Strategic Management Journal
Strat. Mgmt. J., **37**: 2275–2290 (2016)
Published online EarlyView in Wiley Online Library (wileyonlinelibrary.com) DOI: 10.1002/smj.2567
Received 28 September 2014; *Final revision received 25 January 2016*

REPLICATING THE MULTINATIONALITY-PERFORMANCE RELATIONSHIP: IS THERE AN S-CURVE?

HEATHER BERRY[1]* and ASEEM KAUL[2]
[1] *International Business Department, School of Business, George Washington University, Washington, DC, U.S.A.*
[2] *Strategic Management and Entrepreneurship Department, Carlson School of Management, University of Minnesota, Minneapolis, Minnesota, U.S.A.*

Research summary: *We revisit the empirical relationship between multinationality and performance by attempting to replicate the widely cited S-shape relationship reported in Lu and Beamish (2004). Using a longitudinal and comprehensive database on the population of U.S. MNCs from 1989 to 2007, we find no evidence of an S-shaped relationship; nor do we see a moderating effect of intangible assets. Although our results do show a marginally significant U-shaped association between multinationality and performance for a subsample of manufacturing firms, this relationship disappears once we account for the endogeneity of multinationality. Our study contributes to empirical research on the multinationality-performance relationship, highlighting the need for caution in generalizing results across countries and the importance of controlling for the endogeneity of multinationality when assessing its effect on performance.*

Managerial summary: *Our study examines the relationship between a firm's multinationality and its performance. In a much-cited study, Lu and Beamish (2004) found evidence of an S-shaped relationship—with firm performance first decreasing, then increasing, then decreasing again as firms internationalized—in a sample of Japanese firms from 1986 to 1997. We test for the same relationship across all U.S. MNCs from 1989 to 2007, and find no evidence of an S-shaped pattern, or indeed, of any effect of multinationality at an aggregate level. Our study thus suggests that the effect of multinationality may vary with firm capabilities and home country environments, and that managers and academics alike should focus on understanding these specifics, rather than searching for a universal effect of multinationality on performance.*

INTRODUCTION

The effect of multinationality[1] on firm performance is a central question in the strategy and international business literatures. Decades of empirical research seeking to estimate the relationship between a firm's multinationality and its performance at an aggregate level have failed to produce a consistent set of empirical results, with prior studies finding evidence for a variety of different multinationality-performance relationships across a range of samples (Cardinal, Miller, and Palich, 2011; Grant, Jamine, and Thomas, 1988; Hitt *et al.*, 2006; Kirca *et al.*, 2011). These mixed results have, in turn, prompted several recent criticisms of this literature, with some scholars arguing

Keywords: firm performance; multinationality; endogeneity; replication; S-curve
*Correspondence to: H. Berry, International Business Department, Funger 401T, School of Business, George Washington University, Washington DC 20052. E-mail: berryh@gwu.edu
H. Berry and A. Kaul contributed equally.
[1]We use the term multinationality to refer to the extent of a firm's operations outside its home country. Alternate terms used in prior studies for the same construct include "geographic diversification" and "internationalization".

for better empirical approaches that use longitudinal data, account for endogeneity, and recognize the multidimensional nature of multinational activity (Contractor, 2012; Verbeke and Forootan, 2012; Wiersema and Bowen, 2011). Some have even questioned the theoretical premise of a persistent relationship between multinationality and performance (Hennart, 2011).

A key contribution to the literature on the multinationality-performance relationship comes from an article by Lu and Beamish (2004; henceforth L&B). Seeking to reconcile prior empirical findings of both a U-shaped (Capar and Kotabe, 2003; Lu and Beamish, 2001) and an inverted U-shaped (Geringer, Beamish, and Da Costa, 1989; Gomes and Ramaswamy, 1999; Hitt, Hoskisson, and Kim, 1997) relationship between multinationality and performance, L&B offer an S-curve hypothesis, arguing that the relationship will be U-shaped for newly internationalizing firms who face liabilities of foreignness and newness, but inverted-U shaped for well-established multinationals who face diminishing returns to multinationality along with increasing costs of coordination. They confirm this hypothesis using longitudinal data on Japanese firms from 1986 to 1997, showing that firm performance (measured using both accounting and market-based measures) first decreases, then increases, and then decreases again as multinationality increases; a relationship that they interpret as causal. L&B also argue and show that the relationship between multinationality and performance is more positive, the higher the level of intangible assets the firm possesses. In the years since their article was published, L&B's S-shaped hypothesis has come to be widely accepted and cited, though as the authors themselves acknowledge, their results are based on firms from a single country, 'thus raising the concern that the findings might be country-specific' (Lu and Beamish, 2004: 607), and empirical evidence from other studies that have tested for an S-shape provides mixed support (Contractor, Kundu, and Hsu, 2003; Qian *et al.*, 2010; Ruigrok, Amann, and Wagner, 2007).[2] Careful, large sample replication of L&B's study in other samples and settings is therefore critical.

In this article, we undertake such a replication. Using a longitudinal dataset of 2,023 firms over an 18-year period from 1989 to 2007, based on comprehensive survey data from the Bureau of Economic Analysis (BEA) that cover the population of U.S. multinational corporations (MNCs), and adopting measures and methods nearly identical to those used in L&B, we test for both the existence of an S-shaped relationship between multinationality and performance, and a positive moderating role of intangible assets. We are, however, unable to replicate L&B's results in our sample: we find no evidence of an S-shaped relationship between multinationality and performance for US MNCs, nor do we find a positive moderation of this relationship by the firm's intangible assets (as measured by R&D spending). We do see a marginally significant U-shaped association between multinationality and performance in our replication of L&B's main analysis in a subsample of manufacturing firms, but this too disappears when we try to replicate their robustness checks, specifically, when we account for the endogeneity of multinationality. Overall, our results show no evidence of a causal effect of aggregate multinationality on performance.

Our findings contribute to empirical research on the multinationality-performance relationship by using large-sample, longitudinal data to demonstrate the frailty of the S-curve hypothesis that has come to be widely accepted in this literature. These findings emphasize the need for caution while generalizing results across country contexts, and the consequent need for careful replication. They also highlight the importance of accounting for endogeneity when assessing the multinationality-performance relationship, and the need to pay attention to the range of observations within the study sample. In particular, the results from our replication of the robustness tests undertaken by L&B seem to support the theoretical argument that multinationality may have no aggregate effect on firm performance once the endogeneity of multinationality is accounted for, since firms will choose a level of multinationality that is consistent with their weaknesses and strengths (Hennart, 2011). Our study thus contributes to an emerging debate about the value of studying the relationship between multinationality and performance at an aggregate level, and the need for more contingent, contextually grounded, and empirically robust approaches when doing so (Verbeke and Forootan, 2012; Wiersema and Bowen, 2011).

[2]Though these studies test the S-shaped relationship, they are not designed to replicate L&B's findings, and therefore use different measures and methods. They also have more limited coverage, relying on samples of a relatively small number of firms and/or on cross-sectional data analysis.

