Journal of Family Business Strategy 5 (2014) 52–62

Contents lists available at ScienceDirect

Journal of Family Business Strategy

journal homepage: www.elsevier.com/locate/jfbs

Looking forward: The role of multiple regression in family business research

Lucas Hopkins [a,*], Keith E. Ferguson [b]

[a] Middle Georgia State College, United States
[b] Grand Rapids Community College, United States

ARTICLE INFO

Keywords:
Regression analysis
Nonlinear regression
Research methods
Family business
Best practice

ABSTRACT

This article assesses the role of regression analysis in family business research. We discuss four specific types of regression (multiple linear, stepwise, hierarchical, and nonlinear) and review how each type has been used in prior family business research. In order to specify a systematic guide to using regression analysis, specific examples are provided using SPSS. The article concludes by providing several rules of thumb for the use of regression analysis.

1. Introduction

The recent increase in scholarly research focused on family business has amplified the need for a sound understanding of methodological opportunities and more in-depth statistical techniques (Wilson et al., 2014). The origins of family business research began with mostly conceptual contributions from academics and consulting-based insights from advisors (Bird, Welsch, Astrachan, & Pistrui, 2002). Early academic research on family business paid relatively little attention to research designs or sophisticated statistical techniques (Bird et al., 2002). As the field grew, so has the use of empirical methods to test theory (Correll, 1989; Whitmoyer et al., 2014). Since an increasing number of family business researchers utilize empirical techniques, a greater understanding of their uses is warranted as exemplified by this special issue.

One of the most prevalent methodologies in business research is multiple regression analysis (Whitmoyer et al., 2014). The appeal of this methodology is specifically relevant to research in small business and entrepreneurial ventures due to the simplicity with which the researcher can explore the functional relationships among a wide variety of variables of interest (Kellermanns, Eddleston, Barnett, & Pearson, 2008; Lee, 2006). The popularity of this methodology is evident in an analysis of 665 articles published in three of the leading small business journals between 2001 and 2008 (Mullen, Budeva, & Doney, 2009), which shows that 154 (39%) studies used regression analysis.

Given the growth of the family business field and the popularity of regression analysis, it is important for family business researchers to understand and be prepared to apply the numerous options that are available in regression analysis (Bird et al., 2002). This article provides a comprehensive direction to achieve this goal. First, we discuss four types of regression analysis: traditional (linear) multiple regression, hierarchical regression, stepwise regression, and nonlinear regression. Next, we include a systematic explanation of one of the less often used approaches in family business research – nonlinear regression analysis. Last, we offer best practices for researchers using different types of regression analysis. In addition to discussing the assumptions associated with regression analysis, we also offer several rules of thumb when dealing with issues such as sample size selection and model estimation.

2. Theoretical overview

2.1. What is regression analysis?

Regression analysis is a statistical technique used to explore the relationship between metrically measured independent and dependent variables (Cohen & Cohen, 1983). In its simplest form, the appropriate use of regression analysis involves hypothesizing a relationship between a single dependent and a single independent variable, applying the method to examine the relationship, and using fit statistics to evaluate the hypothesized relationship (Sheather, 2009). The inclusion of dummy coded independent variables, which are used to distinguish between two and three groups or conditions (like family and non-family businesses), in traditional regression analysis is also possible. In addition to relying on fit statistics, the researcher also investigates the extent

* Corresponding author. Tel.: +1 4787187918.
E-mail address: lucas.hopkins@maconstate.edu (L. Hopkins).

1877-8585/$ – see front matter © 2014 Elsevier Ltd. All rights reserved.
http://dx.doi.org/10.1016/j.jfbs.2014.01.008

to which the hypothesized relationships between the dependent and independent variables can be accepted (i.e., are statistically significant and meaningful). Most often used for prediction and forecasting, regression analysis has become one of the most widely used statistical tools for analyzing data involving multiple variables (Chatterjee & Hadi, 2012).

Common throughout many business and social science disciplines, regression analysis has been used in a variety of family business studies including family ownership transitions (Morris, Williams, Allen, & Avila, 1997), entrepreneurial orientation (Naldi, Nordqvist, Sjöberg, & Wiklund, 2007), and agency cost differentials between family and non-family firms (Chrisman, Chua, & Litz, 2004). Miller, Le Breton-Miller, Lester, and Cannella (2007) used regression analysis in their study evaluating performance of family firms based on definition of ownership (lone founder with no family involvement, family owned with family involvement, and a broad sample of random firms) because their objective was to apply the most conservative and appropriate techniques for each type of variable. The authors tested 26 variables (types of ownership, shares owned, Tobin's q, firm growth and size, investment, research and development, various economic ratios, and market risk) and found that family owned firms did not outperform similar non-family owned firms, whereas lone founder firms outperformed both family owned firms and other businesses. The authors noted that the definition of a family business significantly impacted the related performance findings. Regression analysis was also used by Sciascia, Mazzola, Astrachan, and Pieper (2013) to explore the relationship between family involvement in the board of directors and sales internationalization. The method was selected because the data included an ordinal dependent variable (sales internationalization as measured by the percent of sales generated from international markets) and a metric independent variable (family involvement in the board of directors as measured by percentage of directors belonging to the controlling family).[1]

