Received: 5 September 2014 | Revised: 17 August 2020 | Accepted: 26 August 2020 | Published on: 23 September 2020

DOI: 10.1002/smj.3236

STRATEGIC MANAGEMENT JOURNAL WILEY

RESEARCH ARTICLE

Using the Shapley Value approach to variance decomposition in strategy research: Diversification, internationalization, and corporate group effects on affiliate profitability

Dmitry Sharapov[1] | **Paul Kattuman**[2] | **Diego Rodriguez**[3] | **F. Javier Velazquez**[3]

[1]Imperial College Business School, London, UK

[2]Cambridge Judge Business School, University of Cambridge, Cambridge, UK

[3]Facultad de Ciencias Económicas y Empresariales, Universidad Complutense de Madrid and GRIPICO, Madrid, Spain

Correspondence
Dmitry Sharapov, Imperial College Business School, Tanaka Building, South Kensington Campus, London SW7 2AZ, UK.
Email: dmitry.sharapov@imperial.ac.uk

Abstract

Research Summary: Variance decomposition methods allow strategy scholars to identify key sources of heterogeneity in firm performance. However, most extant approaches produce estimates that depend on the order in which sources are considered, the ways they are nested, and which sources are treated as fixed or random effects. In this paper, we propose the use of an axiomatically justified, unique, and effective solution to this limitation: the "Shapley Value" approach. We show its effectiveness compared to extant methods using both simulated and real data, and use it to explore how the importance of business group effects varies with group diversification and internationalization in a large, representative sample of European firms. We thus demonstrate the method's superior accuracy and its usefulness in asking and answering new questions.

Managerial Summary: A key contribution of strategic management research to managerial practice is identifying drivers of firm performance that operate at firm, corporation, industry, and national levels. A branch of this research measures the relative importance of factors at these different levels in producing variation in firm performance, thus helping top managers focus efforts on aspects of their businesses most likely to yield performance differences. However, estimates produced by extant methods are sensitive to method used, and to modeling choices made. This paper proposes the use of the "Shapley Value" approach, which is free from such sensitivity, shows its effectiveness compared to extant methods, and uses it to explore how the importance of factors at the level of the business group varies with group diversification and internationalization.

KEYWORDS

corporate group, diversification, internationalization, Shapley value regression, variance decomposition

Variance decomposition methods have been vital in research on whether sources of heterogeneity in firm performance reside at the business unit, corporation, or industry level (e.g., Guo, 2017; McGahan & Porter, 1997; Misangyi, Elms, Greckhamer, & Lepine, 2006; Rumelt, 1991; Schmalensee, 1985). More recently, these methods have also proven useful in evaluating whether a range of other influences on firm performance actually "matter" in explaining its variation across firms, including country and regional effects (Chan, Makino, & Isobe, 2010; Ma, Tong, & Fitza, 2013; Makino, Isobe, & Chan, 2004; McGahan & Victer, 2010), ownership (Fitza & Tihaniy, 2018), and Chief Executive Officers (Crossland & Hambrick, 2007; Fitza, 2017; Quigley & Graffin, 2017).

However, while the methods used for variance decomposition have been improved in a number of ways (for an overview, see Guo, 2017, pp. 1328–1330), most extant approaches share an important limitation: unless the effects under study are orthogonal, the estimates are sensitive to choices regarding the order in which the effects are introduced into the models; which effects are treated as fixed versus random; and which effects are considered to be nested in others.[1] This implies that these methods produce estimates of the share of variance accounted for by different effects that may be lower- or upper-bound estimates, or anywhere in between, depending on the above choices.

In this paper, we draw on the statistical literature (Grömping, 2007; Pintér, 2011; Young, 1985) to propose an axiomatically justified, unique, and effective solution to this limitation: the Shapley Value approach. The Shapley Value for a given effect is its contribution to model explanatory power, averaged (with weights) over all possible sequential orders in which the effects could be introduced into the regression model. In the following sections, we introduce the Shapley Value method, before showing its effectiveness compared to currently used

[1]While random effects variance decomposition methods, which we will refer to as Variance Components Analysis (VCA), do not share this order-dependence limitation when used to estimate models without a nesting structure, this is because they make strong assumptions regarding effect distributions. We provide a further discussion of the VCA approach and compare the results that it generates to those produced by other methods below.

approaches, using both a simulation and an empirical application. The latter employs the Shapley Value approach to explore how the importance of corporate group effects changes depending on the extent of the group's diversification and internationalization, using data on a large and representative sample of European firms. We conclude by discussing how the Shapley Value method could be used in future research, both by improving the reliability of the evidence in long-running debates regarding the sources of heterogeneity in firm performance, as well as by allowing strategy scholars to ask new questions regarding the importance of different effects under different conditions (e.g., Arora et al., 2016).

1 | THE SHAPLEY VALUE METHOD

A fundamental question in strategic management research is why some businesses are successful while others are not. The approach taken in a continuing literature on business performance variation that started with Schmalensee (1985) is based on the view that the key first step is to identify the levels at which important, performance-relevant factors operate. If for example, controlling for all else, performance differences between *businesses* are notably greater than performance differences between *corporate groups, industries*, or *countries* that they belong to, then focusing attention on understanding what business-level features and strategies mark out high performing businesses from poor performers would be productive. As in that literature, the focus of this paper is on estimating the relative *dispersion* importances of the levels at which factors affecting performance operate. The *dispersion* importance of a factor can be estimated based on the extent to which it accounts for *variation* in business performance in a regression model.[2] If the regressor sets were all orthogonal to each other, then the overall variance in profitability will decompose exactly between the regressor sets and yield unique estimates of their dispersion importances. But corporate groups commonly span more than one industry, and more than one country, while industries span all countries. Due to this cross-nesting, firm profitability will not decompose exactly or uniquely between these levels.

The ANOVA and variance components analysis (VCA) approaches used in the early studies (Rumelt, 1991; Schmalensee, 1985) and widely adopted since (e.g., Hawawini, Subramanian, & Verdin, 2003; McGahan & Porter, 1997, 2002), have acknowledged limitations in dealing with this problem. The fixed-effects ANOVA approach assigns the covariance between effects to the effect introduced first into the specification, along with a share of the unique contribution of any omitted but correlated effect to model explanatory power (Grömping, 2007, see Online Appendix S1 for details). Authors have tended to address the resulting indeterminacy by presenting estimates from a number of different paths from the null to the full model, but have generally not reconciled results from different model paths in a consistent manner. Estimated effect contributions thus remain sensitive to the choice of model paths presented.

