APPENDIX[1]

The multinationality-performance relationship

Table A-1 shows the 10 most influential contributions—that is the articles that to date have received the highest number of citations in SSCI—focused on the M-P relationship.[2]

Measures

For our independent variable, we follow L&B and B&K's works and create the *Internationalization* index:

$$Internationalization_i = \frac{1}{2}\left(\frac{NCountries_i}{\max(NCountries)} + \frac{NSubsidiaries_i}{\max(NSubsidiaries)}\right)$$

The internationalization of a firm *i* in each year thus depends on the number of countries (NCountries) and the number of foreign subsidiaries (NSubsidiaries). The first part of the equation is normalized by the maximum number of countries in which any firm has subsidiaries. The second part is normalized by the maximum number of subsidiaries of any firm. In order to test the S-shaped relationship, we also create the squared and cubed term of *Internationalization* and name them *Internationalization squared* and *Internationalization cubed* respectively. The mean number of countries in which a multinational firm operates slightly differs across samples: 4.82 for our sample, 5.57 and 8.11 for L&B and B&K's samples respectively. Similarly, the average number of subsidiaries for multinationals slightly differs as well: 18.25 for our sample, 11.88 and 19.23 in L&B and B&K's samples respectively. The dependent variable, return on assets (*ROA*), is calculated using the ratio of a firm's operating profit to its total assets. This differs slightly from the definition in B&K and L&B's studies, in which ROA is calculated as the ratio of net income (obtained by subtracting taxes and interests from the operating profit). We expect this not to affect our results as this is only a small variation in the way of calculating ROA and we consistently use such measure for all firms in our sample.

The main difference between our analysis and those of L&B and B&K is that we do not include their proxy of a firm's intangible assets (named "R&D Intensity" in their specification) in our model. While they include it to test the moderating effect of intangible assets on the M-P relationship, we restrict our focus on the main effect of *Internationalization* on *ROA*. We also exclude it as control due to the relatively low availability of information on this measure in the Orbis database. We do so also because the coefficient of "Parent R&D intensity" in B&K's analysis is not significant in any of their specifications and the coefficients of their independent variables are barely affected by the inclusion of the interaction term between "Parent R&D Intensity" and *Internationalization*. Despite this, we perform additional analyses (reported in the robustness checks below) to examine whether the exclusion of intangible assets may impact our results. The results obtained in these additional tests allow us to conclude that the exclusion of intangible assets does not raise concerns for our study.

---

[1] This is the online appendix to the article by Pisani, Garcia-Bernardo, and Heemskerk published in the *Strategic Management Journal* and titled "Does it pay to be a multinational? A large-sample, cross-national replication assessing the multinationality-performance relationship."

[2] Verbeke and Forootan (2012) report in their Table 1 the list of 12 most cited articles that empirically investigated the M-P relationship. All 10 articles reported in our Table A-1 were already included in their list.

TABLE A-1

The 10 most influential studies on the M-P relationship[1]

Citation ranking	No. citations in Web of Science	Year	Authors	Journal	M-P relationship hypothesized	M-P relationship empirically shown	Source of data	Sample size	Country of sampled firms	Cross-section versus panel	Pooled cross-section, fixed or random effects
1	1,045	1997	Hitt, Hoskisson, and Kim	*Academy of Management Journal*	Inverted U-shaped	Inverted U-shaped	S&P COMPUSTAT database	295 firms	USA	Cross-section	Not applicable
2	1,028	2000	Zahra, Ireland, and Hitt	*Academy of Management Journal*	Positive	Positive	Survey administered to firms	1,388 firms	USA	Cross-section	Not applicable
3	680	2001	Lu and Beamish	*Strategic Management Journal*	U-shaped	U-shaped	NIKKEI NEEDS tapes database and Japan Company Handbook	164 firms	Japan	Panel (12 years) - Total n. of observations not reported	Pooled cross-section/time-series
4	497	1996	Tallman and Li	*Academy of Management Journal*	Positive	Insignificant	Directory of Multinationals database	192 firms	USA	Cross-section	Not applicable
5	462	2004	Lu and Beamish	*Academy of Management Journal*	S-shaped	S-shaped	NIKKEI NEEDS tapes database and Japan Company Handbook	1,489 firms	Japan	Panel (12 years) - Total n. of observations not reported	Random effects
6	390	2003	Contractor, Kundu, and Hsu	*Journal of International Business Studies*	S-shaped	S-shaped	Directory of the World's Largest Service Companies	103 firms	USA (42% of the sample) plus 9 other undisclosed countries (remaining 58% of the sample)	Panel (6 years) - Total n. of observations=364	Pooled cross-section/time-series
7	354	2000	Palich, Cardinal, and Miller	*Strategic Management Journal*	Inverted U-shaped	Inverted U-shaped	Meta-analyis of 82 studies	Not applicable	Not applicable	Not applicable	Not applicable
8	339	1989	Geringer, Beamish, and Dacosta	*Strategic Management Journal*	Positive	Positive	World Directory of Multinational Enterprises	200 firms	USA (50% of the sample) and Europe (remaining 50%)	Cross-section (Tests of mean differences)	Not applicable
9	289	1999	Gomes and Ramaswamy	*Journal of International Business Studies*	Inverted U-shaped	Inverted U-shaped	Not reported	95 firms	USA	Panel (6 years) - Total n. of observations=570	Pooled cross-section/time-series
10	279	1988	Grant, Jammine, and Thomas	*Academy of Management Journal*	Positive	Positive	Not reported	304 firms	UK	Panel (13 years_	Pooled cross-section

[1] The No. of citations in Web of Science were retrieved on December 5 2017

Another important difference vis-à-vis L&B and B&K's models is that we exclude the variable exchange rate when using our full sample (while we regularly include it in the analysis focused on the subsample of multinational firms provided in the article). We omit this control variable given that reliable information on exchange rates is retrievable for only 71 of the 111 countries considered in our full sample. Thus, its inclusion would imply the reduction of our full sample by approximately 50,000 observations. Having said that, as a robustness check (reported below) we rerun both our random- and fixed-effects models on the subsample for which it is possible to take into account the exchange rate (846,482 firm-year observations) and the results are entirely aligned with the ones obtained excluding this variable.

