors (as expressed by VIFs) while leaving coefficient estimates unbiased.

This result also explains what many researchers in IB have found when they add to a regression model a variable that is highly collinear with one already in the model: significance levels change substantially and the sign of the coefficient of the variable previously in the model changes dramatically. Table 1 shows such a case. If the correlation between *intexp* and *size* is − 0.8, a researcher who omits *size* from the regression model may conclude that there is a significant negative effect of international experience on FDI (model 1, $p < 0.0001$). This result, however, is spurious as the true underlying relationship (as specified above) suggests that the coefficient on *intexp* should be 2. If the collinear variable *size* is added to the regression model, the researcher will find a significant positive relationship between international experience and FDI (model 2, $p < 0.0001$). Given that in reality we do not know the underlying "true" model, it is important not to drop collinear variables, as doing so may lead to spurious, biased, or both spurious and biased results. Researchers – and reviewers – might draw the conclusion that high collinearity makes the coefficient of international experience "unstable". We hope that our illustration has shown this to be flawed thinking. The problem in model 1 is not "instability caused by collinearity" but omission of a relevant collinear variable. Of course, it could happen that two highly collinear variables have some outlier observations such that when controlling for one and adding the other to the model, outliers become highly influential and lead to a flip in the signs of the coefficients. In such a case, the problem is not collinearity but too few observations, because this is what drives the impact of outliers on regression results. This brings us back to Goldberger (1991) equating a multicollinearity problem with a "micronumerosity" one.

Collinearity between Relevant and Irrelevant Independent Variables

So far, our discussion has focused on collinearity between two variables that are both relevant for the dependent construct. Even though theory frequently gives a researcher a good starting point, in the real world, the "true" underlying relationship between independent and dependent constructs is not known. Let us address cases where one of the collinear independent variables is irrelevant, i.e., not part of the "true" relationship. Given that our illustrations up to now show that the coefficients in a complete model are independent of VIFs (and the other way around), we will only discuss potentially spurious results resulting from omission here. The "true" underlying relationship between the three independent variables international experience (*intexp*), firm size (*size*), as well as *slack*, and the dependent variable firm FDI is now

$$\begin{aligned} \mathrm{FDI} &= 2 \cdot intexp + 0 \cdot size + 3 \cdot slack + N(0, 0.5) \\ &= 2 \cdot intexp + 3 \cdot slack + N(0, 0.5). \end{aligned}$$

Figure 6 shows the relative errors of the non-omitted, independent variable for the following regressions (where first *size* and then *intexp* are omitted):

$$\widehat{\mathrm{FDI}'}_1 = \beta'_{0,1} + \beta'_{1,1} \cdot intexp + \beta'_{3,1} \cdot slack + \varepsilon'_1,$$

$$\widehat{\mathrm{FDI}'}_2 = \beta'_{0,2} + \beta'_{2,2} \cdot size + \beta'_{3,2} \cdot slack + \varepsilon'_2,$$

In the left panel of Figure 6, the irrelevant variable, firm size, is omitted from the regression. As one would expect, the omission of an irrelevant variable has no effect on the variance of the remaining coefficients. This also holds true if the omitted variable is collinear with one that is included. On the other hand, and as we have shown previously, if the relevant variable international experience is omitted, the error of the irrelevant variable firm size is deflated. The omission might consequently lead to a spuriously significant impact of firm size on FDI while the true impact should be zero.

Given the potential for error deflation, omitting a collinear variable may even lead to spurious results if coefficients are biased sufficiently far from their true value. This can be especially problematic if the true value of the coefficient is zero, i.e., there is no underlying relationship, as in the case of firm size (*size*) in this example. To show this, we again run both the regression with omitted variables and that including both *intexp* and *size*:

$$\widehat{\mathrm{FDI}}_1 = \beta_{0,1} + \beta_{1,1} \cdot intexp + \beta_{2,1} \cdot size + \beta_{3,1} \cdot slack + \varepsilon_1$$

Figure 7 shows the dependency of the coefficients for international experience (*intexp*) and firm *size* on the correlation between *intexp* and *size* in regressions when the respective other variable is omitted. In the left panel, the coefficients in the correctly specified and mis-specified models are almost the same. This tells us that omitting a collinear but irrelevant variable does not introduce bias in the coefficient of the collinear relevant variable. Omitting a relevant collinear variable, however, deflates errors and biases the coefficient of the irrelevant variable, as shown in the right panel.