While VCA does not share the order-dependence limitation of other methods and does allow for covariances between effects if these are specified in the model, these covariances are assumed to be random. The appropriateness of this strong assumption has been criticized

[2]This notion of relative *dispersion* importance is different from the more common notion of the *level* importance of a factor, which relates to the extent to which the factor accounts for the *expected value* of performance. It is thus possible for a regressor to have high *level* importance, by being influential in determining the expected value of performance, while having low *dispersion* importance, if it does not vary much within the population being studied. We thank the editor for this point.

(Guo, 2017), as has the method's lack of power in finding small but significant effects (Brush & Bromiley, 1997; Hough, 2006). The critique relating to the sum of squares procedure used to estimate the variance components in earlier papers has been overcome through advances in estimation methods. Recent work has taken a random effects approach to variance decomposition using the same maximum likelihood or restricted maximum likelihood estimation techniques as multilevel models (see below), without nesting some effects in others or using fixed effects (Hough, 2006; Marchenko, 2006). However, the reliance on strong assumptions regarding independence and joint normality of the random effects remains.

Alternative approaches have also been used, including simultaneous equation modeling (Brush, Bromiley, & Hendrickx, 1999), nonparametric estimation (Ruefli & Wiggins, 2003), and multilevel modeling (Guo, 2017; Hough, 2006; Misangyi et al., 2006). These are not exempt from criticism (Guo, 2017; Hough, 2006; McGahan & Porter, 2005). The multilevel approach appears to be the most promising, as it explicitly takes the cross-nested structure of variation in firm performance into account. Examples include observations across time being nested within firms, and firms in turn being cross-nested within both business groups and industries (Misangyi et al., 2006), or firms being nested within corporate groups (Majumdar & Bhattacharjee, 2014). Thus, this approach allows for the estimation of random effects variance components like the VCA method but allows for general multilevel error structures in deriving more accurate variance component estimates (Guo, 2017; Hough, 2006). It can produce estimates of the dispersion importance of both random and fixed effects. However the estimates produced by a cross-nested multilevel approach will also depend on choices of which effects are considered to be nested in others, and which cross-nesting effects (e.g., industry or corporate group) are treated as being random versus fixed (see Misangyi et al., 2006, pp. 580–581).[3]

In this paper, we focus on a fully rationalized and unique solution to the general problem of fairly allocating model $\bar{R}^2$ to regressors that may be correlated.[4] The conceptual basis of the Shapley Value allocation, as this approach is called in the statistical literature, lies in interpreting the regression as a cooperative game in which the regressors are players and the model $\bar{R}^2$ is the collective *value* that is to be divided fairly among the regressor-players (Pintér, 2011). The underlying principle, first formulated in transferable utility co-operative game theory as a solution to the problem of sharing gains from cooperation among members of a coalition (Shapley, 1953), is to allocate to each member in the coalition, the marginal contributions that she can make to its joint output. If there is any degree of similarity between coalition members in their ability to contribute to output, then any member's marginal contribution will be sensitive to the set of other members who have already contributed. The Shapley Value method neutralizes this dependence by allocating to each member the expectation of her marginal contribution, taken over all possible sequences in which she can contribute to the coalition. This logic translates in a straightforward way to regression-based decomposition of the variation in firm performance among different (groups of) variables. The Shapley Value allocation of dispersion importance of regressors is general and has been shown to be the only allocation that

[3]While Misangyi et al. (2006) justify their choices based on the guidance provided by the methodological literature, which states that the cross-nested effect with a larger number of observations be treated as random, and the one with a smaller number of observations as fixed, they accept that there is no theoretical justification for this choice. It is also easy to see how following the methodological guidance could lead to different choices across studies, depending on sample properties and the level of granularity at which certain effects, for example, industry, are classified.

[4]When sets of *effects* are involved, $\bar{R}^2$ (adjusted-R^2) corrects R^2 for the degrees of freedom and is the estimator of explained variance that is to be divided up between regressor sets.

meets the axiomatic requirements for a proper decomposition (Grömping, 2007).[5] Online Appendix S1 presents the technical details of the approach and estimator properties.

We now compare the performance of the Shapley Value approach and alternative methods. For this purpose, we first use a simulation with a known data generating process. Next, we proceed to a large-sample empirical application. The latter seeks to answer two questions: how much corporate group effects matter in explaining variance in the profitability of European firms; and of how their importance is affected by the extent of corporate group diversification and internationalization.

2 | SIMULATION

To evaluate the performance of the Shapley Value approach compared to the ANOVA, multilevel, and VCA methods, we use a Monte Carlo simulation. We generate a simulated data set with industry, corporate group, firm, and year effects with defined variance–covariance structures, which together determine the profitabilities of the simulated firms. We then apply the Shapley Value approach, alongside the alternative methods, and compare the results with the known data generating process. Our simulated full model is specified as:

$$\pi_{igt} = \mu + \alpha_i + \beta_g + \gamma_t + \phi_{ig} + \varepsilon_{igt}$$

where π_{igt} is the profitability of corporate group g's business unit in industry i at time t, μ is the overall average profitability, α_i is the profitability component characterizing industry i, and β_g is the profitability component characterizing corporate group g. As described in more detail in Online Appendix S2, these effects are constructed to be correlated with each other. A firm belonging to corporate group g and operating in industry i has the profitability component ϕ_{ig}, itself constructed to be independent of the corporate group and industry effects. γ_t are year-specific profitability components, and ε_{igt} are normally distributed error terms, uncorrelated with any of the other effects. We compare the proportions of total variance allocated to the effects by the Shapley Value approach as well as alternative methods, against the true values following from the parameters used in the data generating process described in Online Appendix S2.