Two of the most frequently cited articles in family business research using regression analysis are Schulze, Lubatkin, Dino, and Buchholtz (2001) and Anderson and Reeb (2003). A search in Google Scholar shows 1038 cites for Schulze et al. (2001) and 2053 cites for Anderson and Reeb (2003). The high citation rate of these two articles was also indicated in a literature of review of highly cited articles in journals featuring family business research (Chrisman, Kellermanns, Chan, & Liano, 2010). Schulze et al. (2001) used regression analysis to explore the agency costs in owner-managed firms. The authors relied on this methodology because the single dependent variable family firm performance as assessed using sales growth was metric and there were over 15 independent variables (family and non-family pay incentives, family ownership and number of family members, strategic planning, CEO tenure, number of outside board members, and transfer intentions), many of which were measured categorically. The application of regression analysis allowed the researchers to predict that family firms experience agency costs due to lack of self-control, altruism, and conflict of interest between owning family members. Anderson and Reeb (2003) also used regression analysis in their study of the relationship between founding-family ownership and firm performance. Their article offers an excellent perspective on the versatility of regression analysis as a variety of regression techniques were used, including instrumental–variable regressions, Fama–MacBeth regressions, and pooled, time series average regression.

The most widely used approach to conducting a multiple regression analysis is ordinary least squares (Wang & Jain, 2003), typically referred to as OLS regression. Ordinary least squares estimates the parameters in a linear model by minimizing the vertical distances between responses that are observed and the responses that are predicted by the linear estimate (Dismuke & Lindrooth, 2006). One result of minimizing the error between responses is that OLS provides the researcher with a more accurate view of the relationship between the independent variables (IVs) and dependent variable (DV).

The most commonly reported regression parameters, and the ones that must be interpreted to understand the meaning of the analysis, are the R^2, the adjusted R^2, the F test, and the standardized beta coefficients. The coefficient of determination, R^2, is a measure of the how well the independent variable explains the variance of the dependent variable. For example, an R^2 of 1.0 means that the variability of the dependent variable is 100% explained by the independent variable; or, in other words, you only need the independent variable to perfectly predict the dependent variable. Whereas the probability of achieving a model with a R^2 of 1 is highly unlikely, the research goal should be to create a model that achieves an R^2 as close to 1 as possible. One way to increase the coefficient of determination is to include additional independent variables. While adding additional independent variables will increase the R^2, when evaluating regression models researchers must also ensure that the added independent variables are meaningful based on the significance and size of the beta weights (Hair, Black, Babin, & Anderson, 2010). Researchers should be careful when adding additional independent variables, as too many independent variables may cause an issue with the modeling of random noise and reduction in the ability to make valid predictions. To avoid this issue, the researcher should employ the adjusted R^2 to determine if the model is indeed accurate and not just apparently so due to the number predictor variables (IVs). The adjusted R^2 considers the number of explanatory variables and only increases if the new IV improves the R^2 more than had it not been included (Draper, Smith, & Pownell, 1966).

When working with multiple regression, the test statistic that is required is called an F-test. The F-score is calculated by dividing the explained variance by the unexplained variance to provide a score that represents the statistical significance of the entire regression equation (Myers, 1990). While a larger F-score is desired, the researcher must not neglect the statistical significance of the F-test. Moreover, if the F-test fails to reach statistical significance, the magnitude of the F-score is irrelevant. Finally, when working with a regression equation, the researcher should note the strength of the relationship between the IV(s) and DV. To allow for this comparison, standardized beta coefficients can be calculated by multiplying the estimated coefficient by the standard deviation of the independent variable, then dividing it by the standard deviation of the dependent variable. The standardized coefficient is then measured in units of standard deviation to assess how strongly each independent variable influences the dependent variable (Schroeder and Stephan, 1986).

Although the core concepts of regression analysis are relatively straightforward, variations of regression analysis are available to help researchers address more specific questions and phenomena. The flexibility and adaptability of regression analysis enables it to be used with most dependence relationships (Hair et al., 2010), thus increasing the importance of choosing the correct variation of this statistical technique. When making this decision, the correct estimation technique is most often contingent on the nature of the independent variables (IVs). For example, while the dependent and independent variables used in a study are based on theory, the researcher often wants to determine the order in which the IVs are entered into the regression equation. As a result, one question faced by researchers is whether to enter the variables simultaneously or sequentially. To further explore these questions, the

[1] The SPSS software includes an option specifically designed to handle an ordinally measured dependent variable.