Finally, we illustrate how an econometrician might make an incorrect interpretation of changes in regression results that may be attributed to "multicollinearity". Table 2 shows regression results for estimations with firm size omitted (model 1), with international experience omitted

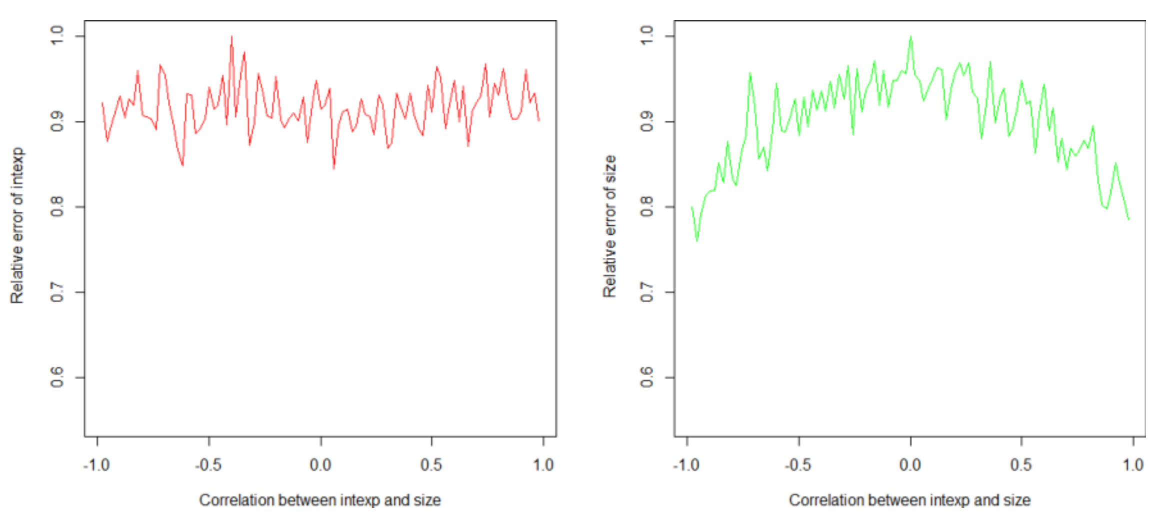

Figure 6 The effect of omitting irrelevant and relevant collinear variables on standard errors. Relationship between correlation of *intexp* and *size* and the (*left panel*) error of coefficient $\beta'_{1,1}$ of international experience (*intexp*) as well as the (*right panel*) error of coefficient $\beta'_{2,2}$ of firm size (*size*) with the respective other variable omitted from regression.

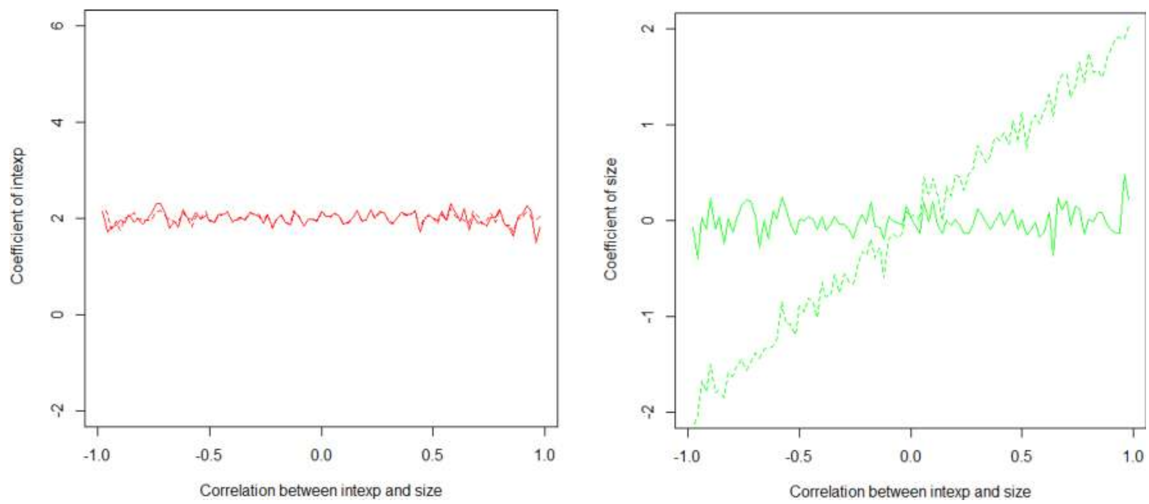

Figure 7 The effect of omitting irrelevant and relevant collinear variables on regression coefficients. Relationship between correlation of *intexp* and *size* and the (*left panel*) coefficient of international experience as well as (*right panel*) coefficient of firm size. *Dashed lines* are for coefficients $\beta'_{1,1}$ and $\beta'_{2,2}$ when the respective other variable (*intexp* and *size*) are omitted (models 1 and 2 in Table 2), and *solid lines* when they are included (coefficients $\beta_{1,1}$ and $\beta_{2,1}$ in model 3 in Table 2).