We examine estimates from the ANOVA, multilevel, VCA, and Shapley Value approaches. For the ANOVA approach, we examine two paths from null to full model (in the spirit of McGahan & Porter, 1997), introducing the effects in the orders: (a) year, industry, corporate group, firm; and (b) year, corporate group, industry, firm. For industry and corporate group effects, we present the ANOVA estimates from both paths as upper and lower bound estimates. For the multilevel approach, we follow the Hierarchical Linear Modeling (HLM) approach of

[5]Pintér (2011), following Young (1985), proved that in regression games, the solution concept satisfies the following three essential requirements of dispersion importance estimators, *if and only if* it is the Shapley Value solution:

1. *Efficiency*: The full model $\bar{R}^2$ must be decomposed exactly among the regressor variables.
2. *Equal treatment*: If two regressors are equivalent in the sense that the full model $\bar{R}^2$ is unchanged regardless of which the two are included in the model, then their dispersion importances must be equal.
3. *Monotonicity*: In comparing two regression models, if a regressor variable contributes more to the explanatory power of the first model than to the explanatory power of the second, then its dispersion importance must be higher in the first model than in the second.

[Correction added on 09 October 2020, after first online publication: the word 'Online' has been inserted before the 'Appendix S1' and 'Appendix S2' citations all throughout the article.]

Misangyi et al. (2006) but provide estimates for models both when corporate group effects are treated as being random and industry effects as fixed, and vice-versa. The VCA results are produced by estimating a crossed-effects model (one where no effect is considered to be nested in another) using Stata's **xtmixed** command. For the Shapley Value approach, we find the weighted average of the contribution of each set of effects to explaining model $\bar{R}^2$ over all possible orders in which the effects can be introduced.

The results based on simulated data for 1,000 firms belonging to 500 corporate groups operating in 250 industries over 4 years, with each corporate group operating in two industries, are presented in Online Appendix S2. It can be seen that the Shapley Value approach provides more accurate estimates of effect contribution to variance in profitability than commonly used alternative methods and is the appropriate method for use in variance decomposition.

3 | EMPIRICAL APPLICATION

Research has shown that being an affiliate of a corporate group can have a significant bearing on firm performance (Almeida & Wolfenzon, 2006; Belenzon & Berkovitz, 2010; Bertrand, Mehta, & Mullainathan, 2002; Carney, Gedajlovic, Heugens, Van Essen, & Van Oosterhout, 2011; Chang & Hong, 2000).[6] A meta-analysis of 141 studies found the relationship between group affiliation and firm performance to be negative and significant on average, with institutional factors and strategic actions at firm and group levels playing important roles in this relationship (Carney et al., 2011). While such meta-analytic estimates are useful to understand the level importance of group affiliation as a driver of firm performance, it is also necessary to consider the dispersion importance of corporate group effects that is *independent* of the variation in other performance drivers at the firm, industry, and country levels. The empirical application of the Shapley Value method that follows demonstrates its usefulness in providing more precise estimates of the importance of diverse sources of performance variation, compared to previously used variance decomposition methods. It also highlights the potential of Shapley Value approach to ask and answer new questions that contribute to our understanding of the implications of corporate group membership for affiliates. We illustrate this by examining how the relative dispersion importance of corporate group effects in explaining firm performance *varies* with the extent of diversification and internationalization of the group.

Diversification and internationalization will be germane to how much corporate groups matter in explaining the variance in affiliate performance. How germane, will depend on the extent to which the potentially homogenizing corporate group influence on group members (e.g., Barney, 1997; Bowman & Helfat, 2001; Caves, 1996; Gulati, Nohria, & Zaheer, 2000; Hitt, Hoskisson, & Kim, 1997; Kali & Sarkar, 2011), will be realized in practice, given the likely costs and difficulties accompanying diversification and internationalization (e.g., Chari, Devaraj, & David, 2008; Crossland & Hambrick, 2007; Hashai, 2015; Levinthal & Wu, 2010; Lu & Beamish, 2004; Nohria & Ghoshal, 1994; Rawley, 2010; Zhou, 2011). To see which influence dominates, we apply the Shapley Value method and compare the resulting estimates of firm, corporate group, industry, country, and year effects against those produced by ANOVA, HLM, and VCA methods. We then use the approach to analyze how the extent of diversification and

[6]While some prior work has used the terminology of "business groups" or "family business groups," we follow Belenzon, Berkovitz, and Rios (2013) in using the concept of "corporate groups," which is codified in European legal, cultural, and economic institutions, and in which groups are defined through equity-ownership ties.

internationalization determines the importance of corporate group effects in explaining variation in firm performance.

3.1 | Data and methods

We chose a setting that enables us to compare the estimates of corporate group effects produced by the Shapley Value approach against those produced by other methods: one that accommodates diversification both within and across corporate group affiliates, as well as corporate group internationalization. Specifically, we study the population of non-financial firms across 25 European countries using the complete version of the Amadeus database maintained by Bureau van Dijk. The database contains balance sheet information and additional data for about 14 million European firms. For comparability with previous studies, we use return on assets (ROA), to measure profitability.[7] We restrict the sample to firms that provide full information on ROA, industrial classification of activity, and their number of employees, over the period 2002–2006, prior to advent of the global financial crisis in 2007. We drop firms corresponding to two specific industries: NACE codes 7415 (Chain services and non-financial holdings) and 7487 (Other Business activities). This is akin to the exclusion of depositary institutions in research using the Compustat database.

The basic statistical unit in our analysis is the firm: a legal unit that reports its own accounts and is legally distinct from other entities that it owns or is owned by. The firm's country is the country in which it reports accounts. Industries are defined according to the European Statistical Classification of Economic Activities (NACE) at the 4-digit level.

In order to account for corporate group influences on the performance of firms in a precise manner, we define a corporate group as the set of firms which, though legally distinct, are bound together by ties of majority share ownership. This is considered to be sufficient to provide a clear basis for effective managerial control (OECD, 2005, p. 49). This definition, along with the features of our data and the European legal environment, allows us to accurately account for the extent of corporate group diversification and internationalization. As we have information on every industry that every firm operates in, our measures of diversification capture diversification both among and within corporate group affiliates. In terms of internationalization, a firm wishing to operate outside of its home country must, in the European context, open a subsidiary in the target host country, which will legally be another firm, and will enter in our dataset as such.

We identify an ownership link when a firm has owns more than 50% of the equity of another firm. This threshold is sufficient to enable the apex owning firm to determine corporate policy of owned firms, by choosing appropriate directors if necessary.[8] This method of delineating corporate groups is in line with research on groups in Europe (Belenzon et al., 2013), and elsewhere (e.g., Cestone & Fumagalli, 2005; Morck, 2005). We impose the additional condition that ownership be above the 50% threshold for at least 2 years, as we do not want to include transient ties which are unlikely to provide the same kinds of benefits as established ones (Gulati et al., 2000, p. 208). Using these criteria, we identify 887,443 links between pairs of firms.