DOI: 10.1002/smj

THE S-CURVE HYPOTHESIS

We begin by offering a brief summary of the study we intend to replicate: Lu and Beamish (2004). In their study, Lu and Beamish (L&B) identify three distinct phases of internationalization. In Phase 1, they argue that firms face substantial liabilities of foreignness and newness, which causes their performance to fall as they first internationalize. As firms expand their international presence, however, these liabilities of foreignness and newness are overcome, and firms see increasing benefits to being international, with the result that firm performance starts to rise with multinationality (Phase 2). Eventually, however, the benefits from multinationality plateau as firms experience diminishing returns, while the coordination costs of operating across multiple countries increase at an increasing rate. As a consequence, increasing multinationality beyond an optimal threshold will cause firm performance to start falling again (Phase 3). Together, these arguments imply an S-shaped relationship between multinationality and firm performance, with performance first falling, then rising, then falling again with increasing multinationality. Further, L&B predict that this relationship shall be positively moderated by the level of a firm's intangible assets, arguing that 'firms with intangible assets should be able to generate abnormal high returns from their foreign direct investments through scale and scope economies and through the exploitation of market imperfections in the trade of intangible assets', and that this effect will hold 'for every internationalizing firm, irrespective of its base level of internationalization' (Lu and Beamish, 2004: 602). These arguments lead them to propose the following two hypotheses, quoted verbatim (Lu and Beamish, 2004: 601 and 602):

> *Hypothesis 1: The relationship between geographic diversification and firm performance is nonlinear, with the slope negative at low levels of geographic diversification, positive at medium levels of geographic diversification, and negative at high levels of geographic diversification.*
>
> *Hypothesis 2: A firm's intangible assets moderate the relationship between geographic diversification and firm performance in such a way that high levels of intangible assets increase the performance gains attributable to geographic diversification.*

L&B test these hypotheses in a sample of 1,489 Japanese firms from 1986 to 1997, which includes 1,059 firms that have at least some foreign direct investment. Specifically, they create an internationalization index by taking (separately) the number of countries in which the firm operates and the number of subsidiaries it has in any given year, dividing that number by the maximum number of countries and the maximum number of subsidiaries in the sample respectively, and then averaging the two ratios to get a single composite index of internationalization that ranges from 0 to 1. They then regress two measures of firm performance—one accounting based (Return on Assets) and the other market based (Tobin's Q)—on this internationalization index, along with a number of relevant control variables. They use a generalized least squares random-effects models as their main specification to account for the longitudinal nature of their data, justifying their choice with Hausman tests, which show that "random-effects models were preferred in all cases" over fixed-effects models (Lu and Beamish, 2004: 603).[3]

Lu and Beamish (2004) find support for both their hypotheses (Lu and Beamish, 2004: 604, table 2). They find an S-shaped relationship between multinationality and performance for both accounting and market-based measures, with a negative and significant coefficient for internationalization, a positive and significant coefficient for internationalization squared, and a negative and significant coefficient for internationalization cubed, supporting Hypothesis 1. They also find some support for Hypothesis 2, with the interaction between internationalization and an R&D-based measure of intangible assets taking a positive and significant coefficient when predicting ROA, and the interaction between internationalization and an advertising-based measure of intangible assets taking a positive and significant coefficient when predicting Tobin's Q. In addition to these main results, L&B also mention a number of robustness checks, though they do not report the results of these tests in the published article. Here, reported verbatim, is their description of the robustness tests they ran:

[3]Presumably meaning that the Hausman test found no significant difference between random-effects and fixed-effects results, allowing the authors to report random-effects models instead of fixed-effects models.

DOI: 10.1002/smj

> We increased the lag structure to 2 years and 3 years and obtained consistent results, although the variance explained (the value of R^2) becomes smaller as the lag increases. The results are robust in the subsample of 1,059 firms with FDI activities. The results are also robust to tests using ROA as the dependent variable in a sample (2,067 firms) not constrained to firms for which we had Tobin's Q data. We also centered the variables on their means to minimize their collinearity; the results were robust to the use of centered variables. Finally, the relationship we were modeling is a complex one in which both assets (such as technological assets) and strategy (such as internationalization) could be endogenous. To handle this endogeneity issue, we used a two-stage least squares procedure (2SLS) in which technological assets and internationalization were entered as endogenous variables. The S-shaped relationship was robust to the 2SLS procedure, while R&D intensity took a positive sign. (Lu and Beamish, 2004: 605–606)

REPLICATION: DATA AND METHODS

Sample and data

In order to replicate L&B's study, we draw on data on US multinationals from the Bureau of Economic Analysis (BEA). The BEA collects detailed data on the worldwide operations of U.S. MNCs through mandatory surveys, which makes these data the most comprehensive that are available for the population of U.S. MNCs.[4] We use firm-level data collected in the BEA surveys of US direct investment abroad from 1989 to 2009 to construct our measures of performance and multinationality, as well as several other variables. Because reporting to the BEA is legally mandated, these data have the advantage that they represent as close to the population of U.S. MNCs as it is reasonably possible to get, effectively eliminating concerns about sampling bias. In addition, the data are collected annually in a consistent manner, allowing us to study within firm changes in multinationality and performance over a long period of time. These data have been widely used in strategy research both at the aggregate industry level (Kobrin, 1991; Nachum and Zaheer, 2005; Wiersema and Bowen, 2011) and at the firm level (Berry, 2013, 2014; Berry and Kaul, 2015; Chung and Yeaple, 2008), though never, to the best of our knowledge, to directly study the relationship between an MNC's aggregate multinationality and its performance.

In the current study, we limit our sample in two ways. First, while we have access to firm-level BEA data from 1989 to 2009, we limit our sample to the period 1989–2007. This is because the patent-based measures we use to account for endogeneity (described in more detail below), are only available to us up to 2006. Second, while the BEA collects data on all foreign affiliates of U.S. MNCs, we consider only those affiliates in which the firm has majority ownership, which we term subsidiaries. Our final sample thus consists of an unbalanced panel of 2,023 firms over 18 years, with a total of 21,297 firm-year observations, making it one of the largest samples used to study the multinationality-performance relationship.

In addition to studying a different home country (U.S. vs. Japan) and a different time period (1989 to 2007 vs. 1986 to 1997), our sample differs from that used by L&B in two ways. First, our sample is limited to multinationals, i.e. firms with at least some foreign investment, while their sample included purely domestic firms as well. We do not see this as a serious concern, however, both because they themselves report that their results hold in the subsample of firms with foreign investments, and because firms who have no foreign investments in our study period would have no variance in multinationality and therefore would not contribute much to a within-firm analysis of the effect of multinationality on performance. We therefore see no reason to expect different results in our multinational-only sample from those we would obtain in a sample that included domestic firms. Second, while Lu and Beamish's primary sample includes only publicly listed firms, ours includes all multinationals both public and private. We see this as a strength of our sample, since it means that our coverage is potentially broader than that of L&B. Further, L&B report robustness of their results to include firms that were

[4]Specifically, the International Investment and Trade in Services Survey Act requires U.S. MNCs to report detailed information on the financial and operating activities of both U.S. parent companies and their foreign affiliates, as well as information on the value of transactions between the parents and affiliates. (See Mataloni and Yorgason, 2006, for a thorough description of definitions and survey methodology used by the BEA.)

DOI: 10.1002/smj

not public, which limits concern about this difference as well.

Measures

Since our primary purpose is to replicate L&B's study, we have tried to develop measures that are as close as possible to their measures. Thus, we use return on assets (*ROA*), calculated as the ratio of the firm's net income[5] to its total assets, as our primary dependent variable. While L&B also use a market-based performance measure (Tobin's Q) in their study, their results for the S-curve are similar using both measures, so we believe that the use of an accounting-based measure is sufficient for replication, and is moreover consistent with the nature of our sample, which includes both public and private firms. The use of *ROA* as the main dependent variable is also consistent with the literature on the multinational-performance relationship more generally, being the measure most commonly used in these studies (Kirca *et al.*, 2011).