In addition to our dependent and independent variables, we include the same controls used by B&K. Thus, we control for a firm's size, its debt-to-equity ratio, its export intensity as well as its extent of product diversification. *Parent size* is measured (as logarithm) using the total turnover of the firm. The variable *Parent export intensity* is calculated as the fraction of revenues allocated to subsidiaries outside the parent country. *Product diversification* is calculated using the Berry-Herfindahl index (1 - Herfindahl index: $1 - \sum_{s \in sector} p_s^2$), where p_s is the fraction of revenues produced in subsidiaries of sector *s,* and sectors are aggregated at the NACE Rev. 2 highest level (letter-level).[3] In addition, we include both industry-fixed effects (at the highest level) and country-fixed effects in the random-effects estimation models, as done by L&B and B&K. As B&K we are unable to control for firm's advertising intensity and thus also exclude it from our analysis. As done by both L&B and B&K, we lag all independent variables by one year.

Robustness checks of the replication using our full sample

Having compared our results with the ones obtained by L&B and B&K in the article (see Table 4), in what follows we look at the robustness of these findings by undertaking a number of additional checks. Thus, in Table A-2 we check the robustness of our findings after dropping each control from the final models (Model 6 in Table 4 for the fixed-effects model estimation and Model 3 in Table A-4 for the random-effects model estimation) as both L&B and B&K do. The S-shaped relationship remains significant across all models, with the exception of the random-effects Model 3, in which the coefficient of the main term of *Internationalization* is not significantly different from zero. We continue with testing the robustness of our findings when using different lags (Table A-3). The S-shaped relationship remains significant for all random-effects estimations, and for all fixed-effects estimation models except for the regression with a three-year lag in the *ROA* (Model 2). Table A-4 shows comparable results for the main and squared models compared with the ones obtained by B&K (as also discussed in the main text focusing on the fixed-effects estimation shown in Table 4), a negative effect of *Internationalization* for the linear model, and a U-shaped relationship for the quadratic model (the models using fixed-effects and testing for the linear, quadratic, and cubic relationships are reported in the article in Table 4 as Models 4, 5, and 6 respectively). Models 4 and 5 in Table A-4 use the mean-centered variables to account for collinearity (in the same vein as L&B and B&K do) and show the robustness of the S-shaped relationship to these additional checks.

[3] For additional information on Eurostat Statistical Classification of Economic Activities refer to: http://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE_REV2

TABLE A-2

Robustness to dropped controls[1]

Model	1 (RE)	2 (FE)	3 (RE)	4 (FE)	5 (RE)	6 (FE)	7 (RE)	8 (FE)
Relationship tested	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic
Dependent variable	ROA	ROA	ROA	ROA	ROA	ROA	ROA	ROA
Internationalization	-0.592	-0.320	0.004	-0.150	-0.687	-0.330	-0.729	-0.335
	(0.052)	(0.077)	(0.047)	(0.076)	(0.050)	(0.075)	(0.052)	(0.076)
	[<0.0001]	[0.00003]	[0.925]	[0.047]	[<0.0001]	[0.00002]	[<0.0001]	[0.00001]
Internationalization squared	4.577	1.889	1.683	1.127	5.298	1.972	5.429	1.999
	(0.477)	(0.533)	(0.395)	(0.519)	(0.497)	(0.527)	(0.507)	(0.528)
	[<0.0001]	[0.0004]	[0.00003]	[0.030]	[<0.0001]	[0.0002]	[<0.0001]	[0.0002]
Internationalization cubed	-7.837	-2.877	-3.669	-1.767	-8.990	-2.934	-9.132	-2.980
	(1.054)	(0.900)	(0.800)	(0.866)	(1.140)	(0.899)	(1.154)	(0.899)
	[<0.0001]	[0.002]	[0.00001]	[0.042]	[<0.0001]	[0.002]	[<0.0001]	[0.001]
Parent size			-0.012	-0.010	-0.012	-0.010	-0.012	-0.010
			(0.0001)	(0.0002)	(0.0001)	(0.0002)	(0.0001)	(0.0002)
			[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]	[<0.000]
Gearing	0.014	0.012			0.015	0.011	0.014	0.011
	(0.0003)	(0.001)			(0.0003)	(0.001)	(0.0003)	(0.001)
	[<0.0001]	[<0.0001]			[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]
Parent export intensity	-0.003	-0.003	-0.007	-0.001			-0.005	-0.004
	(0.002)	(0.003)	(0.002)	(0.003)			(0.002)	(0.003)
	[0.122]	[0.353]	[0.00001]	[0.679]			[0.004]	[0.187]
Parent product diversification	-0.023	-0.006	-0.017	-0.003	-0.023	-0.007		
	(0.001)	(0.002)	(0.001)	(0.002)	(0.001)	(0.002)		
	[<0.0001]	[0.001]	[<0.0001]	[0.144]	[<0.0001]	[0.001]		
Number of observations	889,865	195,205	889,865	195,205	889,865	195,205	889,865	195,205
F statistic	196.86	98.51	236.88	460.43	263.56	557.35	260.14	555.07
R square	0.03	0.001	0.04	0.004	0.04	0.01	0.04	0.01
Fixed or random effects	Random	Fixed	Random	Fixed	Random	Fixed	Random	Fixed

[1] All independent variables are lagged by one period. All models report (robust standard errors in parentheses) and [p-values in brackets].

TABLE A-3

Robustness to other lags[1]