Table 2 The effect of omitting irrelevant and relevant collinear variables on regression results

	(1)	(2)	(3)
Intercept	0.917	8.994	0.892
	(1.596)	(4.054)	(1.597)
International experience	1.988		2.004
	(0.016)		(0.027)
Firm size		1.550	− 0.019
		(0.040)	(0.026)
Slack	2.824	2.276	2.824
	(0.318)	(0.809)	(0.318)
Observations	1000	1000	1000
Adjusted R^2	0.939	0.602	0.938

Regression results for coefficients $\beta'_{i,1}$, and $\beta'_{i,2}$, as well as $\beta_{i,1}$. The correlation between *intexp* and *size* equals 0.8. Standard errors are in parentheses.

(model 2), and with a fully specified model (model 3). In this simulation, the correlation between international experience and firm size is set to 0.8 (as opposed to − 0.8 in the above example). Yet, we would expect, given what is shown in Figures 6 and 7 that omitting an irrelevant correlated variable (*size* in model 1) would not affect the true relationships in model 1. Indeed, the coefficient for international experience is unbiased in model 1. However, if the relevant variable *intexp*, which is correlated with the irrelevant variable *size*, is omitted, a researcher might find spuriously significant effects of *size* on FDI (model 2, $p < 0.0001$). As is shown in model 3, there is no statistically significant relationship between firm size and FDI ($p < 0.5$) despite the spuriously significant positive result on firm size in model 2.

Using simulations, we have highlighted the problems multicollinearity creates in multiple regression analysis. The analytical background of our illustrations is readily available in econometrics texts such as those by Wooldridge (2014) and Greene (2003). The most important conclusion to be drawn is that multicollinearity, while it may inflate standard errors, is not an econometric problem per se. In other words, it does not reduce the degree to which regression results are valid, but only underestimates reliability. The most important issue is the risk of obtaining biased results through the incorrect omission of relevant variables. We hope that we have convincingly shown that this may lead to spurious significance and bias and thus render a regression analysis useless. At the same time, multicollinearity in a regression model is likely to do no worse than inflate standard errors which, although not optimal, will do no more than make results more conservative. Moreover, adding collinear variables to a linear model does not introduce instability. Nonetheless, we agree that the interpretation of coefficients of highly collinear variables is difficult and may require a "joint interpretation", but this is not an econometric problem as such. We highly recommend, if there were any doubt, that collinear variables be included, rather than excluded from models. If all relevant variables are accounted for, the parameter

Table 3 Consequences of different treatments of collinearity in regression models

	Collinear variable included	Collinear variable not included
Collinear variable relevant	(1) Error inflation	(3) Biased coefficients, deflated standard errors
Collinear variable irrelevant	(2) Error inflation	(4) Unbiased results

Overview of consequences if a relevant/irrelevant collinear variable is included/not included in a regression model.

estimates in the full model are unbiased despite collinearity (Cohen, West, & Aiken, 2014).

Table 3 gives an overview of the consequences of including or not including a relevant versus irrelevant collinear variable in a regression model. Naturally, not including an irrelevant variable has no detrimental consequences on regression results (quadrant 4 in Table 3). It is clear that including a collinear variable, regardless of whether it is relevant, leads to error inflation and an increase in VIFs (quadrants 1 and 2 in Table 3), which makes it more difficult for the researcher to identify relevant relationships. It is important to note, however, that not including a relevant collinear variable will lead to both biased coefficients and deflated error terms. This means that there will be spurious findings on any coefficients with non-zero correlations with the excluded collinear variables, and deflated error terms for all those coefficients (quadrant 3 in Table 3).

A "MESSY" EXAMPLE

Our illustrations have focused on isolated cases so far. One reason so few findings from econometric theory are fully used in applied research is that econometric textbooks lack examples that feature the kind of "messy" problems facing IB researchers. We therefore present a more complex (though still very stylized) illustration, that could be found in the empirical section of a research paper. This time the illustration centers on how much FDI, subsidiaries receive from their parent firm abroad. The data are for five countries with a different average FDI inflow per subsidiary. Differences in FDI flows could be driven by differences in GDP (as a proxy for market size, and thus host-country location advantages), and modeled by a random intercept $\gamma_{0,j}$.