Turning to independent variables, we follow L&B and create a composite *Internationalization* index based on number of foreign countries and number of foreign subsidiaries. Following L&B, we divide the number of foreign countries in which the firm operates in a given year by the maximum number of countries in which any firm in our sample operates in that year, undertake a similar calculation for the number of foreign subsidiaries, and then average the two scaled variables to create an overall composite index of *Internationalization* that ranges between 0 and 1. To test the S-curve hypothesis we include both this main *Internationalization* measure, and its square and cubed terms. We also construct a measure of the firm's *R&D Intensity* (i.e. the parent firm's R&D spending as a percentage of its sales) to capture the level of a firm's intangible assets, and include the interaction between *Internationalization* and *R&D Intensity* in order to replicate the test for L&B's second hypothesis.[6] In order to replicate L&B's measures as closely as possible, we do not mean center these variables in our main analysis, but run a separate robustness check with mean centered variables.

In addition to these variables of interest, we include several controls that were also included by L&B. These include controls for firm *Sales* (measured as logarithm), *Debt-to-Equity ratio*, *Product Diversification, Export Intensity*, and the prevailing *Exchange Rate*. In addition, we include industry-fixed effects when running random-effects models, though we drop these when we switch to using fixed-effects (as discussed below). Each of these measures is calculated in the same manner as those used in L&B. As already mentioned, we are unable to control for firm's advertising intensity, because the BEA does not capture this information, but this one variable aside, our measures exactly replicate those used by L&B. As with L&B, we lag all independent variables by one year.

Methods

As mentioned above, L&B use random-effects GLS models for their analysis, justifying this choice using a Hausman test. In our sample, however, a Hausman test comparing otherwise identical models with random-effects and fixed-effects clearly rejects the hypothesis that the two are the same with a p-value of less than 0.01, implying that we need to use fixed-effects rather than random-effects. Thus, while we report both random-effects and fixed-effects below in order to fully replicate L&B's study, we focus primarily on fixed-effects models when discussing our findings. Our assumption is that since the Hausman test was insignificant in L&B's case, their fixed-effects results (which they do not report) would be similar to their random-effects results, making a comparison between their random-effects results and our fixed-effects results a fair one.

In addition to the fixed-effects panel models (run using the xtreg command in STATA) that we use to replicate L&B's main findings, we also use a two-stage least squares (2SLS) model to replicate their final robustness check. Unfortunately, since L&B neither show the results for this analysis in their published article, nor discuss what instruments they used for the first stage, we cannot exactly replicate the procedure they use here. Nevertheless, we attempt to replicate their robustness check, using a 2SLS procedure to account for the endogeneity of *Internationalization* and *Internationalization squared*.

[5]Net income is defined as income minus costs and expenses, which includes income taxes.

[6]L&B also use Advertising Intensity as a second measure of firm intangible assets. We are unable to replicate this measure since the BEA does not capture information on advertising spending. Note, however, that Advertising Intensity has an insignificant moderating effect on ROA in their study, and so is not directly relevant to our replication anyway.

DOI: 10.1002/smj

Table 1. Measures

No	Variable	Description	Source
	Dependent variable		
1	ROA	MNC net income (defined as income minus costs and expenses)/MNC assets	BEA
	Independent variables		
2	Internationalization	$\frac{1}{2}\left[\frac{Countries_i}{\max(countries_i)} + \frac{Subsidiaries_i}{\max(subsidiaries_i)}\right]$ for every firm i in each year. Range is between 0 and 1	BEA (based on Lu and Beamish, 2004)
3	Parent R&D intensity	Parent R&D expenditures/parent sales	BEA
4	Parent size	Log of parent sales	BEA
5	Parent debt-to-equity ratio	Parent debt/(parent total assets – parent debt)	BEA
6	Parent export intensity	Parent exports/parent sales	BEA
7	Parent product diversification	Berry-Herfindahl index of product diversification in parent firm operations	BEA
8	Exchange rate	Annual real effective exchange rate between countries in the Euro area EU(17) and the U.S.	Eurostat
	Instrumental variables		
9	Industry foreign knowledge	Stock of industry-relevant patents with non-U.S. assignees as a proportion of total industry relevant patent stock in MNC's main industry	NBER, Silverman Concordance
10	Industry shipping value by weight	Average product value by weight in MNC's main industry	Pierce and Schott (2009)

We use two instruments to identify the first stage of our 2SLS models. First, we use a measure of *Industry Foreign Knowledge*, i.e. the extent to which technological knowledge relevant to the firm's industry is being developed by firms abroad (i.e. outside the U.S.). We expect firms in industries with greater foreign knowledge to pursue more R&D abroad (and therefore to be more multinational), but do not expect the fact that knowledge in the firm's industry is more global to impact its profitability. To construct our measure of *Industry Foreign Knowledge* we follow prior work (Berry and Kaul, 2015) and calculate the proportion of industry relevant patents filed with the US Patent Office in that year that are assigned to non-U.S. firms, using data from the NBER patent data project, and a concordance between patent class and industry code developed by Silverman (1996).

Our second instrument is the *Industry Shipping Value By Weight* ratio of the industry's products. The higher the value-by-weight ratio of the industry's products, the lower the shipping costs as a proportion of revenues, which are likely to impact overall multinationality (Vernon, 1966) but should not have a separate effect on within-firm performance. In order to construct our measure of the value-by-weight ratio of the industry's products, we use trade data from the U.S. Census Bureau that capture the value and weight of all imports into the US. These data are aggregated up to the industry level using a concordance between product and industry developed by Pierce and Schott (2009) to create a measure of the average ratio of value to weight for every industry in every year.[7] Since we are primarily interested in cross-industry differences in this ratio, we further aggregate across years to create a time-invariant measure of the average value-by-weight ratio of the industry's imports, which serves as our second instrument.[8]

Table 1 describes the various measures used in our analysis. Table 2 provides summary statistics and correlations of these measures.

[7]The data used to create this measure were downloaded from Peter Schott's website in September 2014: http://faculty.som.yale.edu/peterschott/sub_international.htm

[8]We confirm the validity of these instruments empirically, with both instruments significantly predicting *Internationalization* in the first stage (not shown) and the Kleibergen-Paap Wald statistic (reported below) showing that our instruments are sufficiently strong.

DOI: 10.1002/smj

Table 2. Summary statistics

Variable	Mean	Std. Dev.	1	2	3	4	5	6	7
1 ROA	0.04	0.16	1.00						
2 Internationalization	0.10	0.09	0.04	1.00					
3 Parent R&D intensity	0.03	0.19	0.01	0.12	1.00				
4 Parent size	13.81	1.83	0.03	0.32	0.21	1.00			
5 Parent debt-to-equity ratio	2.17	2.67	−0.08	−0.03	0.09	0.20	1.00		
6 Parent export intensity	0.05	0.21	0.01	0.01	0.03	−0.06	−0.03	1.00	
7 Parent product diversification	0.24	0.25	−0.02	0.17	−0.03	0.22	0.07	−0.02	1.00
8 Exchange rate	1.00	0.09	0.01	−0.01	0.01	0.01	0.03	−0.01	−0.03

Table 3. Sample comparison of means and standard deviations across Lu and Beamish (2004) and BEA sample and subsamples

	L&B sample	BEA sample	BEA manufacturing subsample	BEA non-manufacturing subsample
ROA	0.04	0.04	0.04	0.03
	(0.05)	(0.16)	(0.14)	(0.21)
Internationalization	0.04	0.10	0.11	0.08
	(0.07)	(0.09)	(0.18)	(0.07)
Parent R&D intensity	0.01	0.03	0.04	0.02
	(0.02)	(0.19)	(0.07)	(0.12)
Parent size	11.06	13.71	13.76	13.66
	(1.48)	(1.83)	(1.59)	(1.99)
Parent debt-to-equity ratio	3.26	2.17	1.76	2.48
	(6.75)	(2.67)	(2.21)	(2.94)
Parent export intensity	0.10	0.05	0.15	0.03
	(0.15)	(0.21)	(0.27)	(0.41)
Parent product diversification	0.57	0.24	0.35	0.20
	(0.18)	(0.25)	(0.26)	(0.24)
Exchange rate	120.77	1.00	1.00	1.00
	(12.35)	(0.09)	(0.09)	(0.09)
Number of firms	1,489	2,023	982	1,041

Sample means shown in table with (standard deviations in parentheses).