Model	1 (Lag 3, RE)	2 (Lag 3, FE)	3 (Lag 2, RE)	4 (Lag 2, FE)	5 (No lag, RE)	6 (No lag, FE)
Relationship tested	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic
Dependent variable	ROA	ROA	ROA	ROA	ROA	ROA
Internationalization	-0.319	0.033	-0.491	-0.243	-1.114	-1.050
	(0.058)	(0.099)	(0.053)	(0.087)	(0.052)	(0.069)
	[<0.0001]	[0.743]	[<0.0001]	[0.006]	[<0.0001]	[<0.0001]
Internationalization squared	3.004	-0.106	4.032	1.551	7.357	5.127
	(0.507)	(0.744)	(0.460)	(0.615)	(0.539)	(0.542)
	[<0.0001]	[0.887]	[<0.0001]	[0.012]	[<0.0001]	[<0.0001]
Internationalization cubed	-5.798	-0.569	-6.980	-2.393	-11.627	-7.062
	(1.102)	(1.460)	(0.964)	(1.025)	(1.248)	(1.023)
	[<0.0001]	[0.697]	[<0.0001]	[0.020]	[<0.0001]	[<0.0001]
Parent size	0.012	0.002	0.013	0.005	0.022	0.038
	(0.0004)	(0.001)	(0.0003)	(0.001)	(0.0003)	(0.001)
	[<0.0001]	[0.044]	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]
Parent debt-to-equity ratio	-0.010	-0.001	-0.011	-0.004	-0.015	-0.018
	(0.0002)	(0.0003)	(0.0001)	(0.0003)	(0.0001)	(0.0002)
	[<0.0001]	[0.002]	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]
Parent export intensity	-0.004	0.001	-0.002	0.005	-0.007	-0.016
	(0.002)	(0.004)	(0.002)	(0.003)	(0.002)	(0.003)
	[0.066]	[0.795]	[0.288]	[0.145]	[0.00001]	[<0.0001]
Parent product diversification	-0.021	-0.001	-0.021	-0.002	-0.023	-0.011
	(0.001)	(0.003)	(0.001)	(0.002)	(0.001)	(0.002)
	[<0.0001]	[0.625]	[<0.0001]	[0.400]	[<0.0001]	[<0.0001]
Number of observations	492,858	108,115	671,370	147,274	1,195,723	262,299
Lag	3 years	3 years	2 years	2 years	0 years	0 years
F statistic	137.68	2.97	187.38	55.11	435.98	3,275.16
R square	0.04	0.0001	0.04	0.001	0.05	0.03
Fixed or random effects	Random	Fixed	Random	Fixed	Random	Fixed

[1] All independent variables are lagged as specified in the lag row. All models report (robust standard errors in parentheses) and [p-values in brackets].

TABLE A-4

Linear, quadratic, cubic (random-effects) and mean-centered models[1,2]

Model	1 (RE)	2 (RE)	3 (RE)	4 (Mean-centered, RE)	5 (Mean-centered, FE)
Relationship tested	Linear	Quadratic	Cubic	Cubic	Cubic
Dependent variable	ROA	ROA	ROA	ROA	ROA
Internationalization	-0.104	-0.308	-0.651	-0.656	-0.359
	(0.017)	(0.033)	(0.052)	(0.052)	(0.076)
	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]	[0.00001]
Internationalization squared		1.034	5.081	5.149	2.208
		(0.140)	(0.499)	(0.504)	(0.533)
		[<0.0001]	[<0.0001]	[<0.0001]	[0.00004]
Internationalization cubed			-8.659	-8.785	-3.322
			(1.129)	(1.143)	(0.915)
			[<0.0001]	[<0.0001]	[0.0003]
Parent size	0.014	0.015	0.015	0.015	0.011
	(0.0003)	(0.0003)	(0.0003)	(0.0003)	(0.001)
	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]
Parent debt-to-equity ratio	-0.012	-0.012	-0.012	-0.012	-0.010
	(0.0001)	(0.0001)	(0.0001)	(0.0001)	(0.0002)
	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]
Parent export intensity	-0.007	-0.005	-0.003	-0.003	-0.003
	(0.002)	(0.002)	(0.002)	(0.002)	(0.003)
	[0.00001]	[0.002]	[0.047]	[0.043]	[0.313]
Parent product diversification	-0.023	-0.023	-0.023	-0.023	-0.006
	(0.001)	(0.001)	(0.001)	(0.001)	(0.002)
	[<0.0001]	[<0.0001]	[0.000]	[<0.0001]	[<0.0001]
Number of observations	889,865	889,865	889,865	878,499	195,205
F statistic	264.51	263.02	261.80	259.21	478.34
R square	0.04	0.04	0.04	0.04	0.01
Fixed or random effects	Random	Random	Random	Random	Fixed

[1] All independent variables are lagged by one period. All models report (robust standard errors in parentheses) and [p-values in brackets].
[2] The models using fixed-effects and testing for the linear, quadratic, and cubic relationships are reported in the article in Table 4 as Models 4, 5, and 6 respectively.

In addition to the robustness checks undertaken by L&B and B&K, we perform additional tests to verify the robustness of our findings. First, we look at the smaller subsample with available information on intangibles (Models 1 and 2 in Table A-5). For this considerably smaller subset of firm-year data points (31,845) we create a variable named *Intangibles* which corresponds to the ratio of intangible assets to total assets and include it as control. Our findings show that the coefficient of *Intangibles* is not significant in the fixed-effects model (p-value = 0.226) and the main results obtained in Model 6 in Table 4 are confirmed.

Next, we rerun our fully-specified model on the subsample for which it is possible to account for the exchange rate. Information on real effective exchange rates is available from the International Monetary Fund for 71 of the 111 countries considered in our study.[4] The resulting subsample after including the variable *Exchange rate* as control variable comprises 846,482 firm-year observations. The S-curved relationship between multinationality and performance is confirmed in both the random- and fixed-effects estimations (Models 3 and 4 in Table A-5).

After looking at the effect of *Intangibles* and *Exchange rate*, we perform further analyses building on earlier works supporting the adoption of hybrid or multilevel models (Bowen, 2007; Certo *et al.*, 2017; Nielsen and Nielsen, 2010).[5] Thus, we explicitly model the between- and within-firm effects separately (Model 5 in Table A-5) using the Mundlak (1978) correction, in particular the within-between formulation of the model (Bell and Jones, 2014). In this formulation, the heterogeneity bias is explicitly modeled by adding the higher-level mean at the group level of each time-varying covariate, accounting for the between effect. This approach is similar to the hybrid approach introduced by Allison (2005) and explained in Certo *et al.*'s work (2017). As expected—in view of the fact that we find similar results for our random- and fixed-effects estimations—we obtain a significant S-shaped relationship for both the within and between parts of the model. To account for the nested nature of our data, we also repeat our estimation using a multilevel model (Model 6 in Table A-5) in which the lower level is the firm-level while the higher one is the country-level (including industry fixed-effects).[6] The results are entirely aligned with the ones obtained in our random-effects model. When using as higher-level the industry instead of the country (and thus including country fixed-effects in the model; the results of this additional model specification are available upon request), we obtain again very similar results in which the S-curve continues to be empirically validated.

---

[4] For additional information on the data we retrieved to construct our *Exchange rate* variable, refer to: http://data.imf.org/?sk=4C514D48-B6BA-49ED-8AB9-52B0C1A0179B.

