Revisiting Profitability: Firm, Business Group, Industry and Country Effects

Paul Kattuman[*], Diego Rodriguez[†], Dmitry Sharapov[‡] and F. Javier Velázquez[§]

January 2011

Abstract

There is a large body of research in the fields of strategic management and economics that seeks to understand the sources of heterogeneity in firm performance. One approach to this has been to decompose variance in firm-level profitability into firm, supra-firm and industry components. This paper seeks to contribute to this literature in two main ways. First, in contrast to previous literature which has focused almost exclusively on listed corporations in the United States, we analyze differences in profitability between firms (both listed and non-listed), operating in all industries with the exception of the agrarian sector in 25 European countries. These data allow us to investigate the influence of country effects alongside firm, business group, industry and year effects. The size of the database allows us to use a repeated sampling

[*]University of Cambridge, Cambridge Judge Business School, Trumpington Street, Cambridge CB2 1AG, UK. E-mail: p.kattuman@jbs.cam.ac.uk.

[†]Universidad Complutense de Madrid and GRIPICO. E-mail: drodri@ccee.ucm.es.

[‡]University of Cambridge, Cambridge Judge Business School, Trumpington Street, Cambridge CB2 1AG, UK. E-mail: ds361@cam.ac.uk.

[§]Universidad Complutense de Madrid and GRIPICO. E-mail: javel@ccee.ucm.es.

approach in the analysis of the underlying population of firms along country, industry and size dimensions. Our second contribution is methodological. We propose a new method of variance decomposition, the Shapley Value method. Relative to methods used in prior studies, the Shapley Value method obtains more accurate estimates of the contributions of the different effects to the variance in firm profitability in the presence of covariance between effects. We show this by evaluating it against alternatives using Monte-Carlo simulations. Our empirical results suggest that business group effects are the second largest source of variance in firm-level profitability after firm effects, and that the importance of the business group appears to decrease with increasing group size. Industry effects appear to be much smaller than estimated in previous work, while country and year effects are smaller still.

Keywords: profitability, variance decomposition, Shapley value, business group effects, country effects.

Draft version: preliminary and incomplete. Please do not quote, cite or circulate without permission from the authors.

1 Introduction

Understanding the determinants of heterogeneity in firm profitability is arguably one of the most fertile fields of analysis both for industrial economists and strategic management researchers. Schmalensee (1985) initiated a literature using variance decomposition techniques on Federal Trade Commission line-of-business profitability data, and finding significant industry effects, and negligible corporate parent effects (called firm effects by Schmalensee). Many papers followed suit applying increasingly advanced variance decomposition methods to more comprehensive data (the more notable contributions include Brush, Bromiley, and Hendrickx, 1999; Hough, 2006; McGahan and Porter, 1997, 2002; Misangyi et al., 2006; Roquebert, Phillips, and Westfall, 1996; Ruefli and Wiggins, 2003; Rumelt, 1991; Short et al., 2007). The key issue of contention in the debate to which this study seeks to contribute, is about the relative importance of industry, corporate and firm effects in explaining the variation in performance between firms.[1]

While previous studies are in general agreement that the main source of variance in profitability is at the firm level, there is no such consensus regarding the relative importance of corporate parent and industry effects. A number of studies have found industry effects to be larger than corporate parent effects (McGahan and Porter, 1997; Rumelt, 1991; Schmalensee, 1985), others estimate industry and corporate effects to be similar in magnitude (McGahan and Porter, 2002; Misangyi et al., 2006), while yet others find corporate effects to be significantly larger than industry effects (Hough, 2006; Roquebert, Phillips, and Westfall, 1996; Ruefli and Wiggins, 2003). This diversity of results is surprising as all of these papers have used largely similar methods to analyze data on publically listed firms in the United States, sourced for the most part from the Compustat database.

McGahan and Porter (2002) offer a comprehensive review of this literature, reconciling the results of the main previous studies by showing that the size of corporate parent effects depends on whether or not undiversified firms are included in the sample, and providing new evidence.[2] In their concluding remarks the authors state

[1] The titles of key papers in this debate illustrate this : 'Do Markets Differ Much?' (Schmalensee, 1985), 'How Much Does Industry Matter?' (Rumelt, 1991), 'Markets vs Management: What 'Drives' Profitability?' (Roquebert, Phillips, and Westfall, 1996), 'How Much Does Industry Matter, Really?' (McGahan and Porter, 1997), 'Is Performance Driven by Industry- or Firm-Specific Factors? A New Look at the Evidence' (Hawawini, Subramanian, and Verdin, 2003).

[2] For a more wide-ranging review see Bowman and Helfat (2001).

that 'the most direct opportunities for further research reside in exploring new data. Reliable and comparable data on the accounting profits of firms in other parts of the world would yield insight on questions about the relationships between the national economic environment and industrial performance. Data on the profitability of privately held firms would provide results more representative of the entire economy' (p. 849). This paper responds to that exhortation and contributes to the literature in two specific ways.

First, in contrast to above mentioned papers which used data on listed corporations in the United States, we assess the sources of variance in profitability using listed and non-listed firms, located in a wide range of countries. Specifically, we use the Amadeus database of firms operating in all sectors with the exception of the agrarian sector, in 25 European countries, to estimate the influence of firm, business group, industry, country and year effects on differences in firm profitability. Due to the large size of the Amadeus database we use repeated random sampling stratified by size-class, industry and country, and recursive estimation to produce robust results representative of patterns in the underlying population of firms. Our analysis is thus more general than those in the literature, and our European focus provides a useful contrast to the dominance of US data sources in the discussion to date.

Second, we contribute to the literature by using a novel methodology for variance decomposition in the presence of covariance between effects. The Shapley Value method allows us to estimate the contribution of each effect to variance in firm profitability by calculating the weighted average of increases to adjusted R^2 due to the inclusion of the effect, in all possible models ranging from a null ANOVA model to one which includes all effects. The Shapley Value approach provides a more robust way to assess the relative importance of different effects than the consideration of a single path from the null to the full model, or of several paths (as done in previous papers) without aggregating them in a consistent manner. To assess the effectiveness of the Shapley Value approach in an empirical context affected by covariance between effects we carry out a simulation analysis in a Monte Carlo framework, comparing estimates produced by the approach with those produced by the approaches used in the majority of previous studies, ANOVA and Variance Components Analysis (VCA)[3].