As we require information on the industrial classification of each firm at the level of 4-digit NACE for the repeated sampling procedure that we use, we exclude links when (mainly

[7]ROA is defined as the ratio between profits before taxes and fixed assets. Hawawini et al. (2003) use value-based measures of performance as well as ROA in their analysis and find the results to be similar.

[8]This is also possible by controlling more than half the shareholders' voting power indirectly. We are restrictive in requiring controlling ownership.

non-European) subsidiary firms are not in the Amadeus database. We are left with 450,782 links between 628,055 firms. Of these, 28.7% are solely *main* firms, 66.1% are solely *subsidiaries* and the remaining 5.1% are simultaneously *main* (with at least one subsidiary) and *subsidiary*.

From these ownership links, we identify 179,089 corporate groups. The majority of groups are constituted by a unique link between two firms. The average group consists of 2.5 links, but the biggest group has 1,096 links. Overall, 66.1% of all corporate groups have all their firms in the same country, and larger groups are more likely to be internationalized.

To bring our analysis methods to data, we use a stratified random sampling procedure and draw 100 samples of 5,000 firms each that are representative of the underlying population of firms along country, industry, and size dimensions. The stratification criteria are sourced from the Structural Business Statistics (SBS) database of the Statistical Office of the European Commission (Eurostat), which provides information on the numbers of firms in each European Union country and Norway, classified by industry and size-class. The final stage in our sampling procedure is the selection of firms with and without corporate group membership. Half of each of our samples is drawn from the population of group members. As the tracing of corporate groups is completed before sampling, we are able to identify cases of firms belonging to the same corporate group even when they are linked through firms that are not included in the sample.

We use the re-sampling approach for three reasons. First, the Amadeus data is not representative of the underlying population of firms; the database is known to be biased towards larger firms. Re-sampling enables us to overcome this bias and to provide estimates that are representative of the underlying population of firms. Second, the large size of the Amadeus database (14 million firms) makes it computationally infeasible to use all available data to estimate fixed- or random-effect regression models. Third, re-sampling enables us to obtain the bootstrapped sampling distributions, and corresponding standard errors and confidence intervals, as explained below.

To judge statistical significances of the estimates generated by Shapley Value, ANOVA, HLM, and VCA methods, we need their sampling distributions. With our randomly drawn re-samples being representative of a sufficiently large population, we can estimate bootstrapped confidence intervals. This is preferable to using the asymptotic distributions of the estimators for inference. We report simple 95% two-tailed confidence limits (Efron & Tibshirani, 1993). In terms of procedure, we (re)sample (size = 5,000) from the dataset with replacement; calculate the Shapley Value, ANOVA, HLM, and VCA estimates; and repeat this step 100 times, obtaining 100 sets of bootstrapped estimates for each method. These bootstrap statistics are then rank ordered, and the confidence limits are obtained as the 2.5 and 97.5% percentiles. These percentile intervals are nonparametric in that the critical values are obtained by rank, without restrictive assumptions such as normality, and are straightforward to estimate as well as interpret.

4 | RESULTS

Table 1 provides the summary statistics across our 100 (stratified) samples drawn from the database. Each of these samples comprises 5,000 firms, spanning 25 countries and 44 two-digit industries. From any corporate group contained in it, each sample contains between 2 and 14 affiliates, operating in up to 6 industries, and domiciled in up to 7 countries.

Table 2 presents the results. These results take note of potential participation by firms in more than one industry by including an additional dummy variable to indicate each secondary

TABLE 1 Sample summary statistics (average over 100 samples)

	Mean	Minimum	Maximum
Countries	25	25	25
2-digit industries	44	44	44
3-digit industries	192	183	200
4-digit industries	415	400	435
Corporate groups	1,072	1,052	1,089
Firms in corporate groups	2,504	2,495	2,505
Firms per corporate group	2.11	2	14
Industries per corporate group			
2-digits	1.59	1	6
3-digits	1.73	1	8
4-digits	1.8	1	9
Industries per firm (4-digits)	2.21	1	33
Countries per corporate group	1.2	1	7

industry in which each firm operates. This makes our estimates robust to changes in firms' primary industries[9] while also correcting for the likely downward bias on industry effects that can arise from only primary industries being reckoned in regression-based variance decomposition (Bowman & Helfat, 2001, pp. 14–15). Corporate effects are likely to be underestimated in samples that include unaffiliated firms (Bowman & Helfat, 2001). We therefore obtain estimates for samples that include only firms belonging to corporate networks.[10]

Overall, the baseline Shapley Value results presented in the first column suggest that firm effects constitute the most important component in the variance of firm profitability in Europe, with a mean Shapley Value of 34.1%. Corporate group effects are second in importance, accounting for 12.5% of the variance in firm profitability. Industry effects appear to account for only 4.1% of variance in firm profitability. Both country and year effects are of little importance, though the nulls of no country or year effects at all are ruled out by the confidence intervals.

The small magnitude of the Shapley Value for year effects (0.1% of the variance in firm profitability—of the same order of magnitude as the year component reported in the literature) is not surprising, given that the period of analysis was relatively stable economically. The Shapley Value estimate of country effects (0.4% of the variance) is lower than that obtained by Makino et al. (2004), whose data on subsidiaries of Japan headquartered firms span a wider and more heterogeneous set of countries (79 in all). Estimates in McGahan and Victer (2010) of home-country effects lie in the range of 2.6–3.0%, but they find comparably small home country-effects among European firms, falling to 0.3% with a matched Amadeus sample.

Table 2 also presents estimates of effect importance produced by ANOVA, HLM, and VCA approaches. The ANOVA results presented in columns 2 and 3 are the effect contributions to

[9]A limitation of the AMADEUS data that we use is that information on firm-industry affiliation is provided only for the most recent year of the data. Including secondary industries in our estimation of industry effects ensures that our estimates are not affected by cases in which a firm's primary industry changed between the first (2002) and last years of our data (2006), as long as the new primary industry was previously in the firm's set of secondary industries.

[10]When unaffiliated firms are also included, estimates of both industry and corporate group effects are lower, while firm effects are higher. These results are available on request from the authors.