[5] While having some advantages vis-à-vis fixed-effects models, both hybrid and multilevel models present the same methodological challenge that remains to be addressed—i.e., how to account for endogeneity when it can exist with both within- and between-firm variables (Certo *et al.*, 2017). To our knowledge, no published study in the field has in fact used multilevel or hybrid models to examine a causal between-firm relationship using longitudinal (nested) data while also accounting for the endogeneity of the main predictor variable (whose source can be at different levels). To ensure this we reviewed all papers that use multilevel models published by the *Strategic Management Journal*. None of them discusses (and empirically accounts for) the potential endogeneity of the main predictor variable as we point out here.

[6] Despite their several advantages vis-à-vis alternative model estimations, multilevel models have not been frequently used in the fields of strategic management and international business (Certo *et al.*, 2017; Peterson, Arregle, and Martin, 2012). Multilevel models have the benefit to allow for the estimation of the effects of both time-varying and time-invariant variables at different levels of analysis. However, they do not allow to test for differences between within- and between-firm effects as they are embodied in the same coefficient. As also noted by Certo *et al.* (2017), multilevel and random-effects models report essentially the same results.

TABLE A-5

Robustness results to further model specifications[1,2,3]

Model	1 (Including intangibles, RE)	2 (Including intangibles, FE)	3 (Including exchange rate, RE)	4 (Including exchange rate, FE)	5 (Mundlak correction)	6 (Multilevel model)	7 (Heckman second-stage)	8 (Heckman second-stage including prior performance)
Relationship tested	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic	Cubic
Dependent variable	ROA	ROA	ROA	ROA	ROA	ROA	ROA	ROA
Internationalization	-0.159	-0.300	-0.639	-0.213	-0.914; -0.359	-0.652	-0.317	-0.190
	(0.114)	(0.171)	(0.054)	(0.080)	(0.077); (0.076)	(0.328)	(0.044)	(0.039)
	[0.162]	[0.080]	[<0.0001]	[0.008]	[<0.0001]; [0.00001]	[0.047]	[<0.0001]	[<0.0001]
Internationalization squared	0.964	2.639	5.058	1.380	7.996; 2.208	5.067	2.487	1.324
	(0.943)	(1.365)	(0.510)	(0.542)	(0.867); (0.533)	(1.66)	(0.380)	(0.333)
	[0.307]	[0.054]	[<0.0001]	[0.011]	[<0.000]; [0.00004]	[0.001]	[<0.0001]	[<0.0001]
Internationalization cubed	-2.637	-5.933	-8.627	-2.127	-14.686; -3.322	-8.609	-4.449	-2.272
	(2.092)	(2.717)	(1.141)	(0.901)	(2.193); (0.915)	(2.573)	(0.773)	(0.668)
	[0.208]	[0.029]	[<0.0001]	[0.019]	[<0.0001]; [0.0003]	[0.001]	[<0.0001]	[0.001]
Parent size	0.059	0.035	0.014	0.010	0.016; 0.011	0.014	0.018	0.011
	(0.002)	(0.005)	(0.0003)	(0.001)	(0.0003); (0.001)	(0.002)	(0.0004)	(0.0004)
	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]; [<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]
Parent debt-to-equity ratio	-0.016	-0.013	-0.012	-0.010	-0.014; -0.010	-0.012	-0.014	-0.008
	(0.001)	(0.002)	(0.0001)	(0.0002)	(0.0001); (0.0002)	(0.002)	(0.0002)	(0.0002)
	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]; [<0.0001]	[<0.0001]	[<0.0001]	[<0.0001]
Parent export intensity	0.022	-0.007	-0.005	-0.003	-0.002; -0.003	-0.004	-0.006	0.003
	(0.005)	(0.009)	(0.002)	(0.003)	(0.002); (0.003)	(0.004)	(0.001)	(0.001)
	[<0.0001]	[0.405]	[0.009]	[0.342]	[0.407]; [0.313]	[0.405]	[<0.0001]	[0.014]
Parent product diversification	-0.013	-0.006	-0.023	-0.007	-0.035; -0.006	-0.02	-0.031	-0.015
	(0.006)	(0.007)	(0.001)	(0.002)	(0.002); (0.002)	(0.006)	(0.002)	(0.002)
	[0.025]	[0.414]	[<0.0001]	[0.0003]	[<0.0001]; [0.002]	[<0.0001]	[<0.0001]	[<0.0001]
Intangibles	-0.00004	-0.0001						
	(0.000)	(0.000)						
	[0.073]	[0.226]						
Exchange rate			0.0001	0.0003				
			(0.00003)	(0.00003)				
			[0.016]	[<0.0001]				
ROA(t-1)								0.473
								(0.002)
								[<0.0001]
IMR							0.001	0.0004
							(0.001)	(0.001)
							[0.018]	[0.399]
Number of observations	31,845	24,677	846,482	177,755	878,499	889,865	889,865	642,506
Wald Chi square / F statistic	25.86	36.61	371.31	477.94	250.28	521,391.37	13,944.47	61,441.20
R square	0.16	0.01	0.04	0.01	0.04			
Fixed or random effects	Random	Fixed	Random	Fixed	Both	Mixed		

[1] All independent variables are lagged by one period. Models 1-4 and 6 report (robust standard errors in parentheses). All models report [p-values in brackets]. In model 5 each cell shows both the between effect (first number) and the within effect (second number).
[2] Model 6 only contains industry dummies as the two levels considered are firm and country.
[3] In Models 7 and 8 we only report the second-stage results of the Heckman's two-stage estimation procedures. The first-stage results are available upon request.