2 The Decomposition of Accounting Profitability

2.1 The Empirical Model

Alongside firm-level variations, three sources of performance heterogeneity have been the frequently examined: industry level variations; a supra-firm effect usually conceptualized as a corporate parent effect, and year effects. Industry effects capture specific and and persistent outcomes of industry characteristics that lead to different average values of profitability across sectors. For example, differences in entry barriers can generate persistent differences in average profitability between the tobacco and the meat and food processing industries. Corporate parent effects capture the influence on firm-level profitability of membership in business groups controlled by corporate parents with heterogeneous structural or managerial characteristics. The extent of business group diversification and the degree of management structure centralization are examples of such corporate-parent characteristics. A number of papers have considered supra-firm levels of analysis other that the corporate: Chang and Hong (2002) consider business group effects, Short et al. (2007) take the strategic group as a level of analysis and Crossland and Hambrick (2007) look at CEOs as

[3]Called Components of Variance (COV) analysis in earlier papers (e.g. Rumelt, 1991; Schmalensee, 1985).

a source of variation in firm performance. Finally, year effects capture profitability changes through the economic cycle that are common to all firms. Increasing market openness as a consequence of economic integration, exchange rate fluctuations or aggregate demand shocks are examples of causes in that context.

It is useful to discuss the main issues in the literature in terms of a model for profitability for a statistical unit in year y, explained by a set of characteristics. In this paper the statistical unit refers to a firm f, with main activity in industry i, located in a country c, and belonging to a business-group g. We observe the accounting profitability of the firm: π. The specification of the full model in which all these characteristics enter additively is then:

$$\pi_{icgfy} = \mu + \alpha_i + \kappa_c + \beta_g + \phi_f + \gamma_y \tag{1}$$

It is important to understand the differences between the statistical unit used in this paper, the *firm*, and in previous papers using the US Federal Trade Commission line of business data and Compustat reports in order to interpret our results. In this study, as is usual in statistical procedures, each firm is 'assigned' to an industry according to its *main* activity. Although we know whether the firm is product diversified, the information on profitability that we have access to, is for the firm as a whole, not for a specific product or activity. A similar procedure is followed in the FTC database, in which the industry classification corresponding to its main industry is assigned to each *business unit.* However, in Compustat each statistical unit is a *business segment* (i.e. a specific industry) in which a corporation operates. McGahan and Porter (1997) point out that due to this the observed business-segment may be an aggregation of several firms owned by the same corporate parent and operating in the same industry.

A further difference between this study and prior work is that business group in our data are constructed in a 'bottom-up' fashion using information on share ownership links among firms. For that reason our business groups are closer to a concept of firm ownership *networks* than to the concept of *corporations* used in papers based on US data. Thus it is possible, and even likely, that two or more firms in the same business-group are located in the same industry, in contrast to Compustat data. This implies that, though similar in general terms, corporate effects as investigated in the majority of the prior literature, and business group effects as defined in this study are not strictly comparable. While business-group effects can be expected to contribute significantly in explaining profitability variance with either definition, it is likely that the links between business segments in a corporation are stronger, than between firms that share ownership links but are not necessarily are under a common corporation umbrella. If that is the case, the contribution of business group effects as we conceive it should be smaller than the contribution of corporate effects estimated using US data.

The difference in statistical units may also have consequences on the size of industry effects. In particular, in framework defined by Compustat data, the contribution of industry effects may be underestimated insofar as a business-segment comprises different business-units (McGahan and Porter, 1997). We expect the opposite with the structure of the Amadeus database as firms which may operate in numerous different industries are assigned the industry corresponding to their 'main' activity. Data on sales across activities for diversified firms are not available, though we can observe whether industry effects are lower for these firms than for those not diversified.

The last set of effects included are at the firm-level. This introduces a statistical problem: any model that includes firm and year effects will have the same explanatory power as a fully specified model with the five sets of effects (year, country, industry, business group and firm) as firm effects are perfectly multicollinear with country, industry and business group effects. McGahan and Porter (2002) point out that the same issue emerges in the Compustat framework with business-specific effects, which are linear by design with industry and corporate-parents effects. Perfect multicollinearity between firm effects and each one of the other type of effects (except year effects) would not follow if firms could move in the observed period across countries, industries or business-groups. All three possibilities are excluded in our analysis, though for different reasons. First, a firm can not change across countries because in that case it would be, de facto, a different firm. Second, even though a firm can have activities in more than one industry, its main activity through the whole observed period defines its industry. Thus no transient industry effects (as in Rumelt, 1991) will be considered. Finally, business group effects are calculated using the network of ownership links through the whole period (see Section 3.2). Thus, either a firm belongs to the same business group in all years considered or it does not belong to any business group. These two possibilities are similar to the case of diversified/non-diversified corporations in the context of the McGahan and Porter (1997, 2002) papers.

2.2 Approaches to Variance Decomposition

A number of empirical methods have been used in the literature to decompose the observed variation in profitability at the firm level. While variants of the VCA and ANOVA approaches used in the classic studies (Rumelt, 1991; Schmalensee, 1985) have been widely adopted (e.g. Hawawini, Subramanian, and Verdin, 2003; McGahan and Porter, 1997, 2002; Roquebert, Phillips, and Westfall, 1996), more recent papers have experimented with a number of alternative approaches including simultaneous equation modelling (Brush, Bromiley, and Hendrickx, 1999), non-parametric estimation (Ruefli and Wiggins, 2003) and multilevel modelling techniques (Hough, 2006; Misangyi et al., 2006; Short et al., 2007).