TABLE 2 Contributions to explanatory power by effect type (means, bootstrapped 95% confidence intervals in brackets, 100 samples)

<table>
<tr><th></th><th>Shapley</th><th>ANOVA G, I</th><th>ANOVA I, G</th><th>HLM C</th><th>HLM I</th><th>HLM C&I</th><th>VCA</th></tr>
<tr><td>Country</td><td>0.36% [0.20, 0.50]</td><td>0.84% [0.43, 1.33]</td><td>0.84% [0.43, 1.33]</td><td>0.43% [−0.08, 0.87]</td><td></td><td rowspan="2">2.21% [1.21, 4.55]</td><td>2.30% [0.00, 17.97]</td></tr>
<tr><td>Industry</td><td>4.07% [3.43, 4.84]</td><td>6.48% [5.01, 7.80]</td><td>8.92% [7.45, 10.30]</td><td></td><td>1.62% [0.81, 4.14]</td><td>2.39% [1.01, 4.24]</td></tr>
<tr><td>Corporate group</td><td>12.46% [11.53, 13.36]</td><td>25.53% [23.44, 27.59]</td><td>23.09% [21.23, 25.31]</td><td>12.58% [9.20, 12.91]</td><td>11.29% [7.97, 12.26]</td><td>10.84% [7.79, 12.16]</td><td>9.61% [5.57, 13.76]</td></tr>
<tr><td>Year</td><td>0.06% [0.01, 0.13]</td><td>0.04% [−0.004, 0.1]</td><td>0.04% [−0.004, 0.1]</td><td>0.14% [0.02, 0.15]</td><td>0.14% [0.02, 0.15]</td><td>0.14% [0.02, 0.15]</td><td>0.08% [0.00, 0.33]</td></tr>
<tr><td>Firm</td><td>34.11% [32.75, 36.10]</td><td>18.17% [16.47, 20.18]</td><td>18.17% [16.47, 20.18]</td><td>37.14% [36.69, 42.57]</td><td>37.24% [35.48, 41.28]</td><td>37.11% [35.30, 40.93]</td><td>37.84% [32.94, 42.95]</td></tr>
</table>

model explanatory power (increase in $\bar{R}^2$ due to the introduction of the effect in question) based on the two model paths closest to those used in the existing literature. In both cases, year effects are introduced first into the model, then country effects, while firm effects are introduced last. The results presented in column "ANOVA G, I" are estimates from a model path in which corporate group effects are introduced third into the model, while industry effects are fourth; column "ANOVA I, G" corresponds to industry effects being introduced third into the model, before corporate group effects. These results show that the magnitudes of ANOVA estimates of industry and corporate group effects are dependent on the order in which the effects are introduced, with those that are introduced third claiming around 2.5% more of the total variance compared to those introduced fourth. It is interesting to note, as in the simulation results, that the ANOVA estimates for industry and corporate group effects are much larger than those produced by the Shapley Value, HLM, and VCA approaches, while firm effects are estimated to be much smaller. Based on comparison with the Shapley Value and HLM estimates, and the simulation results, it seems likely that this reflects a bias in the ANOVA approach if it is used to consider only some particular model path to the exclusion of others. Among model paths considered here, industry and corporate group effects claim a large share of their covariances with firm effects, leading to the overestimation of the former and the underestimation of the latter.

The next three columns of Table 2 present HLM results for models in which country ("HLM C"), industry ("HLM I"), or both country and industry ("HLM C&I") fixed effects are added to models with random firm and corporate group effects.[11] While the estimates of these alternative models are not very different from one another, firm effect estimates are somewhat higher than

[11]We chose to treat corporate group effects as random effects following both the methodological guidance mentioned in Misangyi et al. (2006) and the results of our simulation, which suggest that doing so produces estimates with lower RMSEs than if we were to treat corporate group effects as fixed and industry effects as random instead.

those produced by the Shapley Value approach, while industry effects appear to be lower, although the differences are far less pronounced compared to the ANOVA approach. The VCA results presented in the last column of the table are broadly similar to those from the HLM models, but with higher estimates for country effects, lower estimates for corporate group effects, and wider confidence intervals compared to all other methods. The wider confidence intervals are likely due to the samples having to be carefully split (ensuring that group members are kept together in the resulting subsamples) in order to achieve convergence in the estimation of crossed-effects models with a large number of random effects. This is a further downside of the VCA approach in practice.

The above results are the baseline, unrestricted estimates of the importance of corporate group effects. We now proceed to examine whether and how corporate group effects change with the extents of diversification and internationalization of the group. Once again, we compare the results yielded by the Shapley Value approach with those produced by alternative methods.

We begin by considering the relationship between group size, in terms of the overall number of the group's affiliates, and the extent of corporate effects. To do this, we analyze subsamples from our data that include corporate groups containing k or more ownership links (affiliates), allowing k to range from 1, the baseline model, to 10, representing the largest groups.

To explore how corporate group influence on affiliate performance changes with the extent of *related* diversification, we estimate the importance of corporate effects in groups spanning k_1 or more NACE 4-digit industries (k_1 ranging from 1 to 10) within the same NACE 2-digit sector. To examine how *unrelated* diversification affects the influence of corporate groups on firm performance, we consider samples including groups spanning k_2 or more 2-digit NACE industries, k_2 ranging from 1 to 10. Finally, to evaluate the evidence in relation to internationalization, we consider samples of groups that span at least k_3 countries, allowing k_3 to range from 1 to 10.

Figure 1 presents the results from the Shapley Value approach (top-left panel), ANOVA when corporate group effects are introduced before industry effects (top-right panel, corresponding to the "ANOVA G, I" column in Table 2),[12] HLM with both country and industry fixed-effects (bottom-left panel, corresponding to the sixth column of Table 2), and VCA (bottom-right panel), calculated using the above defined counter-cumulative sequences of subsamples that differ in the extent group of diversification and internationalization. The bootstrapped 95% confidence intervals of these estimates are shown as light-gray lines on the figures.