FINDINGS

Main results

We begin by comparing our sample to L&B's sample, based on the reported summary statistics. This comparison is shown in Table 3, which compares the means and standard deviations of the firms in the two samples, with our full sample further broken down into manufacturing and non-manufacturing firms. Table 3 shows that the means are comparable across the two samples, lying within one standard deviation of each other across all relevant full sample or subsample comparisons. Although not a statistically significant difference, Table 3 does suggest that the firms in our sample are somewhat more international, with a mean *Internationalization* index of 0.10 and a standard deviation of 0.09, compared to 0.04 and 0.07 for the L&B sample respectively.[9] In part, this is because the firms in our sample are somewhat more multinational on average, operating in an average of 8.11 countries (compared to 3.96 for L&B's sample firms; 5.57 for those with FDI) with an average of 19.23 subsidiaries (compared to 8.45 for L&B; 11.88 for

[9]Since the L&B sample includes domestic firms, it is more meaningful to compare the means for the subsample of their firms with FDI. While L&B do not report this in their article, we can calculate these summary statistics using the fact that *Internationalization* is 0 for domestic firms by definition. The multinational firms in L&B's sample thus have a mean *Internationalization* of 0.06 and a standard deviation of 0.08, still lower than our sample.

DOI: 10.1002/smj

Table 4. Main replication results

Dependent variable Model	ROA 0 (L&B)	ROA 1	ROA 2	ROA 3	ROA 4	ROA 5	ROA 6
Internationalization	−0.39	−0.02	0.05	−0.03	−0.42	−0.04	−0.41
	(−11.15)	(0.20)	(0.07	(0.23)	(0.29)	(0.08)	(0.29)
		[0.92]	[0.46]	[0.89]	[0.16]	[0.63]	[0.16]
Internationalization squared	0.67	0.21		0.21	0.85		0.87
	(4.55)	(0.73)		(0.75)	(0.85)		(0.86)
		[0.77]		[0.78]	[0.32]		[0.31]
Internationalization cubed	−0.40	−0.12		−0.12	−0.44		−0.46
	(−2.75)	(0.52)		(0.57)	(0.62)		(0.62)
		[0.81]		[0.83]	[0.47]		[0.46]
Internationalization × R&D intensity	1.06		0.21	0.15		0.35	0.26
	(2.26)		(0.75)	(0.63)		(0.67)	(0.70)
			[0.78]	[0.81]		[0.61]	[0.71]
Parent R&D intensity	−0.26	0.01	−0.01	−0.01	−0.01	−0.01	−0.01
	(−4.47)	(0.01)	(0.20)	(0.05)	(0.01)	(0.08)	(0.03)
		[0.28]	[0.96]	[0.83]	[0.20]	[0.90]	[0.77]
Parent size	0.01	0.01	0.01	0.01	0.01	0.01	0.01
	(17.54)	(0.01)	(0.01)	(0.01)	(0.00)	(0.00)	(0.00)
		[0.13]	[0.14]	[0.14]	[0.01]	[0.00]	[0.01]
Parent debt-to-equity ratio	0.00	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01
	(9.55)	(0.00)	(0.00)	(0.00)	(0.03)	(0.03)	(0.03)
		[0.01]	[0.01]	[0.01]	[0.72]	[0.70]	[0.72]
Parent export intensity	−0.02	0.01	−0.01	−0.01	0.01	0.01	0.01
	(−2.96)	(0.02)	(1.04)	(0.51)	(0.01)	(0.01)	(0.01)
		[0.56]	[0.99]	[0.98]	[0.33]	[0.33]	[0.32]
Parent product diversification	−0.02	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01
	(−3.95)	(0.01)	(0.01)	(0.01)	(0.02)	(0.02)	(0.02)
		[0.47]	[0.46]	[0.47]	[0.65]	[0.58]	[0.64]
Exchange rate	0.00	−0.12	−0.12	−0.12	−0.12	−0.12	−0.12
	(20.99)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)
		[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Intercept	−0.16	0.13	0.13	0.14	0.25	0.23	0.25
	(−8.09)	(0.03)	(0.03)	(0.03)	(0.05)	(0.05)	(0.05)
		[0.00]	[0.00]	[0.00]	[0.05]	[0.05]	[0.05]
N	(Not reported)	21,297	21,297	21,297	21,297	21,297	21,297
R square	0.10	0.08	0.09	0.09	0.10	0.10	0.10
Wald Chi square	1130.02	94.79	94.69	94.84	8.39	9.08	8.64
Fixed or random effects	Random	Random	Random	Random	Fixed	Fixed	Fixed
Industry dummies	Yes	Yes	Yes	Yes	No	No	No

All independent variables are lagged by one period. The L&B results reported in Model 0 have (t-stats in parentheses) as reported in the original Lu and Beamish (2004) article. All other models report (robust standard errors in parentheses) and [p-values in brackets].

those with FDI). Moreover, since the *Internationalization* index is calculated by dividing each firm's count of countries and subsidiaries by the sample maximum, the higher average for our sample also reflects the greater dispersion of the firms within the range of our sample.

Having confirmed that our sample is comparable on observables to that of L&B, we next turn to our replication of their main results, which is shown in Table 4. Model 0 in Table 4 simply copies the results from L&B that we are attempting to replicate (Lu and Beamish, 2004; table 2, Model 5), showing both the S-curve and a positive interaction between *R&D intensity* and *Internationalization* when predicting ROA. Models 1–3 show the results of our replication using random-effects models (with industry-fixed effects), with Model 1 showing only the S-curve (i.e. including the main, squared, and cubed terms of *Internationalization*) without the interaction with *R&D intensity*, Model 2 showing

DOI: 10.1002/smj

only the interaction between *Internationalization* and *R&D Intensity* (without the squared and cubed terms) and Model 3 being the full model. Models 4–6 then repeat the same analysis with firm-fixed effects.

Overall, Table 4 shows that L&B's results are not replicated in our sample. We see no evidence for a significant effect of either the main, squared, or cubed term of *Internationalization* in any of our specifications (the signs of the coefficients are consistent with L&B's results, but no p-value is less than 0.16), implying that there is no support for an S-curve, or indeed, for any relationship between multinationality and performance in our full sample.[10] We also see no evidence of a significant interaction between *Internationalization* and *R&D intensity* in any of our models. Since the Hausman test fails to support the use of random-effects models in our sample,[11] Model 6 is the most appropriate replication of L&B's study, which continues to show no evidence for either the S-curve relationship or the interaction between *Internationalization* and *R&D intensity*.

Robustness checks

Having compared our findings to L&B's results in the full sample, we next turn to look at the robustness of these findings by undertaking three sets of analyses. First, we check the robustness of our results by staging our replication, i.e. dropping each control variable from our final model to see if it impacts our findings.[12] The results of this analysis are shown in Table 5, which shows that our findings are unaffected when we drop the various controls one by one (Models 7–11)—we continue to find no evidence of either an S-curve, or a significant moderating effect of *R&D intensity*.