The random-effect VCA method does not provide coefficient estimates but uses summary statistics to estimate the components of overall variance of profitability attributable to each effect and thus each effect's importance. This method requires the assumption that effects are randomly drawn from a population. It can also allow for covariance between effects but requires the assumption that these covariances are also randomly drawn. The VCA procedure does not provide an easy means of testing significance of the effects, but by estimating the contribution of each effect to total variance it allows judgements to be made regarding effect importance. This approach has been criticized on a number of grounds, including the appropriateness of the random effects assumption, and its lack of power in finding small but significant effects (Brush and Bromiley, 1997).

In contrast, the ANOVA approach uses fixed effects regression models to carry out exclusion F-tests for the effects in specifications ranging from the null model to a full model including all the effects. These tests allow researchers to make inferences regarding effect significance. If effects are significant, their contribution to explaining the total variance can be estimated by looking at the increases in adjusted R^2 arising due to effect inclusion in the regression specification. Rumelt (1991) and McGahan and Porter (1997) use nested ANOVA, while simultaneous ANOVA is used in McGahan and Porter (2002). A major problem with the ANOVA approach is that its estimates of the significance and importance of effects depend on the order in which these effects are introduced into the model. If there is covariance between effects, as has been discussed in the literature (e.g. Schmalensee, 1985, p. 344), ANOVA will assign the covariance to the effect introduced first into the specification. Aware of this issue and its implications, authors of previous papers have presented estimation results from a number of paths from the null to the full model.[4] However, since Schmalensee (1985) no study has presented the estimates relating to every possible path of variable addition going from the null to the full model. Where results from different model paths have been presented, authors have made no attempt to aggregate these results. This leaves open the possibility that the identification of the contributions of different effects are confounded by the inequitable attribution of the covariance between the effects to one or more of those effects at the expense of others.

Brush, Bromiley, and Hendrickx (1999) use a simultaneous equations method using 2 stage least-squares (2SLS) regressions. The authors argue for its superiority over the ANOVA approach as the 2SLS method uses mostly continuous (rather than dummy) variables, with the result that fewer degrees of freedom are used up leading to more precise estimates. Additionally, the 2SLS method is argued to have more power to identify small but significant effects than VCA. However, the simultaneous equations analysis presented in the paper has drawbacks of its own as pointed out in later work (Hough, 2006). First, it requires that all corporations in the analysis have the same number of segments, potentially reducing the generalizability of the results. Second, the method requires the creation of an artificial segment 'position' classification as separate equations must be estimated for all segments assigned to be segment 1, segment 2 and so on. As a result the analysis produces a number of different estimates of the contribution of effects in explaining variance in performance,

[4]For example, McGahan and Porter (1997) show two specific paths in Table 5. In the first path (A), industry effects are calculated using the residuals of a previous regression in which only year dummies have been introduced. In (B) path the ordering is slightly different and industry effects are included in the third step, after consider consecutively year and corporate-parents effects.

which must then be aggregated.

Ruefli and Wiggins (2003) use a nonparametric approach to split Compustat profitability data at all levels into 3 performance groups that are on average statistically different from one another. Ordinal regression is then used to evaluate whether membership of certain performance categories at corporate or industry level is a significant predictor of performance category membership at the business segment level. Such a method is argued by the authors to make fewer assumptions than VCA or ANOVA methods, and thus more suitable for investigating the effects of managerial control on firm performance. The paper is critiqued in McGahan and Porter (2005); the salient objections are that the approach and the interpretation of it results requires no fewer and no more desirable assumptions than the VCA or ANOVA methods, and that the process of transforming continuous profitability data into categorical variables results in the loss of large amounts of information.[5]

A number of recent papers have used multilevel modelling techniques for variance decomposition (Hough, 2006; Misangyi et al., 2006; Short et al., 2007). This approach allows for the estimation of random effects variance components such as VCA, using superior maximum likelihood or restricted maximum likelihood estimation methods and specified multilevel error structures to derive variance component estimates (Hough, 2006). It also allows the researcher to include continuous covariates into the modelling framework, allowing for more specific hypotheses regarding drivers of firm profitability to be evaluated (see Misangyi et al. (2006) for an exploratory use of this capability).

[5]For the authors' response to the McGahan and Porter (2005) critique see Ruefli and Wiggins (2005).

2.3 The Shapley Value Method

In this paper we propose that in order to get more accurate results in the presence of covariance between effects, regressions corresponding to all possible models ranging from a null model to one including all effects should be run. The results of these regressions, specifically the marginal increase in adjusted R^2 due to effect inclusion, should then be aggregated into a single measure of effect importance in accounting for firm performance. To achieve this we use a concept from co-operative game theory, the Shapley value.

Consider year, country, group, industry and firm effects to be players in a co-operative game, with outcome defined as the proportion of the variance in firm profitability, the adjusted R^2, that is accounted for by the players jointly in all possible coalitions. The Shapley value of any given effect is defined as the weighted average of that effect's contribution to all possible coalitions. More formally,

$$S_j = \sum_M \gamma_n(M)[R^2(M \cup j) - R^2(M)]$$

Here, the Shapley value of effect j is the weighted sum of the differences in adjusted R^2 between a coalition of variables including j, (*i.e.*, $M \cup j$), and the same coalition which includes the other effects in it but not j , (*i.e.*, M) . The weight assigned to the increase in adjusted R^2 due to the introduction of effect j into the coalition is

$$\gamma_n(M) = \frac{m!(n-m-1)!}{n!}$$

where n is the total number of effects being considered in the study and m is the number of effects present in the M^{th} coalition, excluding j.

The Shapley value approach consists of calculating all possible increases in adjusted R^2 due to the inclusion of an effect, assigning a weight to each of these marginal increases, and adding up these weighted results to produce an overall estimated contribution of the effect to explaining the variance in firm profitability.[6] Doing this for all effects allows each effect an equal chance to contribute to R^2 in all possible paths from the null to the full model, thus giving each effect an equal chance to claim any covariance between effects. This technique allows for more accurate estimation of the effect's contribution to overall profitability variance in the presence of covariance between effects than is possible using approaches adopted in extant work.