While the results from the Shapley Value and ANOVA approaches appear to show similar *patterns*, the substantial differences in estimates of effect magnitudes between ANOVA and other approaches persist in this analysis. The ANOVA estimates are substantially larger in all cases. When compared to the Shapley Value and ANOVA results, those produced by the VCA and HLM approaches seem to be less precise, strikingly so in the HLM case, especially once corporate group span increases beyond 2. Interestingly, while the bootstrapped confidence intervals for the ANOVA, VCA, and Shapley Value approaches are largely symmetric, the HLM confidence intervals are highly skewed—bounded by zero on one side but sometimes very large on the other. Overall, these results suggest that neither the ANOVA nor the HLM or VCA approaches are well-suited for the investigation of contingencies affecting the importance of certain effects in explaining variation in firm performance. On that basis, we focus on the Shapley Value results below.

[12]Results from "ANOVA I, G" are similar and are available on request from the authors.

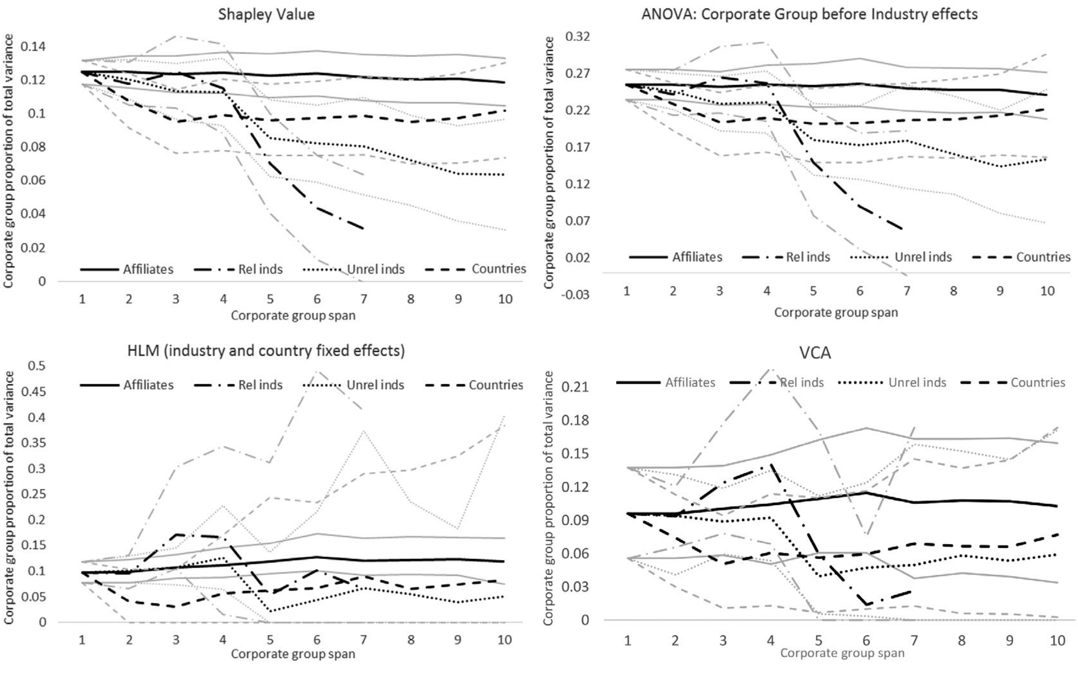

FIGURE 1 Comparison of variance decomposition methods in estimating corporate group effect share of total variance as minimum number of firms/industries/countries which they span increases (means, 100 samples, bootstrapped 95% confidence intervals in gray)

The proportion of variance in firm profitability accounted for by corporate group effects does not change much with an increasing number of ownership links in the group, suggesting that the results below are indeed driven by diversification and internationalization, rather than simply group scale. The proportion of the variance in firm profitability that is accounted for by corporate group effects falls precipitously as groups come to operate in more than four subindustries within the same 2-digit sector. Specifically, the Shapley Value of corporate group effects remains roughly stable for groups operating in between one and four related industries, fluctuating between 12.5 and 11.5%, before declining sharply to 3.1% as the number of subindustries within the 2-digit sector increases from 4 to 7. Operating in an increasing number of unrelated industries is associated with the share of variance in firm profitability accounted for by corporate group effects falling from 12.5% in the unrestricted case to 8.5% for groups spanning at least 5 (unrelated) 2-digit NACE industries. It falls further to 6.3% for groups that span at least 10 unrelated industries. Internationalization is associated with the corporate group effect falling to 10.8% (from 12.5%) for groups spanning at least two countries, and to 9.5% for groups spanning at least 3 countries. Beyond this degree of internationalization, the share of explained by corporate group effects remains stable.

5 | DISCUSSION

This paper has sought to add the Shapley Value method to the toolkit of strategy researchers by establishing its reliability in performing variance decompositions. We highlighted the axiomatic rationale underpinning the method, and the uniqueness and superior properties of its estimates when effects are correlated. We compared it with extant methods and showed its greater accuracy and precision, both with a simulated dataset and in an empirical application. Finally, the Shapley Value approach has also been shown to be better suited for answering novel research questions, such as how the dispersion importance of corporate groups varies with the extent of the group's diversification and internationalization. We now proceed to discuss these contributions, and to consider the opportunities for future research opened up by the Shapley Value approach.

The results from the simulation and the empirical application demonstrate the value of the Shapley Value approach compared to extant variance decomposition methods. The Shapley Value approach provides more accurate measures of effect importance compared to an ANOVA approach, whose estimates are sensitive to the order in which effects are introduced; and also compared to HLM, whose estimates vary more across different samples drawn from the same population—increasingly so as sample size decreases. The Shapley Value estimates are also more accurate and precise than those produced by VCA.

Our analysis of how diversification and internationalization affects corporate group influence demonstrates that beyond providing more accurate answers to old research questions, the Shapley Value method enables scholars to address novel research questions that could not be reliably answered using extant methods. In particular, an analysis of the contingencies that influence effect importance—which, as demonstrated, cannot be reliably accomplished using extant methods—has the potential to illuminate the ongoing debate about sources of variation in firm profitability (e.g., Crossland & Hambrick, 2007; Fitza, 2017; Fitza & Tihaniy, 2018; Guo, 2017; Ma et al., 2013; Quigley & Graffin, 2017). The greater reliability of the Shapley Value approach when applied to datasets with fewer observations should also prove an important advantage for researchers seeking to understand novel organizational phenomena such as the drivers of success in crowdfunding campaigns (e.g., Dushnitsky & Fitza, 2018), the factors influencing the growth trajectories of decentralized autonomous organizations (e.g., Hsieh, Vergne, Anderson, Lakhani, & Reitzig, 2018), and the role of artificial intelligence technologies in shaping the performance of firms and markets (Agrawal, Gans, & Goldfarb, 2019).