Second, we check the robustness of our results in subsamples of firms split by time period and industry sector. It may be that the differences between our findings and those of L&B result from the different time periods we cover, with our data running from 1989 to 2007 while theirs cover the period from 1986 to 1997. To explore this possibility we split our sample into two time periods—one that overlaps with L&B (1989–1997) and one that comes after their study period (1998–2007). The results of this subsample analysis are shown in Models 12 and 13 in Table 6, with Model 12 showing the full model (corresponding to Model 6 in Table 4) for the earlier period, while Model 13 shows the same model for the later period. Comparing across the two time periods, we continue to find no support for either the S-curve or the interaction between *Internationalization* and *R&D intensity* in either of these time periods.

Table 6 also shows the results of splitting our sample into manufacturing and non-manufacturing firms. We do so to partly explore the possibility that the difference between our findings and those of L&B may result from the different mix of industries in their sample relative to ours. Here, we find some evidence for a U-shaped relationship in the manufacturing subsample (Model 14), while seeing no significant relationship between *Internationalization* and *ROA* in the non-manufacturing subsample (Model 15). Although the results in Model 14 provide some evidence for a non-linear relationship between *Internationalization* and *ROA*, replicating at least part of the results reported in L&B, the coefficient of the *Internationalization cubed* term continues to be insignificant, with a p-value of 0.46, meaning the S-curve is not supported; nor do we see a significant interaction between *Internationalization* and *R&D intensity*. Model 16 re-runs Model 14 dropping these insignificant terms, to give us the best fit model for the manufacturing subsample, which continues to show a negative coefficient for *Internationalization*, and a positive coefficient for *Internationalization Squared*, with p-values of 0.06 and 0.02 respectively, implying a significant U-shaped relationship.

In addition to comparing statistical significance, it is also interesting to compare the coefficient estimates from our results to those reported by L&B. To do so, we compare the results in their Model 5 (reported as Model 0 in our Table 4) to the results from Model 16 in Table 6. We base our comparisons off Model 16 rather than Model 6 because it seems pointless to interpret coefficients that are not statistically significant, and because the point estimates from Model 6 imply that performance is constantly declining with multinationality in the relevant range, making comparisons

[10]Robustness tests (not shown) confirm that the coefficient of *Internationalization* remains statistically insignificant even when we drop all interactions and higher order terms.

[11]The Hausman test comparing the specification in Model 6 with fixed and random effects rejects the hypothesis that individual firm level effects are adequately modeled by a random effects model (Hausman statistic 41.14, $p < 0.01$).

[12]We are grateful to the co-editors of the Special Issue for suggesting this analysis.

DOI: 10.1002/smj

Table 5. Robustness to dropped controls

Dependent variable Model	ROA 6	ROA 7	ROA 8	ROA 9	ROA 10	ROA 11
Internationalization	−0.41	−0.44	−0.41	−0.44	−0.48	−0.38
	(0.29)	(0.21)	(0.29)	(0.21)	(0.34)	(0.30)
	[0.16]	[0.15]	[0.16]	[0.16]	[0.16]	[0.17]
Internationalization squared	0.87	0.93	0.88	0.91	0.88	0.80
	(0.86)	(0.87)	(0.87)	(0.87)	(0.81)	(0.87)
	[0.31]	[0.28]	[0.31]	[0.29]	[0.28]	[0.35]
Internationalization cubed	−0.46	−0.49	−0.47	−0.48	−0.42	−0.44
	(0.62)	(0.62)	(0.64)	(0.63)	(0.54)	(0.64)
	[0.46]	[0.43]	[0.46]	[0.45]	[0.48]	[0.49]
Internationalization × R&D intensity	0.26	0.26	0.26	0.19	0.23	0.16
	(0.70)	(0.72)	(0.70)	(0.68)	(0.34)	(0.73)
	[0.71]	[0.73]	[0.71]	[0.78]	[0.76]	[0.83]
Parent R&D intensity	−0.01	−0.01	−0.01	−0.01	−0.01	−0.01
	(0.03)	(0.04)	(0.05)	(0.02)	(0.05)	(0.03)
	[0.77]	[0.82]	[0.83]	[0.67]	[0.85]	[0.77]
Parent size	0.01		0.01	0.01	0.01	0.01
	(0.00)		(0.00)	(0.00)	(0.00)	(0.00)
	[0.01]		[0.01]	[0.01]	[0.01]	[0.01]
Parent debt-to-equity ratio	−0.01	−0.01		−0.01	−0.01	−0.01
	(0.03)	(0.03)		(0.03)	(0.01)	(0.02)
	[0.72]	[0.69]		[0.69]	[0.49]	[0.68]
Parent export intensity	0.01	0.01	0.01		0.01	0.01
	(0.01)	(0.01)	(0.01)		(0.01)	(0.01)
	[0.32]	[0.31]	[0.31]		[0.20]	[0.14]
Parent product diversification	−0.01	−0.01	−0.01	−0.01		−0.01
	(0.02)	(0.02)	(0.02)	(0.02)		(0.03)
	[0.64]	[0.60]	[0.64]	[0.60]		[0.69]
Exchange rate	−0.12	−0.12	−0.12	−0.13	−0.12	
	(0.02)	(0.02)	(0.02)	(0.02)	(0.02)	
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	
Intercept	0.25	0.22	0.25	0.23	0.16	0.13
	(0.05)	(0.03)	(0.05)	(0.03)	(0.03)	(0.04)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
N	21,297	21,297	21,297	21,297	21,297	21,297
R square	0.10	0.09	0.10	0.09	0.09	0.06
Wald Chi square	8.64	8.37	8.39	8.18	6.75	2.14
Fixed or random effects	Fixed	Fixed	Fixed	Fixed	Fixed	Fixed

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

challenging. We undertake two comparisons. First, we consider the threshold values implied by our model vs. theirs.[13] Taking the point estimates of the coefficients in their model, and setting R&D intensity to three percent (which is equal to one standard deviation above its reported mean in their sample),[14] *ROA* is falling for values of *Internationalization* less than 0.44 (i.e. for values of *Internationalization* less than 5.7 standard deviations above the sample mean), then rising till *Internationalization* reaches 0.67 (9 standard deviations above the sample mean), and then falling again. In comparison, our results in Model 16 suggest

[13]In their article, L&B report threshold values and economic magnitudes only for their analysis predicting Tobin's Q (p. 605); we use their reported results to undertake a parallel calculation for their models predicting ROA. To confirm that our calculations are equivalent, we also re-calculate the threshold values and economic magnitudes for their Tobin's Q analysis, and confirm that we get the same values that they report.

[14]We choose this value of R&D intensity rather than the sample mean of one percent because the S-curve is not defined at the mean value. Taking the model reported by L&B, and setting R&D equal to one percent, $\frac{\partial ROA}{\partial Internationalization} = -1.2\ Internationalization^2 + 1.34\ Internationalization - 0.38$ which is a quadratic with only imaginary roots.