We also perform additional analyses to account for the possibility that firms may self-select into the multinational (versus domestic) category due to unobserved firm characteristics and this may bias the results of our estimation. To take into account such potential selection bias in our analysis, we follow common procedure (Bascle, 2008) and use Heckman's (1979) two-stage estimation procedure. In the first stage, we estimate the selection equation using a pooled probit model in which the limited (dummy) dependent variable *MNE* is the propensity to be a multinational versus a domestic firm—i.e., *MNE* scores 1 if the firm is a multinational and 0 if the firm is domestic—and use the set of control variables included in our Model 6 in Table 4 as our independent variables. In this first stage, we also include our main exclusion variable corresponding to the export intensity aggregated at the industry and country levels (*Export intensity by industry, country, and year*). We expect that firms pertaining to a specific industry and country characterized by a higher average export intensity are more likely to internationalize, but do not expect that export intensity aggregated at the industry and country levels has an effect on the profitability of a given company, thus satisfying the exclusion restriction of a two-stage estimation procedure. This is confirmed in our data as the pairwise correlation coefficient between *Export intensity by industry, country, and year* and *MNE* is 0.16 while the one between *Export intensity by industry, country, and year* and *ROA* is 0.01. The first stage allows for the computation of the inverse Mills ratio (*IMR*), which represents our selection parameter that is included in the second stage to account for potential selection bias and thus obtain consistent and unbiased coefficients. In the second stage, *ROA* is modeled as a function of *IMR* and all our independent and control variables as included in Model 6 in Table 4. The second-stage results (reported in Model 7 in Table A-5) are entirely aligned with those obtained without correcting for this potential selection bias. In view of the fact that a firm's prior performance may also influence its likelihood to internationalize, we repeat the entire analysis adding the lagged profitability measure (i.e., *ROA* at time t-1) as control variable in both stages. The second-stage results (reported in Model 8 in Table A-5) are entirely aligned with those shown in Model 7. Thus, these additional analyses provide strong evidence that our findings do not suffer from this potential selection bias.

Finally, another issue is associated with the potential sample selection bias due to the fact that, while the entire Orbis dataset contains data on over three million firms with at least one subsidiary, it was obviously impossible to retrieve the firm-level data that is necessary to replicate L&B and B&K's studies for the vast majority of these firms, especially the information necessary to construct *Parent export intensity* and *Parent product diversification*. If missing values are non-random—i.e., unobservable variables have an effect on the availability of data—our findings may be biased as a result of sample selection bias. To address the potential sample selection bias in our analysis we follow standard procedure (Bascle, 2008) and again make use of the Heckman's (1979) two-stage estimation procedure. Thus, in the first stage we use approximately 1.5 million firm-year data points to model the propensity that a given observation forms part of our full working sample comprising 889,865 firm-year observations. The results of this additional check (available upon request) are entirely aligned with the ones shown in Models 7 and 8 in Table A-5, thus lending support to the conclusion that our findings do not suffer from potential sample selection bias.

TABLE A-6

First-stage regressions[1,2,3]

Model	1 (First-stage regression - Main term)	2 (First-stage regression - Squared term)	3 (First-stage regression - Cubed term)
Dependent variable	Internationalization	Internationalization squared	Internationalization cubed
Board internationalization	0.03	0.48	10.09
	(0.002)	(0.06)	(2.22)
	[<0.0001]	[<0.0001]	[<0.0001]
Board internationalization squared	-0.001	-0.01	-0.36
	(0.000)	(0.002)	(0.08)
	[<0.0001]	[<0.0001]	[<0.0001]
Board internationalization cubed	0.000006	0.0001	0.002
	(0.000)	(0.000)	(0.001)
	[<0.0001]	[<0.0001]	[<0.0001]
Parent size	0.38	5.06	108.72
	(0.04)	(1.21)	(44.86)
	[<0.0001]	[<0.0001]	[0.02]
Parent debt-to-equity ratio	0.03	0.66	23.07
	(0.01)	(0.48)	(21.99)
	[<0.0001]	[0.17]	[0.29]
Parent export intensity	0.46	4.35	34.85
	(0.07)	(2.27)	(82.98)
	[<0.0001]	[0.06]	[0.67]
Parent product diversification	0.64	9.67	261.27
	(0.08)	(2.83)	(114.06)
	[<0.0001]	[0.001]	[0.02]
Exchange rate	0.01	0.14	3.83
	(0.001)	(0.04)	(1.47)
	[<0.0001]	[<0.0001]	[0.01]
Number of observations	102,113	102,113	102,113
F statistic	54.71	17.38	6.73
Fixed or random effects	Fixed	Fixed	Fixed

[1] All models report (robust standard errors in parentheses) and [p-values in brackets].
[2] 2SLS fixed-effects model estimation is used for all models.
[3] As we use both *Internationalization* and *ROA* in percentage points in our 2SLS estimations (thus, e.g., *Internationalization* varies from 0 to 100), we also use *Board internationalization* in percentage points to simplify the interpretation of the coefficients of the first-stage regressions. *Board internationalization* thus ranges from 0 to 100, with a mean of 14.69 and a standard deviation of 25.82.

Additional analyses of the replication using our subsample of multinational firms

In what follows we report the additional analyses and results relative to the replication using a 2SLS fixed-effects model estimation on our subsample of multinational firms (made of 102,113 firm-year observations) whose main findings are shown in the article. First, we provide further information on how we built our instrument *Board internationalization*. The instrument was constructed from Orbis data. It is based on the analysis of the 535,597 unique directors appointed by the 32,835 multinational firms included in our subsample. For every year *y* and firm's board of directors, we consider the directors appointed before *y* and with resignation after *y*. Then, for each firm we calculate the fraction of directors with at least one other board executive position with another foreign firm (domiciled in another country), which corresponds to our measure *Board internationalization*. The construction of this measure thus involved the analysis of a very large number of boards' positions, i.e. more than 5.5 million, most of which related to firms not included in our (sub)sample. While we were able to gather data on all the 535,597 appointed directors and their international interlocks for all the 32,835 multinational firms, the information on the exact year of appointment and resignation within the time period considered was not available for all directors. When such information was not readily available we used additional information provided by Orbis (such as, e.g., the status of the director [resigned/current], the date when the record was last updated in Orbis, the number of confirmation dates) to establish whether in a given year a specific director was already/still appointed in the corresponding board of directors or not.[7] In Table A-6 we provide the first-stage regressions using *Board internationalization*, *Board internationalization squared*, and *Board internationalization cubed* as our three instruments. The second-stage corresponds to Model 7 in Table 5 reported in the article.