Such a data structure is known as "hierarchical", and the literature suggests that "multilevel analysis" (e.g., Hox, 1995), "hierarchical modeling" (e.g., Steenbergen, 2012), or "random coefficient models" (e.g., Alcácer, Chung, Hawk, & Pacheco-de-Almeida, 2018) are ways to consider data structured into two (or more) levels. These methods provide significant opportunities for IB research where data are often naturally hierarchical. Yet, ignoring the hierarchical data structure in econometric specifications leads to misspecification, and may be the reason for the spurious findings of country-level variables influencing firm-level decisions. The general idea is that, similar to the multicollinearity issues discussed in this paper, inappropriate treatment of clusters may introduce bias into regression results. It is important to note, however, that the reason for this bias is an unconsidered association between observations, whereas with multicollinearity, the source of bias is the potential omission of a variable. As in the multicollinearity case, the reason is bias resulting from non-independent error terms. The model specification shown below accounts for differing means in the predicted dependent variable (i.e., the $\gamma_{0,j}$) for the different countries *j*. This specification is what some scholars call a fixed-effect hierarchical model (Hox, 1995), or a random intercept model (Alcácer et al., 2018). It allows for a constant term that is country-specific by including a dummy for all countries but one (to avoid the dummy-variable trap).

We again present a stylized example to illustrate the broader points we want to make. In this stylized example, FDI inflows into subsidiaries are a function of six variables. First, market size in a subsidiary country can be important to determine how much firms will be prepared to invest in the subsidiary, at least in the context of market-seeking investment (e.g., Driffield & Munday, 2000). Second, FDI inflows will increase as a function of the length of time during which a subsidiary has been operating in a target market (e.g., Asmussen, Pedersen, & Dhanaraj, 2009). Third, we speculate, in line with the concept of subsidiary-specific advantage (Rugman & Verbeke, 2001), that affiliates with substantial slack resources, which they can use for further investments, will receive less FDI, because slack allows them to invest further without additional funds from headquarters. Fourth, if subsidiaries are run by more experienced CEOs, they will receive more FDI because those CEOs are supposedly more competent to scale up the business (e.g., Reuber & Fischer, 1997). Fifth, similar to the example in the section before, a subsidiary's ability to generate intangible assets will increase investment in this subsidiary. Finally, political stability in the subsidiary country will also positively influence FDI inflows (e.g., Habib & Zurawicki, 2002). The "true" model specification for FDI inflows into subsidiary i in country j takes the following form:

$$\text{FDIin}_{i,j} = \gamma_{0,j} + 3 \cdot customers_i + 4 \cdot age_i - 5 \cdot slack_i + 6 \cdot CEOexperience_i + 7 \cdot intangibleassets_i + 800 \cdot politicalstability_j + N(0,1).$$

The data that we use, for illustrative purposes only, are again simulated. In contrast to the previous examples, we create data that are to a different extent collinear, and which have a more complex data structure, with clusters on the country level. Values for a subsidiary's number of customers, age, slack, CEO experience, and intangible assets (all dimensions of subsidiary level) are drawn from a multivariate distribution with defined correlations (see Table 4). Political stability is simulated to be independent of the subsidiary-level variables beyond the country clusters. Political stability varies only on the country level (j). To illustrate how ignoring the clustered data structure can bias regression results, we introduce the variable "subsidiary registration number" into the model. This variable is quite obviously unrelated to FDI inflows, but we will see how ignoring the data structure may make an econometrician believe the "subsidiary registration number" may be an important determinant of FDI inflows. This variable is simulated to be independent from all other variables, and clustered on the country level.

Since the sample size is not very large (500 observations), even the variables drawn from independent distributions will have non-zero correlations with the other variables. The correlation matrix in Table 4 reflects the actual correlations, with substantial partial correlation between the variables drawn from a joint distribution (number of customers, age, slack, CEO experience, and intangible assets). Descriptive statistics of the dependent and independent variables are given in Table 4 as well.

In the next step, we show what results a researcher might obtain when fitting different models to the simulated data. Table 5 shows regression results. We ran five OLS models (models 1–5) and two fixed effect models that account for country-level differences in the mean amount of FDI received by subsidiaries. Models 1 and 2 show that omitting relevant variables that are not independent of those in the "true" model wreaks havoc with the conclusions that are drawn from linear modeling, as is illustrated in the second section of this paper. From the results of model 1, a researcher could conclude that a subsidiary's number of customers has a positive effect on inward FDI ($p < 0.000000051$). If subsidiary age is controlled for, however, the conclusion from model 2 would be that the number of customers has a negative effect on the FDI a subsidiary receives ($p < 0.047$). Thus, the direction of the effect changes because a previously omitted relevant variable is now included.