DOI: 10.1002/smj

Table 6. Robustness results for time period splits and manufacturing and non-manufacturing subsamples

Dependent variable Model	ROA 12 1989–1997	ROA 13 1998–2007	ROA 14 Manufacturing firms	ROA 15 Non-manufacturing firms	ROA 16 Manufacturing firms (best fit)
Internationalization	−0.73	−0.52	−0.31	0.11	−0.08
	(0.94)	(0.49)	(0.17)	(0.09)	(0.04)
	[0.44]	[0.29]	[0.08]	[0.23]	[0.06]
Internationalization squared	0.79	0.90	0.55	−0.10	0.12
	(1.75)	(1.47)	(0.22)	(0.10)	(0.05)
	[0.65]	[0.54]	[0.01]	[0.32]	[0.02]
Internationalization cubed	−0.32	−0.50	−0.22	−0.06	
	(1.10)	(1.36)	(0.29)	(0.15)	
	[0.77]	[0.66]	[0.46]	[0.68]	
Internationalization × R&D intensity	0.32	0.24	0.22	0.15	
	(1.10)	(0.33)	(0.52)	(0.57)	
	[0.77]	[0.47]	[0.67]	[0.80]	
Parent R&D intensity	−0.01	−0.01	−0.01	−0.01	0.01
	(0.01)	(0.01)	(0.02)	(0.02)	(0.00)
	[0.24]	[0.46]	[0.56]	[0.58]	[0.07]
Parent size	0.01	0.01	0.01	0.01	0.01
	(0.01)	(0.01)	(0.10)	(0.02)	(0.00)
	[0.00]	[0.08]	[0.00]	[0.67]	[0.00]
Parent debt-to-equity ratio	−0.01	−0.01	0.01	−0.01	0.01
	(0.01)	(0.06)	(0.11)	(0.02)	(0.20)
	[0.08]	[0.88]	[0.93]	[0.54]	[0.96]
Parent export intensity	0.01	0.01	0.02	0.01	0.03
	(0.02)	(0.01)	(0.01)	(0.06)	(0.01)
	[0.53]	[0.28]	[0.16]	[0.86]	[0.05]
Parent product diversification	0.01	−0.01	−0.01	−0.01	−0.01
	(0.02)	(0.01)	(0.05)	(0.03)	(0.04)
	[0.61]	[0.23]	[0.85]	[0.75]	[0.82]
Exchange rate	−0.16	−0.11	−0.14	−0.08	−0.14
	(0.02)	(0.02)	(0.02)	(0.03)	(0.02)
	[0.00]	[0.00]	[0.00]	[0.00]	[0.00]
Intercept	0.22	0.32	0.27	0.21	0.27
	(0.12)	(0.09)	(0.06)	(0.09)	(0.06)
	[0.00]	[0.00]	[0.00]	[0.02]	[0.00]
N	10,002	11,295	9,482	11,815	9,482
R square	0.07	0.08	0.11	0.04	0.11
Wald Chi square	3.48	2.79	10.51	2.81	12.17
Fixed or random effects	Fixed	Fixed	Fixed	Fixed	Fixed

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

that for firms in our sample *ROA* falls till they achieve *Internationalization* of 0.33 (1.2 standard deviations above the sample mean) and rises subsequently. This threshold value corresponds to the 89th percentile of our sample, i.e. only 11 percent of our observations lie in the range where multinationality has a positive relationship with performance.

Second, we compare the economic magnitude of the predicted effects on *ROA*. Looking at their Tobin's Q results, L&B report that Tobin's Q drops by about 0.35 from the base case to the first threshold value, rises by 0.5 (a net increase of 0.15 relative to the base case) to the second threshold value, and falls thereafter (Lu and Beamish, 2004: 605). While they do not report similar numbers for the *ROA* results, a parallel calculation, based on Model 5 in their table 2 (Model 0 in Table 4 shown above), shows a predicted drop in *ROA* by about 6.2 percentage points (corresponding to about 1.42 standard deviations in sample *ROA*) from the base case of no multinationality to the first threshold value, followed by an increase of about 0.3 percentage points (0.05 standard deviations) to the second threshold value (meaning a net decrease

DOI: 10.1002/smj

of 5.9 percentage points relative to the base case), with *ROA* declining steadily thereafter.[15] The corresponding values from Model 16 are a predicted drop in *ROA* by 1.33 percentage points (0.08 standard deviations) below the base case to the threshold value of *Internationalization*, followed by a steady rise in *ROA*, reaching a maximum of 4 percentage points (0.25 standard deviations) above the base case of no multinationality, with multinationality having a net positive effect on performance for values of *Internationalization* greater than 0.67.

Finally, Table 7 replicates the robustness checks reported by L&B.[16] Models 17–20 replicate their tests using lags of two and three years between dependent and independent variables, and confirm that we fail to replicate the S-curve and the interaction between *Internationalization* and *R&D intensity* even with these longer lags. These models do continue to show a marginally significant U-shaped relationship between multinationality and performance for our manufacturing subsample, with p-values of 0.12 and 0.09 for *Internationalization*, and 0.07 and 0.01 for *Internationalization Squared* in Models 18 and 20 respectively.

Models 21 and 22 then re-run our results using mean-centered variables to account for collinearity (as L&B report they do in their article) for the full sample and the manufacturing subsample respectively. Our finding of a U-shaped relationship is statistically stronger in the manufacturing subsample once this correction is applied, with Model 22 showing p-values of 0.04 and 0.01 for *Internationalization* and *Internationalization Squared* respectively, but we still do not replicate L&B's other findings.

Finally, Model 23 replicates L&B's attempt to use a 2SLS model to account for the endogeneity of multinationality. We focus our replication here exclusively on the manufacturing subsample, because the two instruments we use—*Industry foreign knowledge* and *Industry shipping value by weight*—are only meaningful for manufacturing industries, and because we are primarily concerned with testing the robustness of findings that have proven to be significant thus far. For similar reasons, we also limit this analysis to testing the effect of *Internationalization* and *Internationalization squared* terms, i.e. of confirming the robustness of the U-shaped relationship observed in Model 16. Model 23 reports the results of the second stage of the 2SLS model (the first stage regressions, with *Industry Foreign Knowledge* instrumenting for *Internationalization* and *Industry shipping value by weight* for *Internationalization squared* are not shown, but available upon request) and shows that the U-shaped relationship in Model 16 disappears once we account for the endogeneity of multinationality. Both *Internationalization* and *Internationalization squared* have coefficients within one standard error of zero in Model 23, implying that there is no evidence of a truly causal effect of multinationality on performance in our models. Overall, Table 7 confirms our inability to replicate L&B's findings of an S-curve and a moderating effect of intangible assets, while further showing that we are also unable to replicate the consistency of results across various robustness checks that they found. It also confirms the absence of a causal effect of aggregate multinationality on performance, with the one statistically significant result we do find (in our manufacturing subsample) ceasing to be significant once we account for endogeneity.

CONCLUSION AND DISCUSSION

Our empirical results fail to replicate the relationship between multinationality and performance found by Lu and Beamish (2004). We see no evidence for an S-curve in our sample of US multinationals from 1989 to 2007, with the coefficient of the *Internationalization cubed* term never attaining statistical significance in any of our models. Nor do we find any evidence of a significant moderating effect of *R&D intensity* on the relationship between multinationality and performance. We do find some evidence of a U-shaped association between multinationality and performance in our manufacturing industry subsample, with this relationship being statistically significant at conventional levels in models that use fixed-effects and account for collinearity by using mean centered variables, but this result is not especially robust, and, importantly, disappears once we control for the endogeneity of multinationality. Our results thus suggest that while there may be a

[15] Again, these calculations are for a firm with R&D intensity of three percent. For a firm with R&D intensity of one percent (the mean value in L&B's sample), ROA consistently declines with internationalization, since the threshold values of the S-curve are imaginary.

[16] Since L&B do not report the actual results of these tests in their article, simply mentioning that they have run these tests and confirmed robustness, we are unable to compare coefficient estimates for our replications of these tests to their original estimates.