To gauge instrument strength in our empirical setting, we begin by noting that all three instruments have significant coefficients in all three first-stage equations and the first-stage F-statistics testing the hypothesis that the coefficients of the instruments equal 0 are always larger than 10 (133.72, 34.94, and 11.36 respectively), exceeding the rule of thumb cutoff for weak instruments proposed by Staiger and Stock (1997). While Staiger and Stock's (1997) rule and especially the more recently developed Stock and Yogo's (2005) test for weak instruments have been widely used—also in strategy research as corroborated by, e.g., Semadeni, Whiters, and Certo (2014) and Bascle (2008)—there is ongoing research in the econometrics field to develop better approaches to test for instrument strength. This is especially because Stock and Yogo's (2005) test relies heavily on the assumption of homoskedasticity, while data used in practice often violate this assumption. Building on these latest developments, we rely on the test developed by Montiel Olea and Pflueger (2013),[8] identified by Andrews, Stock, and Sun (2019) as currently the preferred test for detecting weak instruments, being robust to heteroskedasticity, time series autocorrelation, and clustering in the case of linear instrumental variables regressions with one endogenous regressor

[7] We follow standard procedure in network analysis research (using the non-parametric method implemented in the R package missForest) to determine whether in a given year a specific director was already/still appointed in the corresponding board of directors or not based on the available information. Further information on the procedure followed is available upon request.

[8] Pflueger and Wang (2015) introduce the Stata routine that implements the test for weak instruments by Montiel Olea and Pflueger (2013).

and one or more instruments.[9] Their approach appears of particular relevance in the context of our study given the longitudinal nature of our data.

The effective F-statistic of Montiel Olea and Pflueger (2013) that we obtain when using our three instruments is 184.80 (the corresponding critical value for τ = 5% equals 23.70), thus rejecting the null hypothesis that our instruments are weak in our empirical setting. Given that our three endogenous regressors are the main, squared, and cubic terms of *Internationalization*, while our three instruments correspond to the main, squared, and cubed terms of *Board Internationalization*, to further gauge our main instrument strength we also perform the Montiel Olea-Pflueger test in the simplest case of a just-identified model with a single endogenous regressor (*Internationalization*) and a single instrument—i.e., our main instrument *Board Internationalization*. We obtain an effective F-statistic of Montiel Olea and Pflueger (2013) of 82.57 (the corresponding critical value for τ = 5% equals 37.42). This result further corroborates that our instrument *Board Internationalization* is sufficiently strong to account for the endogeneity of *Internationalization* in our empirical setting.

B&K evaluate the strength of their two instruments when testing for a quadratic relationship between multinationality and performance using the Kleibergen-Paap Wald statistic, which in settings with a single endogenous variable is equivalent to the non-homoskedasticity-robust F-statistics—being equal to the effective F-statistic of Monfiel Olea and Pflueger in the case of a single instrument (Andrews *et al.*, 2019). To provide a close comparison between our and B&K's instruments using the same approach, we also calculate the Kleibergen-Paap Wald statistic when testing for the quadratic relationship and using two instruments, i.e., *Board internationalization* and its squared term. We obtain a Kleibergen-Paap Wald statistic of 10.55, while the one obtained by B&K is 10.51. This result shows that our instruments perform very similarly to the ones used by B&K, thus lending additional support to our conclusion that they are sufficiently strong.

As emphasized by Andrews *et al.* (2019), to date there is still no known analog of the Montiel Olea and Pflueger approach for settings with multiple endogenous regressors, thus calling for more research on this econometric issue. To examine whether our results hold whether or not the instruments are weak, especially when testing for the quadratic and cubic relationships, we follow Andrews *et al.* (2019) and rely on the Anderson-Rubin (AR) test that is robust to weak identification—i.e., its probability of incorrectly rejecting the null hypothesis remains well-controlled also when the instruments are weak. The results we obtained show that the coefficients of the endogenous variables on the right-hand side of our main equations are never jointly significant in any of our model specifications;[10] thus, by never rejecting the null hypothesis that these coefficients equal zero, the AR test further confirms our main finding that once we account for endogeneity we do not find any causal relationship between multinationality and performance.

[9] To date, the case with one endogenous regressor and one or more instruments is the one for which research on developing procedures for detecting weak instruments is most advanced. For this case, as mentioned in the text, Montiel Olea and Pflueger (2013) develop a test that is robust not only to heteroskedasticity, but also to serial autocorrelation and clustering. Refer to the forthcoming article by Andrews *et al.* (2019) for a comprehensive and up to date assessment of the literature on weak instruments in instrumental variable regressions. The forthcoming article and its supplementary materials are available for download here: https://scholar.harvard.edu/stock/publications/weak-instruments-iv-regression-theory-and-practice

[10] The weak-instrument-robust AR test Chi square statistic equals 1.40 (p-value = 0.71) when testing for the linear, quadratic, and cubic relationships.

Being our system not over-identified in the case of the cubic relationship, we cannot directly test the exogeneity condition of our instruments. Given that in our main model estimations we also examine the existence of a (within-firm) linear and quadratic relationship between multinationality and performance, we follow Semadeni *et al.* (2014) and Bascle (2008) to test for the exogeneity of our three instruments in such cases. This is because the equations are over-identified when considering the linear and quadratic relationships (as we have 3 instruments for 1 and 2 endogenous regressors respectively). In both cases the Hansen's J-statistic test does not lead to a rejection (for the linear case: Hansen J-statistic = 1.19 with p-value = 0.55; for the quadratic case: Hansen J-statistic = 0.99 with p-value = 0.32), thus corroborating that our instruments can be considered as exogenous in our setting. This also alleviates concerns about an over-identification problem when focusing on the linear and quadratic relationships.

To further assess the validity of our main instrument *Board Internationalization* in relation to its exogeneity, we also conduct the sensitivity analysis developed by Conley, Hansen, and Rossi (2012).[11] To do so, we focus on the linear association and test the sensitivity of the effect of *Internationalization* on *ROA* by relaxing the exclusion restriction concerning *Board Internationalization*, thus dealing with the potential endogeneity of our main instrument in our empirical setting. The results obtained show that *Internationalization* remains statistically not significant (Coef. = -1.13; p-value = 0.85) even when considering this level of uncertainty on the plausibility of the exclusion restriction of *Board Internationalization*, thus lending additional support to the main findings of our instrumental variable regressions reported in the article.

We also run a dynamic panel data model as an alternative way of assessing causality between multinationality and performance. To do so, we use the estimator outlined by Arellano and Bover (1995) and fully developed by Blundell and Bond (1998), also known as system GMM. While this estimator allows some regressors to be endogenous, it does not assume that good instruments are available outside the immediate dataset; thus, it assumes that the only available instruments are internal, i.e., based on lags of the instrumented variables. Accordingly, using this dynamic panel data model estimator—though presenting several issues whose in-depth assessment goes beyond the purpose of this work—allows us to verify whether our main findings are robust to a different way of accounting for the endogeneity of multinationality. We run the model specifying lags 2 and 3 as instruments for our endogenous variables. We repeat the same models using as lags 2-3-4 and 2-3-4-5. Table A-7 reports the key coefficient estimates obtained when testing for the linear, quadratic, and cubic relationships and using the above-mentioned lag structures. As shown in Table A-7, we find no evidence in support of any causal relationship between multinationality and performance in any of the models considered, thus lending further support to our main findings shown in the article.