The effect of the number of customers on FDI inflows disappears in model 3 ($p < 0.44$) when we

Table 4 Descriptive statistics and partial correlations for the variables in the "messy" illustration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
FDIin (1)	1	0.216	0.614	− 0.336	0.435	0.242	− 0.550	0.674
Customers (2)	0.216	1	0.479	0.325	− 0.142	− 0.070	− 0.095	0.107
Age (3)	0.614	0.479	1	− 0.257	0.410	0.031	− 0.014	0.025
Slack (4)	− 0.336	0.325	− 0.257	1	− 0.010	− 0.072	− 0.012	0.022
CEO experience (5)	0.435	− 0.142	0.410	− 0.010	1	0.028	0.057	− 0.055
Intangible assets (6)	0.242	− 0.070	0.031	− 0.072	0.028	1	0.023	− 0.035
"Sub. Reg. No." (7)	− 0.550	− 0.095	− 0.014	− 0.012	0.057	0.023	1	− 0.815
Political stability (8)	0.674	0.107	0.025	0.022	− 0.055	− 0.035	− 0.815	1
Mean	78.565	4.957	2.979	0.176	6.036	0.541	321.718	0.016
SD	15.376	1.015	0.980	0.992	0.988	0.516	107.250	0.013
Min	41.164	2.259	− 0.247	− 2.963	3.359	− 1.162	− 11	0.005
Max	125.490	8.047	6.025	2.884	9.192	2.235	543	0.040

Table 5 Regression results for the "messy" illustration

	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	56.790	50.774	47.765	8.022	36.832	16.232	− 0.619
	(3.362)	(2.720)	(2.735)	(5.010)	(3.075)	(5.137)	(0.599)
Number of customers	3.552	− 1.172	0.522	4.489	3.359	3.186	3.176
	(0.642)	(0.587)	(0.675)	(0.759)	(0.446)	(0.076)	(0.076)
Company age		10.108	8.491	3.433	3.920	3.778	3.787
		(0.607)	(0.684)	(0.838)	(0.491)	(0.084)	(0.083)
Slack			− 2.990	− 5.522	− 5.118	− 5.216	− 5.211
			(0.626)	(0.641)	(0.375)	(0.064)	(0.064)
CEO experience				5.881	6.014	6.134	6.129
				(0.640)	(0.374)	(0.064)	(0.064)
Intangible assets	7.701	6.448	6.363	6.549	6.801	7.070	7.078
	(1.261)	(1.014)	(0.993)	(0.919)	(0.537)	(0.092)	(0.091)
"Subsidiary registration number"					− 0.080	− 0.002	− 0.001
					(0.003)	(0.001)	(0.001)
Political stability							800.414
							(6.221)
Country fixed effect	NO	NO	NO	NO	NO	YES	YES
Observations	500	500	500	500	500	500	500
Log likelihood	− 2045	− 1934	− 1923	− 1884	− 1615	− 765	− 744

Regression result for stepwise modeling of a "messy" case. DV is FDI inflows. Models 1 through 5 are OLS, models 6 and 7 are fixed effect hierarchical models. The underlying simulated relationship is presented above. The registration number is irrelevant in the true model. Standard errors are in parentheses.

Table 6 VIFs for the "messy" illustration

	(1)	(2)	(3)	(4)	(5)
Customers	1.005	1.311	1.809	2.673	2.691
Age	1.005	1.306	1.729	3.038	3.041
Slack		1.011	1.484	1.820	1.823
CEO experience			1.011	1.797	1.797
Intangible assets				1.011	1.012
Subsidiary registration number					1.013

control for subsidiary slack, which has a negative effect on FDI inflows ($p < 0.0000024$). Yet, when considering CEO experience, and hence all relevant independent variables except country-level variables (model 4), it turns out that there is in fact a significant positive relationship between the number of customers and the amount of FDI a subsidiary receives ($p < 10^{-8}$), as in the "true" model. At the same time, it appears from model 4 that the coefficient of subsidiary age was biased upwards in models 2 and 3, and that CEO experience is another significant predictor of FDI received by a subsidiary ($p < 10^{-15}$). A researcher concerned with multicollinearity might well be troubled by increasing collinearity, as shown by the increase in VIFs from model 1 to model 4 (Table 6). As mentioned above, proposed cutoffs for VIFs (from 20 to 3) do not help in resolving the underlying econometric concerns. Yet, as we have argued and illustrated, collinearity itself is not an econometric problem in the sense that it would violate assumptions necessary for regression models to work. Consequently, a researcher would be well advised to include all relevant variables irrespective of their collinearity or VIFs, so as to avoid misspecification and consequently bias (see the Table 3 discussion).