DOI: 10.1002/smj

Table 7. Replicating robustness checks

Dependent variable	ROA	ROA	ROA	ROA	ROA	ROA	ROA
Model	17	18	19	20	21	22	23
	Two year lag full sample	Two year lag manufacturing sample	Three year lag full sample	Three year lag manufacturing sample	Mean-centered full sample	Mean-centered manufacturing sample	2SLS - manufacturing sample
Internationalization	−0.12	−0.06	−0.14	−0.07	−0.26	−0.22	0.09
	(0.50)	(0.04)	(0.61)	(0.04)	(0.21)	(0.11)	(0.43)
	[0.81]	[0.12]	[0.81]	[0.09]	[0.23]	[0.04]	[0.83]
Internationalization squared	0.19	0.32	0.62	0.31	0.73	0.48	−0.27
	(0.68)	(0.13)	(2.58)	(0.13)	(0.68)	(0.20)	(0.69)
	[0.78]	[0.07]	[0.81]	[0.01]	[0.28]	[0.01]	[0.70]
Internationalization cubed	0.21	−0.19	−0.20	−0.18	−0.46	−0.22	
	(0.33)	(0.40)	(0.85)	(0.36)	(0.62)	(0.32)	
	[0.70]	[0.63]	[0.81]	[0.62]	[0.46]	[0.67]	
Internationalization × R&D intensity	0.36	0.28	0.34	0.29	0.26	0.22	
	(0.71)	(0.36)	(1.70)	(0.35)	(0.70)	(0.32)	
	[0.58]	[0.44]	[0.84]	[0.40]	[0.71]	[0.68]	
Parent R&D intensity	−0.01	−0.01	−0.01	0.01	−0.01	−0.01	0.02
	(0.02)	(0.05)	(0.02)	(0.02)	(0.05)	(0.02)	(0.01)
	[0.68]	[0.83]	[0.60]	[0.67]	[0.83]	[0.67]	[0.13]
Parent size	0.01	0.01	0.01	0.01	0.01	0.01	0.04
	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.00)	(0.01)
	[0.01]	[0.00]	[0.00]	[0.00]	[0.01]	[0.00]	[0.00]
Parent debt-to-equity ratio	0.01	−0.01	0.01	0.01	−0.01	0.01	−0.01
	(0.03)	(0.02)	(0.59)	(0.02)	(0.02)	(0.11)	(0.05)
	[0.73]	[0.60]	[0.09]	[0.60]	[0.72]	[0.93]	[0.86]
Parent export intensity	0.01	0.01	0.01	0.01	0.01	0.01	−0.02
	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.01)	(0.03)
	[0.22]	[0.50]	[0.17]	[0.46]	[0.32]	[0.16]	[0.54]
Parent product diversification	0.01	0.01	0.01	0.01	−0.01	−0.01	−0.01
	(0.01)	(0.03)	(0.00)	(0.11)	(0.02)	(0.11)	(0.03)
	[0.12]	[0.72]	[0.07]	[0.72]	[0.64]	[0.93]	[0.72]
Exchange rate	−0.01	−0.04	0.03	−0.04	−0.12	−0.14	−0.18
	(0.01)	(0.05)	(0.02)	(0.01)	(0.01)	(0.02)	(0.04)
	[0.21]	[0.01]	[0.17]	[0.01]	[0.01]	[0.01]	[0.01]
Intercept	0.22	0.24	0.12	0.27	0.21	0.27	
	(0.05)	(0.05)	(0.07)	(0.06)	(0.05)	(0.06)	
	[0.00]	[0.00]	[0.02]	[0.00]	[0.00]	[0.00]	
N	19,962	9,066	18,620	8,631	21,297	9,482	9,482
R square	0.06	0.08	0.06	0.08	0.10	0.11	
Wald Chi square or F (for 2SLS)	3.21	4.79	2.84	3.79	8.39	8.18	13.97
Kleinbergen-Paap rk Wald F statistic							10.51
Fixed or random effects	Fixed	Fixed	Fixed	Fixed	Fixed	Fixed	Fixed

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

DOI: 10.1002/smj

U-shaped relationship between multinationality and performance, at least for manufacturing firms, this is an association rather than a causal relationship. Note, moreover, that our finding of a U-shaped relationship between multinationality and performance in a sample of large, well-established multinationals is inconsistent with L&B's argument that the U-shaped relationship will apply to newly internationalizing firms only (Lu and Beamish, 2001, 2004).

What explains these differences in findings? While our answer to that question must necessarily be speculative, one possibility is that these differences reflect the greater diversity and experience of the U.S. MNCs in our sample. As L&B point out, the internationalization of Japanese firms between the mid-1980s and late-1990s that they study marked a period of rapid expansion by a set of firms that had only recently gone international. In contrast, many of the US multinationals in our sample have a long history of foreign operations, a fact reflected in both the higher average level and greater dispersion of multinationality in our sample compared to that in L&B, as shown in Table 3 above. It may be, therefore, that firms in L&B's sample were more apt to over- or under-internationalize, with the less informed choices that these firms made having a causal effect on their performance that they had not predicted. In contrast, the firms in our sample may be doing a better job of systematically choosing the appropriate level of multinationality given their resources and capabilities, so that we see neither the kind of geographic overreaching that results in the third phase of the S-curve, nor a causal effect of multinationality once the endogeneity of their choices is accounted for. This explanation for the discrepancy in our findings seems even more likely when we consider that the third phase of L&B's S-curve only begins at very high values of multinationality, suggesting that this result likely reflects the performance of a small handful of the Japanese firms in L&B's sample that were far more internationally diversified than their peers.

Whatever its cause, the discrepancy between our findings and those of L&B suggests the frailty of the S-curve relationship, and emphasizes the need for caution in generalizing their finding to other contexts. Our study thus contributes to empirical research on the multinationality-performance relationship by showing that the results from one of the most widely cited studies on this relationship may not replicate outside its original context. In doing so, moreover, we highlight the importance of accounting for the endogeneity of multinationality while estimating its effect on performance, with our replication of the endogeneity correction attempted by L&B finding no evidence of a causal effect once we account for the firm's choice of multinationality. In addition, our study also suggests the need to more carefully consider the relevant range of observations in non-linear models, with non-linear effects that only hold for extreme values of observations within the sample being especially unlikely to prove robust.

The results of our attempts to account for the endogeneity of multinationality are interesting not only because they replicate the robustness checks conducted by L&B, but because they speak to a growing body of work that has questioned the theoretical validity of a causal effect of multinationality on performance (Hennart, 2011; Shaver, 1998; Verbeke and Forootan, 2012). This work argues that since multinationality is a strategic choice, we should not expect to see any effect of multinationality on performance, except in cases where firms make a mistake in the level of multinationality that they choose (Hennart, 2011). The results of our endogeneity-corrected analysis provide preliminary support for this argument, showing that the relationship between multinationality and performance disappears once we account for the endogeneity of multinationality. More generally, these results, coupled with our finding that even the association between multinationality and performance does not replicate across countries, throw the very quest for a generalizable aggregate effect of multinationality on performance open to debate. It may be that the study of global strategy is better served by attempts to look at the effect of specific types of multinational activity on firm performance (Wiersema and Bowen, 2011), and the ways in which these effects are contingent upon country contexts, rather than continuing to look for a universal relationship between multinationality and performance.