As mentioned previously, all our results are obtained requesting heteroscedasticity-robust standard errors. We also repeat the 2SLS estimation requesting not only heteroskedastic but also

[11] Conley *et al.*'s (2012) methodology builds on earlier econometric works that have developed sensitivity analyses for instrumental variables studies, see, e.g., Small (2007). A sensitivity analysis examines the impact of plausible amounts of invalidity of the proposed instruments on inferences of the relevant parameters. Clarke and Matta (2018) introduce the Stata routine that implements the "plausibly exogenous" estimation developed by Conley *et al.* (2012). To run our sensitivity analysis, we adopt the so-called "local-to-zero" estimation approach.

TABLE A-7

The coefficients of the M-P relationship when using dynamic panel data model estimation[1,2]

Relationship tested	Cubic				Quadratic			Linear	
Coefficient / Wald Chi square	Main term	Squared term	Cubed term	Wald Chi square	Main term	Squared term	Wald Chi square	Main term	Wald Chi square
Lags 2-3	0.15 (0.18) [0.43]	-0.006 (0.01) [0.63]	0.0001 (0.0002) [0.76]	1474.12	0.09 (0.11) [0.36]	-0.001 (0.003) [0.55]	1480.15	0.07 (0.05) [0.16]	1476.47
Lags 2-3-4	0.15 (0.18) [0.40]	-0.006 (0.01) [0.65]	0.0001 (0.0002) [0.80]	1475.45	0.10 (0.10) [0.33]	-0.002 (0.003) [0.55]	1481.07	0.07 (0.05) [0.16]	1476.17
Lags 2-3-4-5	0.16 (0.18) [0.38]	-0.006 (0.01) [0.63]	0.0001 (0.0002) [0.77]	1474.88	0.10 (0.10) [0.33]	-0.002 (0.003) [0.58]	1481.21	0.06 (0.05) [0.17]	1475.84

[1] All dynamic panel data models are run using system GMM estimators on the subsample of 102,113 firm-year observations used to run our 2SLS fixed-effects model estimations shown in the article. Roodman (2009) introduces the Stata routine that implements dynamic panel data model estimation.
[2] All independent variables are lagged by one period. All models report (robust standard errors in parentheses) and [p-values in brackets]. Both *Internationalization* and *ROA* are in percentage points.

autocorrelation consistent standard errors and statistics that are therefore robust to both arbitrary heteroskedasticity and arbitrary autocorrelation and the results obtained do not change. Additionally, we perform the Cumby-Huizinga (1992) general test for autocorrelation which fails to reject the null hypothesis according to which there is no serial correlation (Chi square = 0.69; p-value = 0.41), thus further alleviating our concerns of the potential presence of serial correlation in our setting. To address the issue that prior profitability may influence the level of internationalization of a given firm, we also repeat our main model (Model 7 in Table 5) including the lagged profitability measure (i.e., *ROA* at time t-1) as additional control variable. The results obtained when adding this variable into our fixed-effects model specification (available upon request) are entirely aligned with the ones reported in Table 5, thus lending additional support to our conclusion that the results we obtained using the above-mentioned instrumental variable approach account for the potential endogeneity of multinationality.

Finally, our replication focuses on the time period 2009-2016 because 2016 is the last available year for which we were able to collect all the information necessary to run our 2SLS model. We did not collect data prior to 2009 as Orbis provides much less information for earlier years and this would have impacted our ability to closely replicate L&B and B&K's work while also covering the variety of contexts that are instead assessed in the present work. To account for potential time effects in the period considered, we also repeat our main replication model (Model 7 in Table 5) including year dummies and the results obtained do not show any relevant change vis-à-vis the ones reported in Table 5.

Table A-8 shows the results corresponding to the additional analyses discussed in the article in relation to the variety of contexts included in our setting. While Models 1 and 2 focus on

TABLE A-8

Additional analyses[1,2]

Model	1 (Japanese listed multinational firms)	2 (Japanese listed, large multinational firms)	3 (U.S. multinational firms)	4 (U.S. large, manufacturing firms)	5 (U.S. large, listed, manufacturing firms)	6 (Multinational firms in other industry groups)	7 (Multinational SMEs)
Relationship tested	Cubic	Cubic	Cubic	Quadratic	Quadratic	Cubic	Quadratic
Dependent variable	ROA	ROA	ROA	ROA	ROA	ROA	ROA
Internationalization	-9.97	-12.88	-9.51	3.43	2.67	-3.00	12.10
	(35.04)	(48.29)	(13.99)	(3.11)	(3.12)	(7.95)	(25.29)
	[0.78]	[0.79]	[0.50]	[0.27]	[0.39]	[0.71]	[0.63]
Internationalization squared	1.10	1.45	1.48	-0.25	-0.22	0.25	-4.97
	(4.07)	(5.44)	(1.88)	(0.32)	(0.32)	(0.77)	(9.30)
	[0.79]	[0.79]	[0.43]	[0.42]	[0.48]	[0.75]	[0.59]
Internationalization cubed	-0.03	-0.04	-0.04			-0.01	
	(0.12)	(0.16)	(0.04)			(0.01)	
	[0.80]	[0.79]	[0.40]			[0.66]	
Parent size	1.04	1.74	1.95	10.63	11.83	1.82	1.56
	(3.67)	(6.50)	(6.06)	(8.32)	(8.34)	(1.18)	(0.39)
	[0.78]	[0.79]	[0.75]	[0.20]	[0.16]	[0.12]	[<0.0001]
Parent debt-to-equity ratio	-1.20	-1.31	-1.75	-0.85	-0.79	-1.14	-0.98
	(0.57)	(0.71)	(0.75)	(0.89)	(0.94)	(0.43)	(0.11)
	[0.03]	[0.07]	[0.02]	[0.34]	[0.40]	[0.01]	[<0.0001]
Parent export intensity	-1.94	-2.76	-5.66	3.95	4.62	-3.42	-0.99
	(3.67)	(4.40)	(7.56)	(5.67)	(5.78)	(2.91)	(1.07)
	[0.60]	[0.53]	[0.45]	[0.49]	[0.42]	[0.24]	[0.35]
Parent product diversification	4.18	7.06	-6.88	-3.86	-3.71	-0.18	-0.73
	(20.61)	(30.00)	(7.45)	(4.70)	(4.97)	(1.81)	(1.22)
	[0.84]	[0.81]	[0.36]	[0.41]	[0.46]	[0.92]	[0.55]
Exchange rate	-0.11	-0.13	-0.12	0.10	0.11	0.01	0.01
	(0.37)	(0.54)	(0.17)	(0.24)	(0.24)	(0.06)	(0.03)
	[0.76]	[0.81]	[0.47]	[0.67]	[0.66]	[0.86]	[0.84]
Number of observations	3,355	3,154	9,559	4,274	4,119	8,278	59,204
Wald Chi square	66.85	39.89	33.59	41.06	42.08	90.61	114.68
Fixed or random effects	Fixed	Fixed	Fixed	Fixed	Fixed	Fixed	Fixed