Our results suggest that corporate groups that span a greater number of industries and countries account for a smaller proportion of the variation in profitability of affiliate firms, particularly so as the extent of related diversification increases. The increase in within-group variance in profitability accompanying related diversification is therefore significantly greater than any increase in between-group variance. This implies that even the best-managed corporate groups struggle to effectively exploit their group-level advantages in a large number of related industries in the face of adjustment and coordination costs. It is particularly noteworthy that the effect of increasing unrelated diversification on the proportion of variation in the profitability of affiliates accounted for by the corporate group is smaller than that associated with increasing related diversification. This suggests that, in the case of increasing unrelated diversification, either within-group heterogeneity in affiliate profitability increases by less, or that between-group heterogeneity (in the cross-section) increases by more. An explanation for the former could relate to the reallocation of capital from better performing affiliates to poorer-performing ones. The latter could result from decisions on the extent of a group's unrelated diversification sometimes being made by those unable to manage unrelated businesses effectively. Both possibilities have been discussed in the literature on corporate finance focused on potential agency problems in conglomerates (see Maksimovic & Phillips, 2007, for a review). In future research it may be fruitful to examine the types of corporate structures under which either or both issues would be most pronounced.

Interestingly, the relationship between corporate group effects and the extent of group internationalization is different. In the European setting, the sizes of group effects fall as groups come to span a minimum of two, and then of three countries, but remain largely stable with further internationalization. Whether this finding holds for groups operating across more institutionally heterogeneous countries is a fruitful international management research question.

The Shapley Value approach has limitations. First, it is computationally expensive. As 2^K coalitions can be constituted out of K regressors, the number of regressions required to find the Shapley Value increases exponentially with the number of regressors. Second, the size and power properties of tests of hypotheses using the approach requires more research. Third, while the Shapley Value approach satisfies the essential and desirable property of proper and exact decomposition into non-negative dispersion importances, with any regressor having a non-zero coefficient in the full model always receiving non-zero dispersion importance, the converse of this—that the share allocated to a regressor with coefficient equal to zero in the full regression should be zero—is not satisfied. However, this unsatisfied feature is not a desirable property when there is model uncertainty and potential mediation effects. Finally, regressors with high Shapley Values are natural candidates to prioritize when the objective is to influence performance. But it must be noted that Shapley Values are based on the full set of regressors in the model. The correlation structure among the regressors must therefore not be ignored in drawing practical conclusions on ways to enhance performance.

In this paper, we have sought to show that the Shapley Value method improves the reliability of estimates apportioning heterogeneity in firm performance. It also allows strategy scholars to ask new questions regarding contingencies driving effect importance. We hope that the Shapley Value approach will be a valuable addition to the methodological toolbox of strategic management research.

ACKNOWLEDGEMENTS

The authors are grateful for the suggestions and guidance of associate editor Sea-Jin Chang and three anonymous reviewers. The paper was also shaped by feedback from Gerry McNamara and seminar participants at the Academy of Management Annual Meeting 2011, VI Jornadas de Integración Económica 2010, Universidad Complutense de Madrid, and Cambridge Judge Business School.

REFERENCES

Agrawal, A., Gans, J., & Goldfarb, A. (Eds.). (2019). *The economics of artificial intelligence: An agenda*. Chicago, IL: University of Chicago Press.

Almeida, H., & Wolfenzon, D. (2006). Should business groups be dismantled? The equilibrium costs of efficient internal capital markets. *Journal of Financial Economics*, *79*, 99–144.

Arora, A., Gittelman, M., Kaplan, S., Lynch, J., Witchell, W., & Siggelkow, N. (2016). Question-based innovations in strategy research methods. *Strategic Management Journal*, *37*, 3–9.

Barney, J. B. (1997). *Gaining and sustaining competitive advantage*. Reading, MA: Addison-Wesley.

Belenzon, S., & Berkovitz, T. (2010). Innovation in business groups. *Management Science*, *56*(3), 519–535.

Belenzon, S., Berkovitz, T., & Rios, L. A. (2013). Capital markets and firm organization: How financial development shapes European corporate groups. *Management Science*, *59*(6), 1326–1343.

Bertrand, M., Mehta, P., & Mullainathan, S. (2002). Ferreting out tunneling: An application to Indian business groups. *Quarterly Journal of Economics*, *117*, 121–148.

Bowman, E. H., & Helfat, C. E. (2001). Does corporate strategy matter? *Strategic Management Journal*, *22*, 1–23.

Brush, T. H., & Bromiley, P. (1997). What does a small corporate effect mean? A variance components simulation of corporate and business effects. *Strategic Management Journal*, *18*, 825–835.

Brush, T. H., Bromiley, P., & Hendrickx, M. (1999). The relative influence of industry and corporation on business segment performance: An alternative estimate. *Strategic Management Journal*, *20*, 519–547.

Carney, M., Gedajlovic, E. R., Heugens, P. P. M. A. R., Van Essen, M., & Van Oosterhout, J. H. (2011). Business group affiliation, performance, context, and strategy: A meta-analysis. *Academy of Management Journal*, *54*(3), 437–460.

Caves, R. E. (1996). *Multinational enterprise and economic analysis* (2nd ed.). Cambridge, MA: Harvard University Press.

Cestone, G., & Fumagalli, C. (2005). The strategic impact of resource flexibility in business groups. *RAND Journal of Economics*, *36*(1), 193–214.

Chan, C. M., Makino, S., & Isobe, T. (2010). Does subnational region matter? Foreign affiliate performance in the United States and China. *Strategic Management Journal*, *31*, 1226–1243.

Chang, S. J., & Hong, J. (2000). Economic performance of group-affiliated companies in Korea: Intragroup resource sharing and internal business transactions. *Academy of Management Journal*, *43*(3), 429–448.

Chari, M. D. R., Devaraj, S., & David, P. (2008). The impact of information technology investments and diversification strategies on firm performance. *Management Science*, *54*(1), 224–234.

Crossland, C., & Hambrick, D. C. (2007). How national systems differ in their constraints on corporate executives: A study of CEO effects in three countries. *Strategic Management Journal*, *28*, 767–789.