2.4 Simulation

To examine the performance of the Shapley value method in comparison to ANOVA and VCA, we use a Monte Carlo simulation approach. We generate a dataset of industry, corporate parent, business segment and year effects with defined variance and covariance structures and apply different methods to extract results that can be compared to the data generating process. Our full model is :

$$r_{ikt} = \mu + \alpha_i + \beta_k + \gamma_t + \phi_{ik} + \varepsilon_{ikt}$$

where r_{ikt} is the profitability of corporate parent k's business unit in industry i, μ is the overall average profitability, α_i and β_k are industry and corporate parent dummy variables that are correlated with each other, φ_{ik} are business segment dummy variables (n interaction of α_i and β_k), γ_t are year dummy variables and ε_{ikt} is a normally distributed error term which is uncorrelated with any of the effects. We

[6]See Lipovetsky and Conklin (2001) for an example of the application of the Shapley Value approach in a marketing context.

evaluate the performance of the different methods by comparing the proportions of total variance assigned to the effects by the methods, with the true values implicit in the parameters used in our data generating process.

Consider the variance decomposition:

$$Var(r_{ikt}) = Var(\alpha_i) + Var(\beta_k) + Var(\gamma_t) + Var(\phi_{ik}) + 2Cov(\alpha_i, \beta_k) + Var(\varepsilon_{ikt})$$

If we know the variance and covariance parameters of the data generating process we can calculate the true proportions of total variance attributable to each effect. For year and segment effects these are $\frac{Var(\Upsilon_t)}{Var(r_{ikt})}$ and $\frac{Var(\phi_{ik})}{Var(r_{ikt})}$ respectively. To calculate the true proportion of total variance attributable to the industry and corporate parent effects we must divide their covariance between them. The fairest way to do so is to split the covariance term evenly between the two effects. Thus the true proportions of total variance attributable to industry and corporate parent effects are $\frac{Var(\alpha_i)+Cov(\alpha_i,\beta_k)}{Var(r_{ikt})}$ and $\frac{Var(\beta_k)+Cov(\alpha_i,\beta_k)}{Var(r_{ikt})}$ respectively. Having calculated these proportions, we can compare them to the proportions of total variance attributable to each effect as estimated by each variance decomposition method in the literature, including Shapley value. The method most suitable for this type of analysis will be the one that produces estimates closest to the true values.

We generate data for 500 corporate parents operating in 250 industries over 4 years, with each corporate parent operating in two industries, resulting in 1,000 business segments in the following way. First we draw 250 industry effects and 250 average corporate parent effects within an industry from a bivariate normal distribution with mean zero and standard deviation 5 for both effects, and a fixed correlation between them. To examine the accuracy of the methods when used on data with different correlation structures we allow the correlation in the data generating process to vary from 0 to 0.9.

Next we generate 2 individual corporate parent effects for each industry by adding a normally distributed error term with mean 0 and variance 1 to the average corporate parent effects. We now have 500 corporate parents, each operating in a primary industry. To assign a secondary industry to each corporate parent while maintaining the correlation structure between industry and corporate parent effects, we generate a hypothetical secondary industry effect for each corporate parent by adding an error term drawn randomly from a standard normal distribution to their primary industry effect. We then match this hypothetical secondary industry effect to the closest existing industry effect that is different from the corporate parent's primary industry effect. The data now consists of 1,000 observations of corporate parent profitability in 250 industries, making up 1,000 business segments.

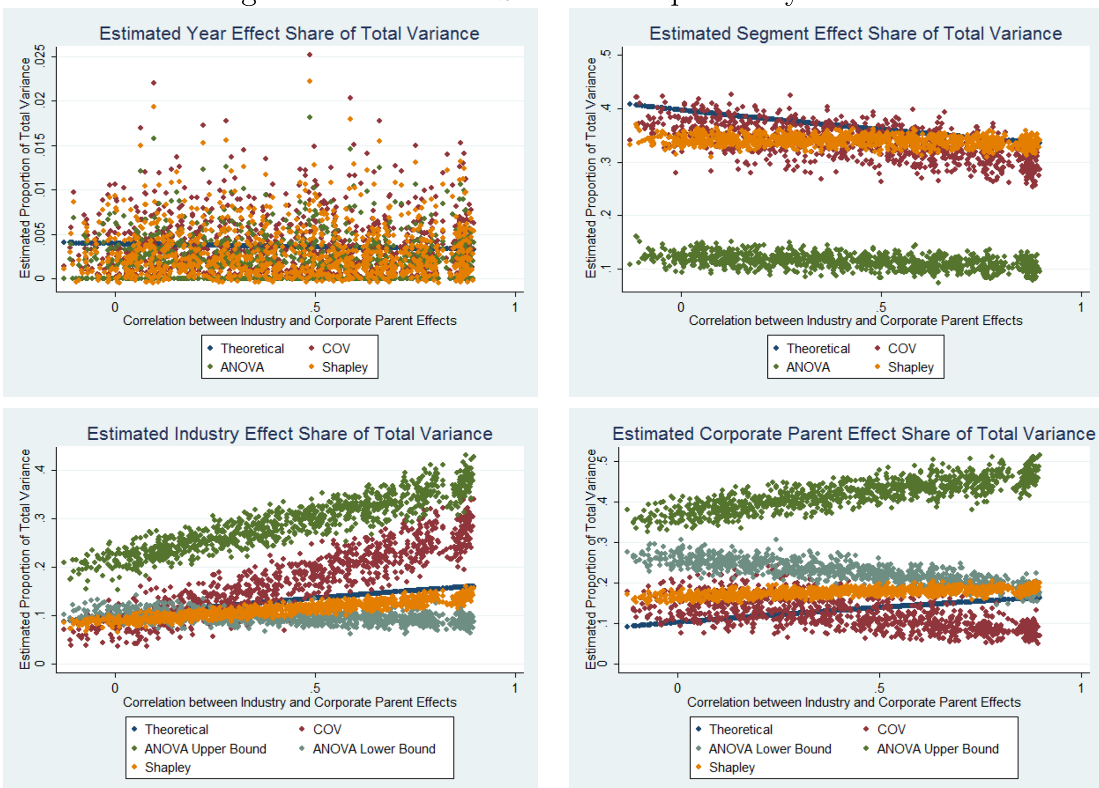
We next generate the business segment effects by drawing these from a normal distribution with mean 0 and variance 100. The business segment effects are thus drawn independently from the industry or corporate parent effects. To construct the year effects we create an additional 3 copies of the dataset and assign a year effect drawn randomly from a normal distribution with mean 0 and variance 1 to each of the 4 copies of the data. The final step is to construct the profitability measure as the sum of industry, corporate parent, business segment and year effects, plus a normally distributed error term with mean zero and variance 100. We run the data generating process 100 times for each value of the correlation between industry and corporate parent effects ranging from 0 to 0.9 in steps of 0.1, making 1,000 runs of the simulated data generating process in total.