The addition of a country intercept ($\gamma_{0,j}$) represents – at first sight – only a slight change to the econometric specification. However, it ensures the non-violation of the assumption of independently distributed error terms. Violation of this assumption is not a major problem if one is only interested in firm-level variables, and if these are independent of each other. Yet, if the researcher adds a country-specific variable (such as national culture or a variable capturing national institutions), as is often the case in IB research, this misspecification may become substantial and may lead to spurious support for hypotheses related to these variables. Models 5–7 in Table 5 show how this may come about.

In model 5, the researcher investigates whether there might be some country-level determinant of FDI inflows into subsidiaries. By adding the variable "subsidiary registration number" (which the researcher might somehow believe is relevant to FDI inflows), the OLS regression shown in model 5 of Table 5 suggests that the higher the subsidiary's registration number, the less FDI it receives ($p < 10^{-15}$). This effect seems to be statistically highly significant. However, model 6, which is a hierarchical model that accounts for country clusters in the error terms, shows that this result is spurious, and that it disappears if one uses country-level fixed effects ($p < 0.06$). If one accounts for the country-level variable "political stability" (model 7), the spurious effect of "subsidiary registration number" is even less significant ($p < 0.11$).

Some researchers argue that introducing fixed effects in a regression will eliminate the influence of all variables that work at that level. For example, Delmar and Shane (2003, p. 1171) argue in the realm of ventures:

> […] fixed effects regression partials out the effect of venture-level factors, such as the quality of the venture opportunity, and allows for an unbiased estimate of the relationship between business planning and the outcome under investigation.

In their case, the cluster-level variable is ventures. While it is true that the estimate obtained using country-fixed effects is unbiased by country-level clusters (Griliches, 1986), this does not mean that fixed effects eliminate all influence of variables at that level. In model 7 in Table 5, we see that political stability does show up as highly significant ($p < 10^{-16}$), and with a coefficient that is only insignificantly different from the "true" coefficient of 800, which represents the relationship between political stability and FDI inflows presented, where we define the "true" relationship between FDI inflows and its above determinants. While country-level fixed effects ensure that firm-level variables are unbiased by the omission of country-level variables and eliminate spurious results of country-level variables, they do not eliminate the possibility of finding statistical support for effects driven by country-level independent variables.

In conclusion, this "messy" example shows two things. First, omitting highly collinear variables introduces bias into regression result. Because this bias may lead to spurious findings or spurious non-findings (as illustrated in greater detail in the previous section), researchers that omit collinear variables risk that all findings from a regression analysis lose their validity. In addition, one must bear in mind that clusters in the data may also lead to spuriously significant results if clustering is not treated correctly in econometric specifications.

FINDINGS AND RECOMMENDATIONS

We have illustrated how regression results change when we consider correlated variables in a regression. More specifically, we have investigated how correlation affects error inflation; what omitting correlated variables does to regression results; and how these results are different in scenarios where the omitted variable is relevant for explaining variation in the dependent variable, as well as in scenarios where it is not. We have investigated both the effects of omissions on standard errors in regressions, and on the point estimates (the coefficients).

In the "messy" illustration, we extended the isolated analysis on multicollinearity to correspond to a more realistic scenario that an IB researcher may be confronted with: We used several variables, which are to different degrees correlated, and we introduced clustering in the data structure. In such a case, several econometric problems co-exist, and the applied researcher must take care of several potential biases at the same time. In order to facilitate the correct treatment of the different effects, the following insights need to be taken on board:

1. *Multicollinearity does not introduce bias*. Collinearity itself is not an econometric problem in the sense that it would violate the assumptions necessary for regression models to work. If one variable that is highly collinear with another one, or with several other ones, is added to a regression model, its collinearity does not affect regression coefficients. In all but extreme cases of collinearity (e.g., 0.8 and above, see Allison, 1999, p. 64) researchers are advised to add, rather than omit, potentially relevant collinear variables.

2. *Significant results under high variance inflation underestimate, rather than overestimate, statistical significance*. Variance inflation factors are indicators of standard errors that are too large, not too small. In a regression model, VIFs represent the degree to which regression coefficient variance is too large, i.e., larger than in a model with only independent variables with zero partial correlation. Consequently, significant results under high variance inflation may be taken to be conservative.
3. *Coefficient instability is not a consequence of multicollinearity*. If coefficients change when a potentially collinear variable is added to a regression model, it indicates that the model was misspecified prior to the collinear variable being added. As over-specification only inflates error terms while on the other hand under-specification introduces bias to coefficients and error terms, we strongly advise researchers who are uncertain about a variable to add it rather than to drop it.
4. *The higher the partial correlation between the variables, the more problematic it is to omit one*. If two variables are perfectly independent, omitting one will not affect the results of a regression model. As soon as there is some correlation between an omitted and an included variable, however, coefficients will be biased and error terms may be too low, potentially leading to biased and spuriously significant results.
5. *Ignoring clusters in data may lead to spurious results*. While hierarchical models definitely have considerable potential, not accounting for clusters in the data may lead researchers to conclude that relationships between variables exist when in reality they do not. Whether one is interested in cross-level relationships or not, it is crucial to account for clusters in the data.
6. *Accounting for country clusters does not pick up all country-level variation*. If a researcher uses hierarchical modeling, country fixed effects, or random intercept models, the effects of country-level independent variables on e.g., firm-level dependent variables can still be detected. While using different intercepts across countries avoids finding spurious results from country-level variables, it does not eliminate the potential influence of all country-level explanatory variables. This is the case especially for time-variant country-level variables.