Dushnitsky, G., & Fitza, M. (2018). Are we missing the platforms for the crowds? Comparing investment drivers across multiple crowdfunding platforms. *Journal of Business Venturing Insights*, *10*, e00100.

Efron, B., & Tibshirani, R. J. (1993). *An introduction to the bootstrap*. London: Chapman & Hall.

Fitza, M. (2017). How much do CEOs really matter? Reaffirming that the CEO effect is mostly due to chance. *Strategic Management Journal*, *38*, 802–811.

Fitza, M., & Tihaniy, L. (2018). How much does ownership form matter? *Strategic Management Journal*, *38*, 2726–2743.

Grömping, U. (2007). Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician*, *61*(2), 139–147.

Gulati, R., Nohria, N., & Zaheer, A. (2000). Strategic networks. *Strategic Management Journal*, *21*, 203–215.

Guo, G. (2017). Demystifying variance in performance: A longitudinal multilevel perspective. *Strategic Management Journal*, *38*, 1327–1342.

Hashai, N. (2015). Within-industry diversification and firm performance—An S-shaped hypothesis. *Strategic Management Journal*, *36*, 1378–1400.

Hawawini, G., Subramanian, V., & Verdin, P. (2003). Is performance driven by industry- or firm-specific factors? A new look at the evidence. *Strategic Management Journal*, *24*, 1–16.

Hitt, M. A., Hoskisson, R. E., & Kim, H. (1997). International diversification: Effects on innovation and firm performance. *Academy of Management Journal*, *40*, 767–798.

Hough, J. R. (2006). Business segment performance redux: A multilevel approach. *Strategic Management Journal*, *27*, 45–61.

Hsieh, Y.-Y., Vergne, J.-P., Anderson, P., Lakhani, K., & Reitzig, M. (2018). Bitcoin and the rise of decentralized autonomous organizations. *Journal of Organization Design*, *7*, 14.

Kali, R., & Sarkar, J. (2011). Diversification and tunneling: Evidence from Indian business groups. *Journal of Comparative Economics*, *39*, 349–367.

Levinthal, D., & Wu, B. (2010). Opportunity costs and non-scale free capabilities: Profit maximization, corporate scope and profit margins. *Strategic Management Journal*, *31*, 780–801.

Lu, J. W., & Beamish, P. W. (2004). International diversification and firm performance: The S-curve hypothesis. *Academy of Management Journal*, *47*(4), 598–609.

Ma, X., Tong, T., & Fitza, M. (2013). How much does subnational region matter? Evidence from multinational corporations subsidiaries in a large emerging economy. *Journal of International Business Studies*, *44*(1), 66–87.

Majumdar, S. K., & Bhattacharjee, A. (2014). Firms, markets, and the state: Institutional change and manufacturing sector profitability variances in India. *Organization Science*, *25*, 509–528.

Makino, S., Isobe, T., & Chan, C. M. (2004). Does country matter? *Strategic Management Journal*, *25*, 1027–1043.

Maksimovic, V., & Phillips, G. M. (2007). Conglomerate firms and internal capital markets. In B. Espen Eckbo (Ed.), *Handbook of corporate finance* (pp. 423–477). North-Holland: Amsterdam.

Marchenko, Y. (2006). Estimating variance components in Stata. *Stata Journal*, *6*, 1–21.

McGahan, A. M., & Porter, M. E. (1997). How much does industry matter, really? *Strategic Management Journal*, *18*, 15–30.
McGahan, A. M., & Porter, M. E. (2002). What do we know about variance in accounting profitability? *Management Science*, *48*, 834–851.
McGahan, A. M., & Porter, M. E. (2005). Comment on 'industry, corporate, and segment effects and business performance: A non-parametric approach' by Ruefli and Wiggins. *Strategic Management Journal*, *26*, 873–880.
McGahan, A. M., & Victer, R. (2010). How much does home country matter to corporate profitability? *Journal of International Business Studies*, *41*, 142–165.
Misangyi, V. F., Elms, H., Greckhamer, T., & Lepine, J. A. (2006). A new perspective on a fundamental debate: A multilevel approach to industry, corporate, and business unit effects. *Strategic Management Journal*, *27*, 571–590.
Morck, R. (2005). *How to eliminate pyramidal business groups: The double taxation of inter-corporate dividends and other incisive uses of tax policy*. Cambridge: MIT Press.
Nohria, N., & Ghoshal, S. (1994). Differentiated fit and shared values: Alternatives for managing headquarters-subsidiary relations. *Strategic Management Journal*, *15*(6), 491–502.
OECD. (2005). *Measuring globalisation: OECD handbook on economic globalisation indicators*. Paris: OECD Publishing.
Pintér, M. (2011). Regression games. *Annals of Operations Research*, *186*, 263–274.
Quigley, T. J., & Graffin, S. D. (2017). Reaffirming the CEO effect is significant and much larger than chance: A comment on Fitza (2014). *Strategic Management Journal*, *38*, 793–801.
Rawley, E. (2010). Diversification, coordination costs, and organizational rigidity: Evidence from microdata. *Strategic Management Journal*, *31*, 873–891.
Ruefli, T. W., & Wiggins, R. R. (2003). Industry, corporate, and segment effects and business performance: A non-parametric approach. *Strategic Management Journal*, *24*, 861–879.
Rumelt, R. R. (1991). How much does industry matter? *Strategic Management Journal*, *12*, 167–185.
Schmalensee, R. (1985). Do markets differ much? *American Economic Review*, *75*(2), 341–351.
Shapley, L. S. (1953). A value for n-person games in contributions to the Theory of Games II. In H. Kuhn & A. W. Tucker (Eds.), *Annals of mathematics studies* (Vol. 28, pp. 307–317). Princeton, NJ: Princeton University Press.
Young, H. P. (1985). Monotonic solutions of cooperative games. *International Journal of Game Theory*, *14*, 65–72.
Zhou, Y. M. (2011). Synergy, coordination costs, and diversification choices. *Strategic Management Journal*, *32*, 624–639.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of this article.

How to cite this article: Sharapov D, Kattuman P, Rodriguez D, Velazquez FJ. Using the Shapley Value approach to variance decomposition in strategy research: Diversification, internationalization, and corporate group effects on affiliate profitability. *Strat Mgmt J*. 2021;42:608–623. https://doi.org/10.1002/smj.3236