Given the parameters of our data generating process, we can calculate the components of the variance decomposition. $Var(\alpha_i) = 25$, $Var(\beta_k) \approx 26$ (variance of the average corporate parent effect plus variance of the error term used to generate individual corporate parent effects, plus some small distortion due to matching process used to select second industry), $Var(\gamma_t) = 1$, $Var(\phi_{ik}) = 100$, $2Cov(\alpha_i, \beta_k) = 2 * \rho(\alpha_i, \beta_k) * \sigma(\alpha_i) * \sigma(\beta_k) = 2 * \rho(\alpha_i, \beta_k) * 5 * \sqrt{26}$ and $Var(\varepsilon_{ikt}) = 100$.

In addition to estimates from the Shapley Value method we also examine estimates from the ANOVA and VCA approaches. For the ANOVA approach we take 2 paths from null to full model in the spirit of McGahan and Porter (1997), introducing the effects in the orders year, industry, corporate parent, segment; and year, corporate parent, industry, segment. For industry and corporate parent effects we present the ANOVA estimates from both paths as upper and lower bound ANOVA estimates. For the VCA analysis we use restricted maximum likelihood estimation rather than expected sum-of-squares calculations as used in earlier papers (e.g. Rumelt, 1991; Schmalensee, 1985) as it produces estimates with superior properties (Hough, 2006). Our VCA analysis is thus also similar to a multilevel variance components estimate without covariates.

The simulation results can be seen in Figure 1 on the following page. The ANOVA method appears to significantly underestimate segment effects compared to its true contribution to variance. Additionally, the ANOVA upper and lower bounds diverge significantly both from each other and the true values for industry and corporate effects. The VCA method results (labelled COV on the graphs) are far superior to the ANOVA for the segment effect but appear to deviate significantly from the theoretical values of the industry and corporate parent effects with increasing correlation between the two. In contrast to the results of using these approaches, the Shapley Value estimate of proportion of total variance attributable to each effect appears to be consistently close to the true value for every effect. These simulation results support our claims that the Shapley Value approach is appropriate for use in variance decomposition and that it is likely to provide more accurate estimates of effect

Figure 1: Estimated Shares of Explanatory Power



contribution to variance in profitability than frequently used alternative methods.

3 Data and Sampling

We use the full version of the Amadeus database collected and maintained by Bureau van Dijk. The database contains balance sheet and additional data for about 14 million firms in forty European countries.[7] Several profitability measures are available and we use return on assets (ROA) as our measure for comparability with previous studies.[8] We restrict the data to those firms that provide full information on ROA,

[7]Amadeus database currently covers all European Union countries, Belarus, Bosnia-Herzegovina, Croatia, Iceland, Liechtenstein, Macedonia, Moldova, Montenegro, Norway, Russian Federation, Serbia, Switzerland and Ukraine.

[8]Hawawini, Subramanian, and Verdin (2003) use value-based measures of performance as well as ROA in their analysis and find the results to be similar.

industrial classification of their activity and the number of employees for the period 2001-2006, thus creating a balanced panel.[9] Although it is possible to include more recent data, this would result in the loss of information for some countries. This six-year period is slightly smaller than Roquebert, Phillips, and Westfall (1996), who used a 7 year-period, but larger than Rumelt (1991), who used a 4-years series, and similar to McGahan and Porter (1997, 2002), in which the average time series of each economic unit was 5.7 years in a unbalanced panel for a 14-years period. Industries are defined at four-digits NACE, a level comparable to four-digit SIC classification used in most studies of US data, and a firm's country is defined as the country in which it is located and reports.

3.1 The Construction of Business Groups

We have mentioned earlier some characteristics of the Amadeus database that differentiate it from the Compustat and FTC databases. Our basic statistical unit is the firm. We have information about firms' subsidiaries (other firms), if they exist, and the corresponding shares of ownership. Using this information we define a business group as the set of companies connected by ownership links. Given that the on-line version of the Amadeus database only offers a snapshot of the most recent available information on ownership structure (the last wave), we extracted and used seven waves (one a year) to measure the shape of the business groups more precisely.

We define an ownership link when the *main* firm has ownership of more than 50% of a *subsidiary* firm.[10] An additional condition imposed is that ownership links greater than 50% should be observed for at least two years. We identify 887,443

[9] The information on the number of employees is required because we control for size-classes in the sampling procedure, as we next describe.

[10] In some cases there is no information about the percentage of ownership. We do not count these links in constructing our business groups.

pairs of different links surpassing the 50% threshold in the period 2002-2006.[11] The main firm can have one or more subsidiaries, and it can also be controlled (i.e. its shares can be owned) by some other firm. Thus the business group is defined as the network of paired firms connected by share ownership that surpasses the 50% threshold. This is similar to the definition of Korean business groups in Chang and Honj (2002), though the chaebols definition uses a 30% threshold.[12]

We use information on the industrial code for each firm (main or subsidiary), defined by the 4-digits NACE rev 1.1 classification. In doing this we lose some links in which the subsidiary firm is not included in the Amadeus database. This mostly applies to non-European subsidiary firms. We also drop some industries that increase the size of the network, but are purely instrumental companies. In particular, we drop such firms in financial industries (divisions 65 to 67), two particular business-service industries linked to financial services (7415 and 7487), and non-service sectors.[13] The final number of links thus fall to 450,782, between a total of 628,055 firms. Of these firms 28.7% are solely *main* firms, 66.1% are solely *subsidiaries* and the remaining 5.1% are simultaneously *main* (they have at least one subsidiary) and *subsidiary* (they have one main firm) firms.