As a replication, our study is subject to several limitations. First, while we have attempted to replicate the methods and measures used by L&B as faithfully as possible, we are unable to replicate their analysis completely, given the differences between our secondary data sources and theirs. In particular, because our sample includes both public and private firms, and cannot be directly matched to stock market information, we are unable

DOI: 10.1002/smj

to replicate L&B's study with market performance as a dependent variable. Though we do not find an S-shaped relationship between multinationality and *ROA*, therefore, it is possible that such a relationship may exist between multinationality and Tobin's Q. Even if this were true, however, it would only mean that the S-curve relationship was extremely susceptible to the choice of dependent variable, and would not detract from our overall finding of a failure to replicate. Second, it is true that, just like the L&B study we replicate, our empirical analysis is limited to a single country (the United States), and only shows that the S-curve does not replicate to that home country context. It may be that the U.S. is an outlier in this regard—only additional replications in other country contexts can tell us how truly generalizable the S-curve hypothesis is (or is not), and we hope that our study will pave the way for other such replications. Finally, being a replication, our study suffers from many of the same limitations that plagued the original; specifically, it uses highly aggregated measures of both multinationality and performance (Wiersema and Bowen, 2011). Future work should look at the relationship between a firm's multinational activities and its performance at a more disaggregated level, including more consideration for the subsidiary and value chain activity levels of analysis (Berry, 2015; Gong, 2003). Future work could also look for more sophisticated ways of accounting for the endogeneity of multinationality, perhaps relying on changes in government policies or other natural experiments. The instrumental variable strategy we use to replicate the robustness check from prior work, while valid, is certainly far from perfect.

To conclude, we replicate Lu and Beamish's (2004) study of the relationship between multinationality and performance using an 18-year panel that covers the population of U.S. MNCs, but find no evidence of either the S-shaped relationship between multinationality and performance, or the moderating effect of intangible assets, that they found. While we do see some evidence for a U-shaped association between multinationality and performance for our manufacturing firm subsample, this finding disappears once we account for the endogeneity of multinationality, suggesting that there is no true causal effect of multinationality on performance in our sample. Our study thus contributes to a substantial body of empirical research on the multinationality-performance relationship (Kirca *et al.*, 2011) highlighting the need for caution when generalizing results from single country studies, and stressing the importance of accounting for the endogeneity of multinationality (Verbeke and Forootan, 2012) when assessing its effect on performance.

ACKNOWLEDGEMENTS

The statistical analysis of firm-level data on U.S. multinational companies was conducted at the Bureau of Economic Analysis (BEA) U.S. Department of Commerce under arrangements that maintain legal confidentiality requirements. Views expressed in this article do not reflect those of the BEA or the Department of Commerce. We thank Special Issue editors Rich Bettis, Connie Helfat and Myles Shaver, as well as two anonymous reviewers, for their comments and suggestions. We appreciate helpful comments from Raymond Mataloni and William Zeile Zeile on the BEA data, and are grateful to participants at the Academy of Management, Academy of International Business, and Strategic Management Society conferences, as well as the Applied Economics Seminar at the University of Minnesota for their feedback on earlier versions of this article. Berry thanks and acknowledges Mack Institute for Innovation Management funding for this research.

REFERENCES

Berry H. 2013. When do firms divest global operations? *Organization Science* **24**(1): 241–261.

Berry H. 2014. Global integration and innovation: multi-country knowledge generation within MNCs. *Strategic Management Journal* **35**(6): 869–890.

Berry H. 2015. Knowledge inheritance in global industries: the impact of parent firm knowledge on the performance of foreign subsidiaries. *Academy of Management Journal* **58**(5): 1438–1458.

Berry H, Kaul A. 2015. Global sourcing and foreign knowledge seeking. *Management Science* **61**(5): 1052–1071.

Capar N, Kotabe M. 2003. The relationship between international diversification and performance in service firms. *Journal of International Business Studies* **34**(4): 345–355.

Cardinal LB, Miller CC, Palich LE. 2011. Breaking the cycle of iteration: forensic failures of international diversification and firm performance research. *Global Strategy Journal* **1**(1-2): 175–186.

Chung W, Yeaple S. 2008. International knowledge sourcing: evidence from US firms expanding abroad. *Strategic Management Journal* **29**(11): 1207–1224.

Contractor FJ. 2012. Why do multinational firms exist? A theory note about the effect of multinational expansion on performance and recent methodological critiques. *Global Strategy Journal* **2**(4): 318–331.

Contractor FJ, Kundu SK, Hsu CC. 2003. A three-stage theory of international expansion: the link between multinationality and performance in the service sector. *Journal of International Business Studies* **34**(1): 5–18.

Geringer JM, Beamish PW, Da Costa RC. 1989. Diversification strategy and internationalization: implications for MNE performance. *Strategic Management Journal* **10**(2): 109–119.

Gomes L, Ramaswamy K. 1999. An empirical examination of the form of the relationship between multinationality and performance. *Journal of International Business Studies* **30**(1): 173–183.

Gong Y. 2003. Subsidiary staffing in international enterprises, agency, resources and performance. *Academy of Management Journal* **46**(6): 728–739.

Grant RM, Jamine AP, Thomas H. 1988. Diversity, diversification, and profitability among British manufacturing companies, 1972–1984. *Academy of Management Journal* **31**(4): 771–781.

Hennart JF. 2011. A theoretical assessment of the empirical literature on the impact of multinationality on performance. *Global Strategy Journal* **1**(1-2): 135–151.

Hitt MA, Hoskisson RE, Kim H. 1997. International diversification: effects on innovation and firm performance in product-diversified firms. *Academy of Management Journal* **40**(4): 767–777.

Hitt MA, Tihanyi L, Miller T, Connelly B. 2006. International diversification: antecedents, outcomes, and moderators. *Journal of Management* **32**(6): 831–867.

Kirca AH, Hult GTM, Roth K, Cavusgil ST, Perry MZ, Akdeniz MB, Deligonul SZ, Mena JA, Pollitte WA, Hoppner JJ, Miller JC, White RC. 2011. Firm-specific assets, multinationality, and financial performance: a meta-analytic review and theoretical integration. *Academy of Management Journal* **54**(1): 47–72.

Kobrin S. 1991. An empirical analysis of the determinants of global integration. *Strategic Management Journal* **12**(S1): 17–31.

Lu JW, Beamish PW. 2001. The internationalization and performance of SMEs. *Strategic Management Journal* **22**(6-7): 565–575.

Lu JW, Beamish PW. 2004. International diversification and firm performance: the S-curve hypothesis. *Academy of Management Journal* **47**(4): 598–608.

Mataloni R, Yorgason D. 2006. Operations of US multinational companies: preliminary results from the 2004 benchmark survey. *Survey of Current Business* **86**: 37–68.

Nachum L, Zaheer S. 2005. The persistence of distance? The impact of technology on MNE motivations for foreign investments. *Strategic Management Journal* **26**(8): 747–767.

Pierce JR, Schott PK. 2009. Concording U.S. harmonized system categories over time. NBER Working paper 14837. National Bureau of Economic Research, Cambridge, MA.

Qian G, Khoury TA, Peng MW, Qian Z. 2010. The performance implications of intra- and inter-regional geographic diversification. *Strategic Management Journal* **31**(9): 1018–1030.

Ruigrok W, Amann W, Wagner H. 2007. The internationalization-performance relationship at Swiss firms: a test of the S-shape and extreme degrees of internationalization. *Management International Review* **47**(3): 349–368.

Shaver JM. 1998. Accounting for endogeneity when assessing strategy performance: does entry mode choice affect FDI survival? *Management Science* **44**(4): 571–585.

Silverman BS. 1996. Technological assets and the logic of corporate diversification. PhD diss., University of California at Berkeley, Berkeley, CA.

Verbeke A, Forootan MZ. 2012. How good are multinationality-performance (M-P) empirical studies? *Global Strategy Journal* **2**(4): 332–344.

Vernon R. 1966. International investment and international trade in the product cycle. *Quarterly Journal of Economics* **80**(2): 190–207.

Wiersema MF, Bowen HP. 2011. The relationship between international diversification and firm performance: why it remains a puzzle. *Global Strategy Journal* **1**(1-2): 152–170.

DOI: 10.1002/smj