CONCLUSION

International Business scholars engaged in empirical research usually cannot conduct controlled experiments. Consequently, they have difficulty obtaining ex ante independent observations, and hence face the challenge of correlated variables. While multicollinearity is well analyzed in the econometrics literature, illustrations are lacking of how it might affect research results in the kind of "messy" empirical contexts with which IB researchers must often contend. We do provide such illustrations. As the underlying econometric theory is readily available from econometrics texts, we do not repeat the derivations here. However, the treatment of collinearity has continued to be a sensitive issue in empirical IB research. In a large number of papers published in JIBS, authors explicitly compute variance inflation factors (VIFs), and use these as a means of choosing which variables to include in regression models. In at least 10% of these papers, the authors explicitly state that variables were dropped from models as a consequence of high VIFs.

Our aim was to illustrate how multicollinearity affects error terms in regression models so that IB researchers can effectively specify empirical models and interpret regression results computed for collinear variables. Although we recognize that high multicollinearity makes it difficult to interpret two coefficients that were specified as being independent, we suggest that it is possible to do so by engaging in cautious interpretation. Thus, high collinearity, either expressed by pairwise correlation or high VIFs, does not justify omitting relevant variables from a regression, and we show that the consequences of doing this can be biased coefficients and spurious results.

Our paper contributes to highlighting misconceptions about collinearity that cause problems in some IB research. We have shown that collinearity does not introduce bias and does not cause instability of results. The example we gave of a "messy" case illustrates both the kind of results researchers can expect to see, if empirical models are under-specified, and how these results are likely to change under different specifications. The simulation approach we adopted may help researchers connect findings from econometric theory with their own quantitative research. Our approach does have some limitations. First, we did not thoroughly address other specifications and related econometric problems – hierarchical data structures for instance. Second, we did not deal with the problem of highly impactful outliers in models suffering from high multicollinearity, beyond suggesting the obvious solution of collecting more data or eliminating these outliers in the first place. Third, we did not extend the discussion beyond regression analysis. All three issues deserve further treatment.

REFERENCES

Alcácer, J., Chung, W., Hawk, A., & Pacheco-de-Almeida, G. 2018. Applying random coefficient models to strategy research: Identifying and exploring firm heterogeneous effects. *Strategy Science,* 3(3): 533–553.

Allison, P. D. 1999. *Multiple regression: A primer*. Newbury Park: Pine Forge Press.

Asmussen, C. G., Pedersen, T., & Dhanaraj, C. 2009. Host-country environment and subsidiary competence: Extending the diamond network model. *Journal of International Business Studies,* 40(1): 42–57.

Cohen, P., West, S. G., & Aiken, L. S. 2014. *Applied multiple regression/correlation analysis for the behavioral sciences*. London: Psychology Press.

Cuervo-Cazurra, A., Andersson, U., Brannen, M. Y., Nielsen, B. B., & Reuber, A. R. 2016. *From the editors: Can I trust your findings? Ruling out alternative explanations in international business research*. Berlin: Springer.

Delmar, F., & Shane, S. 2003. Does business planning facilitate the development of new ventures? *Strategic Management Journal,* 24(12): 1165–1185.

Driffield, N., & Munday, M. 2000. Industrial performance, agglomeration, and foreign manufacturing investment in the UK. *Journal of International Business Studies,* 31(1): 21–37.

Enright, M. J. 2009. The location of activities of manufacturing multinationals in the Asia-Pacific. *Journal of International Business Studies,* 40(5): 818–839.

Goldberger, A. S. 1991. *A course in econometrics*. Cambridge: Harvard University Press.

Greene, W. H. 2003. *Econometric analysis*. New Delhi: Pearson Education India.

Griffiths, W. E., Judge, G. G., Hill, R. C., Lütkepohl, H., & Lee, T.-C. 1985. *The theory and practice of econometrics*. New York: Wiley.