We use an algorithm to define business groups as networks of connected links. The total number of identified business groups is 179,089. Table 1 shows the size distribution of business groups (i.e., networks) according to the number of links.

[11] This represents a 42.1% of initially observed 2,107,422 pairs of firms linked by some degree of ownership. If we do not impose the condition that ownership links above 50% should be observed for at least two years the number of such links increases to 1.4 million.

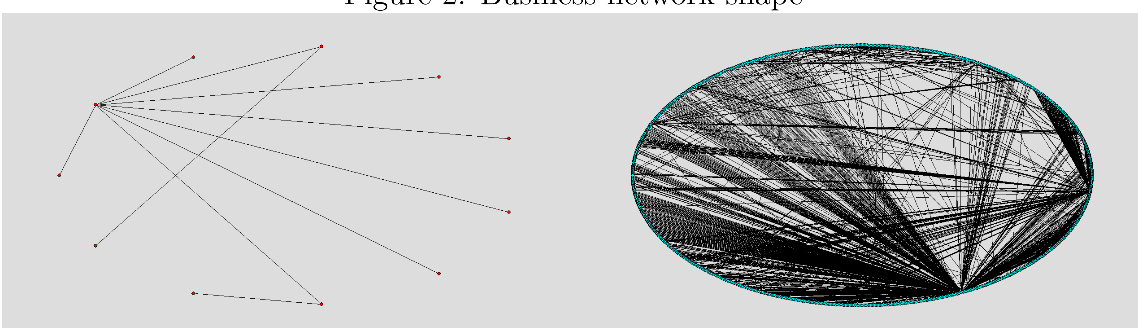
[12] Chang and Hong (2002) obtain a strong effect for Korean business groups (*chaebols*) to explain the variance of profits. The economic relevance of these conglomerates (40 percent of Korea's total output in 1996) is unusual when compared to other national economies. Our analysis for a wide set of European countries allows us to include a heterogeneous set of institutional settings, though no European economy has a structure similar to the Korean one.

[13] This is similar to the exclusion of depositary institutions in papers using the Compustat database.

Table 1: Business Group Size (Number of Observed Links)

Number of links in network	Number of observed links	Percentage
1	113,626	25.21
2	59,964	13.3
3	37,131	8.24
4	26,524	5.88
5	20,275	4.5
6	15,390	3.41
7	11,998	2.66
8	10,648	2.36
9	9,153	2.03
10 or more	146,073	32.00
Total	450,782	100

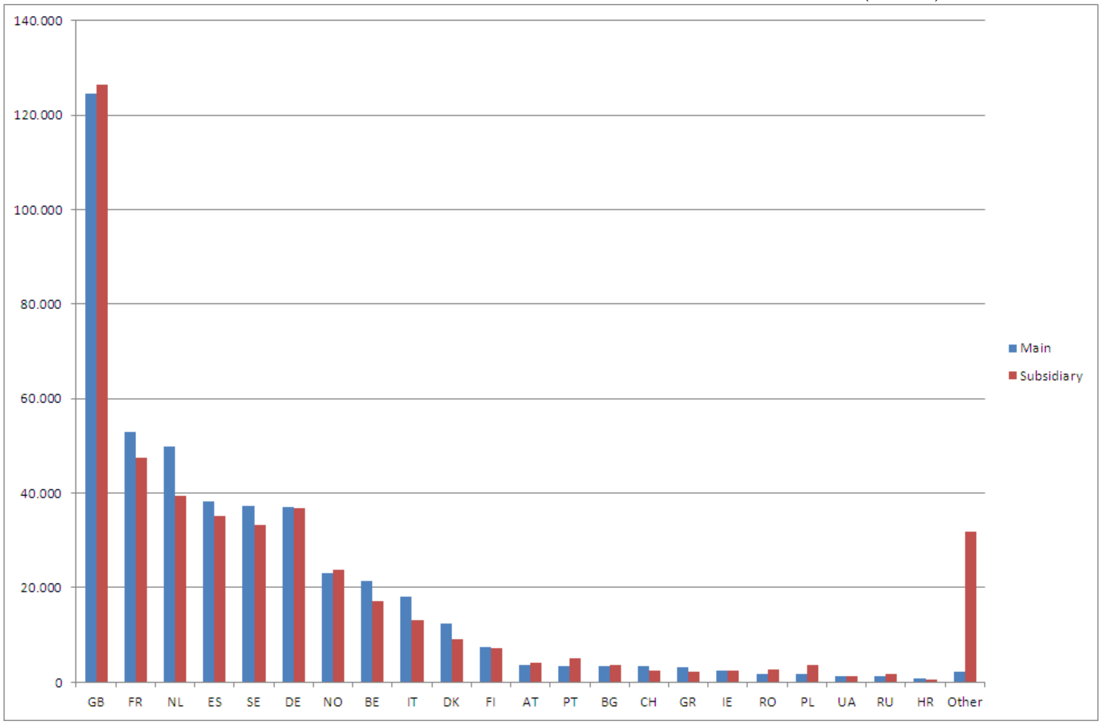
Figure 2: Business network shape



As can be observed, the majority of business groups are constituted by only one main and one subsidiary firm. These one-link groups account for 25.2% of the all observed links. For almost half of observed links, the relevant business group is constituted by three links or fewer, with the average business group consisting of 2.5 links. The distribution of bigger groups is fairly smooth with the biggest business group consisting of 1,096 links. The shape of the business group depends on the specific structure of links in the network. Figure 2 shows the shape of two business groups. The first (left panel) corresponds to a randomly chosen network with 10 links, and the second (right panel) corresponds to the biggest business group.