[1] All independent variables are lagged by one period. All models report (robust standard errors in parentheses) and [p-values in brackets]. 2SLS fixed-effects model estimation is used for all models. Both *Internationalization* and *ROA* are in percentage points.

[2] In Model 6 we provide the results obtained when restricting our focus to firms excluded from both manufacturing and service industry groups. The other industry group comprehends firms pertaining to the following sectors: Agriculture, Forestry and Fishing (A), Mining and quarrying (B), Electricity, gas, steam and air conditioning supply (D), Water supply; sewerage, waste management and remediation activities (E), and construction (F). More information relative to each sector can be found here: https://ec.europa.eu/eurostat/ramon/nomenclatures/index.cfm?TargetUrl=LST_NOM_DTL&StrNom=NACE_REV2

subsamples of our Japanese setting, Models 3-5 focus on our U.S. setting. Model 6 tests for the presence of a cubic relationship in the other industry groups beyond the service and manufacturing ones, while Model 7 tests for the existence of a quadratic relationship between multinationality and performance in the case of multinational SMEs. The results obtained show no significant effect of the main and squared terms of *Internationalization*, thus failing to replicate the results of Lu and Beamish (2001), who instead found a U-curved relationship when focusing on SMEs.

REFERENCES

Allison PD. 2005. *Fixed Effects Regression Methods for Longitudinal Data Using SAS*. SAS Institute: Cary, NC.

Andrews I, Stock JH, Sun L. 2019. Weak instruments in IV regression: Theory and practice. *Annual Review of Economics*. Forthcoming. Available at: https://scholar.harvard.edu/stock/publications/weak-instruments-iv-regression-theory-and-practice

Arellano M, Bover O. 1995. Another look at the instrumental variable estimation of error-components models. *Journal of Econometrics* **68**: 29-51.

Bascle G. 2008. Controlling for endogeneity with instrumental variables in strategic management research. *Strategic Organization* **6**(3): 285-327.

Bell A, Jones K. 2014. Explaining fixed effects: Random effects modeling of time-series cross-sectional and panel data. *Political Science Research and Methods* **3**(1): 133-153.

Blundell R, Bond S. 1998. Initial conditions and moment restrictions in dynamic panel data models. *Journal of Econometrics* **87**: 115-43.

Bowen HP. 2007. The empirics of multinationality and performance. In *Research in Global Strategic Management, Vol. 13: Regional Aspects of Multinationality and Performance* Rugman A (ed.), Elsevier: The Netherlands.

Certo ST, Withers MC, Semadeni M. 2017. A tale of two effects: Using longitudinal data to compare within- and between-firm effects. *Strategic Management Journal* **38**(7): 1536-1556.

Clarke D, Matta B. 2018. Practical considerations for questionable IVs. *The Stata Journal* **18**(3): 663-691.

Conley TG, Hansen CB, Rossi PE. 2012. Plausibly exogenous. *The Review of Economics and Statistics* **94**(1): 260-272.

Cumby RE, Huizinga J. 1992. Testing the autocorrelation structure of disturbances in ordinary least squares and instrumental variables regressions. *Econometrica* **60**(1): 185-195.

Heckman JJ. 1979. Sample selection bias as a specification error. *Econometrica* **47**(1): 153-161.

Lu JW, Beamish PW. 2001. The internationalization and performance of SMEs. *Strategic Management Journal* **22**(6-7): 565-586.

Montiel Olea JL, Pflueger C. 2013. A robust test for weak instruments. *Journal of Business & Economic Statistics* **31**(3): 358-369.

Mundlak Y. 1978. On the pooling of time series and cross section data. *Econometrica* **46**(1): 69-85.

Nielsen BB, Nielsen S. 2010. A multilevel approach to understanding the multinationality – performance relationship. In *The Past, Present and Future of International Business & Management (Advances in International Management, Volume 23)*. Devinney TM, Pedersen T, Tihanyi L (eds.): Emerald Group Publishing Limited, pp. 527 - 557.

Peterson MF, Arregle JL, Martin X. 2012. Multilevel models in international business research. *Journal of International Business Studies* **43**(5): 451-457.
Pflueger C, Wang S. 2015. A robust test for weak instruments in Stata. *The Stata Journal* **15**(1): 216-225.
Roodman D. 2009. How to do xtabond2: An introduction to difference and system GMM in Stata. *The Stata Journal* **9**(1): 86-136.
Semadeni M, Withers MC, Certo ST. 2014. The perils of endogeneity and instrumental variables in strategy research: Understanding through simulations. *Strategic Management Journal* **35**(7): 1070-79.
Small DS. 2007. Sensitivity analysis for instrumental variables regression with overidentifying restrictions. *Journal of the American Statistical Association* **102**(479): 1049-1058.
Staiger D, Stock J. 1997. Instrumental variables regression with weak instruments. *Econometrica* **65**: 557-586.
Stock J, Yogo M. 2005. Testing for weak instruments in linear IV regression. In Andrews DWK (ed.), *Identification and Inference for Econometric Models*, Cambridge University Press: New York, pp. 80-101.
Verbeke A, Forootan MZ. 2012. How good are multinationality-performance (M-P) empirical studies? *Global Strategy Journal* **2**(4): 332-344.