Griliches, Z. 1986. Economic data issues. In Z. Griliches & M. D. Intriligator (Eds), *Handbook of econometrics*, Vol. 3: 1465–1514. Amsterdam: North-Holland.

Habib, M., & Zurawicki, L. 2002. Corruption and foreign direct investment. *Journal of International Business Studies,* 33(2): 291–307.

Hox, J. J. 1995. *Applied multilevel analysis*. Amsterdam: TT-publikaties.

Kalnins, A. 2018. Multicollinearity: How common factors cause type 1 errors in multivariate regression. *Strategic Management Journal,* 39(8): 2362–2385.

Meyer, K. E., & Sinani, E. 2009. When and where does foreign direct investment generate positive spillovers? A meta-analysis. *Journal of International Business Studies,* 40(7): 1075–1094.

Muethel, M., & Bond, M. H. 2013. National context and individual employees' trust of the out-group: The role of societal trust. *Journal of International Business Studies,* 44(4): 312–333.

Nielsen, B. B., & Raswant, A. 2018. The selection, use, and reporting of control variables in international business research: A review and recommendations. *Journal of World Business,* 53: 958–968.

Read, D., & Read, N. L. 2004. Time discounting over the lifespan. *Organizational Behavior and Human Decision Processes,* 94(1): 22–32.

Reuber, A. R., & Fischer, E. 1997. The influence of the management team's international experience on the internationalization behaviors of SMEs. *Journal of International Business Studies,* 28(4): 807–825.

Rogerson, P. 2001. *Statistical methods for geography*. Thousand Oaks: Sage.

Rugman, A. M., & Verbeke, A. 2001. Subsidiary-specific advantages in multinational enterprises. *Strategic Management Journal,* 22(3): 237–250.

Steenbergen, M. R. 2012. Hierarchical linear models for electoral research: A worked example in Stata. ELECDEM. www.elecdem.eu/media/universityofexeter/elecdem/pdfs/istanbulwkspjan2012/Hierarchical_Linear_Models_for_Electoral_Research_A_worked_example_in_Stata.pdf. Accessed February 15, 2014.

Verbeke, A., Amin Zargarzadeh, M., & Osiyevskyy, O. 2014. Internalization theory, entrepreneurship and international new ventures. *Multinational Business Review,* 22(3): 246–269.

Verbeke, A., & Yuan, W. 2013. The drivers of multinational enterprise subsidiary entrepreneurship in China: A new resource-based view perspective. *Journal of Management Studies,* 50(2): 236–258.

Wang, C., Hong, J., Kafouros, M., & Wright, M. 2012. Exploring the role of government involvement in outward FDI from emerging economies. *Journal of International Business Studies,* 43(7): 655–676.

Wooldridge, J. M. 2014. *Introduction to econometrics: Europe, Middle East and Africa Edition*. Boston: Cengage Learning.

Zhao, M., Park, S. H., & Zhou, N. 2014. MNC strategy and social adaptation in emerging markets. *Journal of International Business Studies,* 45(7): 842–861.

ABOUT THE AUTHORS

Thomas Lindner is an Assistant Professor at the Institute for International Business at WU Vienna. His research covers the influence of the external environment on firm internationalization, international corporate finance, and cross-national distance. Thomas' research predominantly uses quantitative methodologies.

Jonas Puck is Full Professor of International Business at WU Vienna University of Economics and Business. His current research interests lie in the overlaps of global strategy, finance, and political science. Jonas serves on the Editorial Boards of the *Journal of International Business Studies*, *Long Range Planning*, *European Management Journal*, *Journal of World Business*, and the *Multinational Business Review*. His research has been published in journals

such as the *Journal of International Business Studies*, *Journal of Management Studies*, *Global Strategy Journal*, and *Journal of World Business*, among others .

Dr. Alain Verbeke is a Professor of International Business Strategy and holds the McCaig Research Chair in Management at the Haskayne School of Business, University of Calgary. He was elected as the Inaugural Alan M. Rugman Memorial Fellow at the Henley Business School, University of Reading, and is Editor-in-Chief of the Journal of International Business Studies. He is also an Adjunct Professor at the Solvay Business School, Vrije Universiteit Brussel. He has published several books and numerous refereed articles, including many pieces in leading scholarly journals such as *JIBS*, *JMS* and *SMJ*. His research agenda consists of rethinking and augmenting the core paradigms in strategic management and international business, especially internalization theory, which is a joint transaction cost economics and resource-based view of the firm, focused on the governance of new resource combinations. He has particular expertise in managing headquarters – subsidiary relationships and broader governance challenges in large multinational enterprises.