The majority of the links (83.5%) occur between two firms that are located in the

Figure 3: Country of the main and subsidiary firm (links)



same country, with this being most likely in single link groups. 92.0% of single link business groups have the subsidiary in the same country as the main firm. Overall, 66.1% of groups have all their firms in the same country. The probability that linked firms are located in different countries increases with business group size as expected. The biggest group, consisting of 1,096 links, has 712 links in which the main and the subsidiary firms are located in different countries. Figure 3 displays the numbers of main and subsidiary firms with links across countries. The average size of multinational groups (business groups in which at least one firm is located in a different country) is 6.2 links, larger than the average 2.5 links across all groups.

Finally, although these business groups are identified as networks of firms, they do not capture strategic networks. These 'represent an attempt to achieve shared goals through collective efforts by multiple participants, each of which also have their own strategic interests that are not necessarily always aligned' (Wincent et al., 2010,

p. 599). The analysis of strategic networks belongs to a related literature that has followed a different approach, usually focusing on a small sets of networks in specific industries.

3.2 The Sampling Procedure

We adopt a re-sampling procedure whereby a large number of different samples of firms are drawn from the Amadeus database for analysis. The large size of the Amadeus database makes it computationally impossible to use all the available data with firm dummy variables in single regressions. An additional reason for the re-sampling method is that Amadeus data is not representative of the population of firms across countries, industries or size-classes. This raises questions about whether the results of analyses performed using Amadeus data generalize to the whole population of firms.[14] In particular, it is well-known that the distribution of firms included in the Amadeus database has a bias towards larger sized firms. As far as we know, prior research using the Amadeus database have not attempted to correct for this.

To draw samples that are representative of the underlying population of firms we use a stratified random sampling procedure with stratification by country, industry and size-class. We fix the stratification criteria for sampling using the Structural Business Statistics (SBS) database maintained by the Statistical Office of the European Commission (Eurostat) (and freely available from its web server). The SBS provides information on the number of firms in each EU-27 country and Norway according to industry (NACE Revision 1.1) and size-classes (<10, 10-19, 20-49, 50-249, 250 and more employees). Use of the SBS structure to define our sampling procedure results in the restriction of the Amadeus data we use to firms in European Union

[14] As mentioned previously, papers based on the Compustat database have been careful to point out that the firms covered by Compustat are not representative of the population of US firms in general.

Table 2: Sample summary statistics

	Average	Minimum	Maximum
Countries	25	25	25
Industries	415	400	435
Business groups (networks)	1072	1052	1089
Countries per group	1.20	1	7
Industries per group	1.80	1	9
Firms per group	2.11	2	14
Firms in business groups	2504	2495	2505

countries and Norway.

The final stage in our sampling procedure is the selection of firms with and without business group membership. Half of each of our samples are drawn from the population of business group members. In order to be able to distinguish business group and firm effects in these draws, we drop all those business groups for which only one firm has been included in the sample. The other half of each of our samples are drawn from the population of firms that are not members of business groups. As the construction of the business groups is done before sampling, we are able to identify cases of firms belonging to the same business group even when they are linked through firms that are not included in the sample.[15]

We carry out the Shapley Value analysis on 100 samples, each consisting of 5000 firms. Table 2 provides summary statistics relating to the number of industries, business groups and business group composition across all our samples.

[15]For example, a German firm could have a subsidiary in Russia, and this firm in turn could have a subsidiary in Sweden. Even though Russian firms are not included in our sampling procedure, both the German and Swedish firms would be considered as belonging to the same business group.

4 Results

Table 3 on the next page presents the results obtained using three levels of aggregation for industry effects: two, three and four-digits industries in the NACE rev 1.1 classfication. This provides us a useful sense of the sensitivity of results to the level of industrial aggregation. The figures presented in the table are Shapley Values for each effect - the weighted average of the effect's contribution to model adjusted-R^2 in all possible paths from the null to the full regression model.

As can be seen, the level of industry aggregation affects mainly the size of the industry effect. The Shapley Value for industry effects is 2.5 time greater when using four-digit NACE industries than when using 2-digit NACE industries. Keeping this in mind and focussing on the four-digit NACE results for comparability with previous work, we note that firm effects constitute the most important component in accounting for variance in firm profitability with a Shapley Value of 38.6%. Business group effects appear to be second in importance, accounting for 6% of the variance in firm profitability. Industry effects appear to account for only 2.8% of variance in firm profitability, an estimate that is significantly lower than previous estimates of industry effects using US data. Finally, both country and year effects appear to be of little importance, accounting only for 0.5% and 0.1% of variance in firm accounting profitability respectively.

Most of the prior empirical analyses of micro-data focuses on the manufacturing sector, for data availability reasons. We are in position to depart from this. The results in Table 3 are for firms across all sectors, with the exception of the small number of previosly mentioned industries that are excluded. Some previous papers have compared the results of the analysis for firms in manufacturing and service sectors (e.g. McGahan and Porter, 1997, 2002). We follow suit with Table 4 that

Table 3: Contribution to Explanatory Power by Effect Type (Average of Shapley Values for 100 Samples)

	2-digit industries	3-digit industries	4-digit industries
Country	0.0059	0.0059	0.0058
Industry	0.0120	0.0187	0.0284
Group	0.0601	0.0599	0.0601
Year	0.0007	0.0007	0.0007
Firm	0.4018	0.3953	0.3856
Total	0.4805	0.4805	0.4805

Table 4: Contribution to Explanatory Power by Effect Type: Manufacturing vs. Services (Average of Shapley Values for 100 Samples)

	Manufacturing	Services
Country	0.0054	0.0068
Industry	0.0217	0.0249
Group	0.0529	0.0662
Year	0.0010	0.0016
Firm	0.4276	0.3997
Total	0.5086	0.4991

reports results for samples from manufacturing and from services industries. To keep sample sizes similar to those used in the general analysis we draw two new sets of 100 samples each of 5000 firms, one for manufacturing industries and the other for services. As earlier, half the sampled firms belong to business groups, and the industry classification is four-digit NACE.

As can be seen, all types of effects other that firm effects appear to contribute more to overall variance in firm profitability inthe services sector than in the manufacturing sector. The difference in Shapley Values of the group effect is the largest: group Shapley Value in the services sector is estimated at 6.6%, and in the manufacturing sector, 5.3%.

Finally, Roquebert, Phillips, and Westfall (1996) and McGahan and Porter (2002) report that the size of the corporate effect decreases with increasing diversification of firms. In the context of the Compustat database, samples with increased diversification are generated by limiting the sample to corporations with increasing numbers of business segments (McGahan and Porter (2002) present results for all corporations, those with at least 2 segments and those with at least 3 segments). We carry out the Shapley Value analysis on samples which include business groups with at least two links and compare the results to that from the analysis of samples including business groups of all sizes. The results are in Table 5. As expected, and consistent with previous work, the size of the group effect decreases marginally with the imposition of the restriction that business groups must consist of two or more links: the estimated Shapley Value drops from 6% to 5.9%. The size of the industry effect increases with this restriction, rising from 2.8% to 3.1%. There does not appear to be any significant change in the magnitudes of the country or firm effects. Year effects more than double but account for only 0.2% of variance in firm profitability.

Table 5: Contribution to Explanatory Power by Effect Type: Business Group Size (Average of Shapley Values for 100 Samples)

	All sizes	At least two links
Country	0.0058	0.0054
Industry	0.0284	0.0311
Group	0.0601	0.0590
Year	0.0007	0.0017
Firm	0.3856	0.3844
Total	0.4805	0.4816

5 Conclusion

This paper offers two main contributions to the literature analyzing the heterogeneity in firm-specific accounting profitability. First, in contrast the vast majority of previous work which has analyzed the profitability of listed US companies, we analyze firms operating in in 25 European countries. In order to be true to the underlying population of firms, we use a stratified (along size, industry and country dimensions) random sampling procedure with re-sampling. Second, we use a novel approach to variance decomposition - the Shapley Value method. The determination of weighted averages of the contributions of each effect across all possible model specifications, inspired by co-oprative game theory, helps us to deal with the issue of multicollinearity between explanatory variables that have plagued previous studies. We also show that this method outperforms both ANOVA and VCA methods widely used in the prior literature using Monte-Carlo simulations.

Our results point to the importance of business group effects. These effects are the second largest influence on firm profitability (Shapley Value between 5.9% and 6.6%) after firm effects (Shapley Value between 38.4% and 42.8%). The business group effect appears to decrease with the size of the business group. Our estimates of industry effects (between 2.2% and 3.1% of variance in firm profitability) are significantly smaller than those found in prior work. Differences across countries appear to have little relevance in explaining heterogeneity in firm profitability. Year effects are negligible. Manufacturing and services sectors differ: firm effects are smaller, and the other effects are larger in the services sector.

References

Bowman, Edward H., and Helfat, Constance E., 2001. "Does Corporate Strategy Matter?" *Strategic Management Journal* 22:1–23.

Brush, Thomas H, and Bromiley, Philip, 1997. "What Does a Small Corporate Effect Mean? A Variance Components Simulation of Corporate and Business Effects." *Strategic Management Journal* 18:825–835.

Brush, Thomas H, Bromiley, Philip, and Hendrickx, Margaretha, 1999. "The Relative Influence of Industry and Corporation on Business Segment Performance: An Alternative Estimate." *Strategic Management Journal* 20:519–547.

Chang, Sea Jin, and Hong, Jaebum, 2002. "How Much Does the Business Group Matter in Korea?" *Strategic Management Journal* 23:265–274.

Crossland, Craig, and Hambrick, Donald 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.

Hawawini, Gabriel, Subramanian, Venkat, and Verdin, Paul, 2003. "Is Performance Driven by Industry- or Firm-Specific Factors? A New Look at the Evidence." *Strategic Management Journal* 24:1–16.

Hough, Jill R., 2006. "Business Segment Performance Redux: A Multilevel Approach." *Strategic Management Journal* 27:45–61.

Lipovetsky, Stan, and Conklin, Michael, 2001. "Analysis of Regression in Game Theory Approach." *Applied Stochastic Models in Business and Industry* 17:319–330.

McGahan, Anita M., and Porter, Michael E., 1997. "How Much Does Industry Matter, Really?" *Strategic Management Journal* 18:15–30.

McGahan, Anita M., and Porter, Michael E., 2002. "What Do We Know About Variance in Accounting Profitability?" *Management Science* 48:834–851.

McGahan, Anita M., and Porter, Michael 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.

Misangyi, Vilmos F., Elms, Heather, Greckhamer, Thomas, and Lepine, Jeffrey 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.

Roquebert, Jaime A., Phillips, Robert L., and Westfall, Peter A., 1996. "Markets vs. Management: What 'Drives' Profitability?" *Strategic Management Journal* 17:653–664.

Ruefli, Timothy W., and Wiggins, Robert R., 2003. "Industry, Corporate, and Segment Effects and Business Performance: A Non-Parametric Approach." *Strategic Management Journal* 24:861–879.

Ruefli, Timothy W., and Wiggins, Robert R., 2005. "Response to McGahan and Porter's Commentary on 'Industry, Corporate, and Segment Effects and Business Performance: A Non-Parametric Approach'." *Strategic Management Journal* 26:881–886.

Rumelt, Richard P., 1991. "How Much Does Industry Matter?" *Strategic Management Journal* 12:167–185.

Schmalensee, Richard, 1985. "Do Markets Differ Much?" *American Economic Review* 75:341–351.

Short, Jeremy C., Ketchen, David J. Jr., Palmer, Timothy B., and Hult, G. Tomas M., 2007. "Firm, Strategic Group, and Industry Influences on Performance." *Strategic Management Journal* 28:147–167.

Wincent, Joakim, Anokhin, Sergey, Örtqvist, Daniel, and Autio, Erkko, 2010. "Quality Meets Structure: Generalized Reciprocity and Firm-Level Advantage in Strategic Networks." *Journal of Management Studies* 47:597–624.