Industrial Marketing Management xxx (xxxx) xxx–xxx

Contents lists available at ScienceDirect

Industrial Marketing Management

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

Endogeneity bias in marketing research: Problem, causes and remedies

Ghasem Zaefarian[a,*], Vita Kadile[a], Stephan C. Henneberg[b], Alexander Leischnig[c]

[a] University of Leeds, Leeds University Business School, Maurice Keyworth Building, Leeds LS2 9JT, United Kingdom
[b] Queen Mary University of London, School of Business and Management, Business Ecosystems Research Group, London E1 4NS, United Kingdom
[c] University of Bamberg, Feldkirchenstraße 21, 96052 Bamberg, Germany

A B S T R A C T

Endogeneity bias represents a critical issue for the analysis of cause and effect relationships. Although the existence of endogeneity can produce severely biased results, it has hitherto received only limited attention from researchers in marketing and related disciplines. Thus, this article aims to sensitize researchers intending to publish in the *Industrial Marketing Management* (IMM) journal to the topic of endogeneity. It outlines the problem of endogeneity bias, and provides an overview of potential sources, i.e. omission of variables, errors-in-variables, and simultaneous causality. Furthermore, the article shows ways to deal with endogeneity, including techniques based on instrumental variables as well as instrument-free approaches. Our methodological contribution relates to providing researchers aiming to publish in IMM with an initial overview of the causes of and remedies for endogeneity bias, which should be considered in designing research projects as well as when analysing data to obtain insights into cause and effect relationships (causal models).

1. Introduction

An increasing number of articles in marketing as well as in related fields such as international business, supply chain management, and operations management have recently pointed to issues associated with endogeneity (Guide & Ketokivi, 2015; Jean, Deng, Kim, & Yuan, 2016; Shugan, 2004). Endogeneity constitutes a critical problem for research as it compromises key conditions for claiming causality (Antonakis, Bendahan, Jacquart, & Lalive, 2010, 2014) and both the direction and the size of its bias are difficult to predict in advance (Hamilton & Nickerson, 2003). A failure to consider and correct for endogeneity in research practice can lead to biased and inaccurate results, and poses the risk of drawing incorrect conclusions about cause and effect relationships between concepts of interest. Even though the issue is much more predominant in naturally occurring data (e.g. regularly and automatically collected customer data at the point of purchase or via web browsing) as opposed to market research data (e.g. data collected through survey questionnaires), and is less of a problem for experimental data (e.g. Anderson & Simester, 2004), any study involving questionnaire or survey design is potentially subject to endogeneity bias (Toubia, Simester, Hauser, & Dahan, 2003).

Endogeneity is most commonly described in the context of ordinary least squares (OLS) estimation, and refers to a situation in which an independent (explanatory) variable correlates with the structural error term (also referred to as 'disturbance term' or 'residual') in a model (Kennedy, 2008; Wooldridge, 2002). In such a situation, the error term is not random and the estimation is inconsistent, which implies that the coefficient estimate of the independent variable fails to converge to the true value of the coefficient in the population as sample size increases. When an independent variable correlates with the error term, the coefficient estimate includes the effect of the respective independent variable on the dependent variable *as well as* the effects of all unobserved factors that correlate with the independent variable and explain the dependent variable, thus rendering its interpretation problematic, or even useless (Antonakis et al., 2010, 2014). If this correlation is ignored, the estimated effect of the observed variable is likely to be biased. This bias is referred to as the *endogeneity bias* (Chintagunta, Erdem, Rossi, & Wedel, 2006).

Endogeneity is a major concern in many areas of marketing and related research, which rely on employing regression-based analyses with the aim to draw causal inferences (Jean et al., 2016). In essence, endogeneity may affect the causal inferences that researchers make with regard to the hypothesized associations between variables, and failure to account for this may lead to spurious findings resulting in misleading theoretical as well as managerial implications (Semadeni, Withers, & Certo, 2014). Against this background, editors and reviewers of various disciplines in the area of management studies increasingly point to endogeneity as a likely alternative explanation for results

* Corresponding author.
E-mail addresses: G.Zaefarian@leeds.ac.uk (G. Zaefarian), V.Kadile@leeds.ac.uk (V. Kadile), S.Henneberg@qmul.ac.uk (S.C. Henneberg), Alexander.Leischnig@uni-bamberg.de (A. Leischnig).

http://dx.doi.org/10.1016/j.indmarman.2017.05.006
Received 15 November 2016; Received in revised form 12 May 2017; Accepted 24 May 2017

provided in manuscripts they process, and therefore endogeneity considerations become more and more of a (contributing) reason for manuscript rejection (e.g. Guide & Ketokivi, 2015; Larcker & Rusticus, 2010; Shugan, 2004). In spite of the fact that several approaches to address endogeneity have been available for almost three decades, only fairly recently have some of these remedies been applied in studies published in marketing journals (Hamilton & Nickerson, 2003), and the number of researchers proactively correcting for endogeneity still remains very low.

The *Industrial Marketing Management* (IMM) journal has made significant theoretical and empirical contributions to the field of industrial and B2B marketing, as well as supply chain management research. In many respects, the articles published in IMM are rigorous in terms of method, e.g. by assessing several sources of bias such as non-response and common method variance, and by incorporating measurement validity and reliability analyses. However, the issue of endogeneity arguably is a blind spot that has not been sufficiently addressed in research published in IMM to date. So far we have only found a dozen or so papers published in IMM that tackle the issue of endogeneity in their empirical analyses, with the first study being published by Streukens, Hoesel, and de Ruyter (2011). We therefore believe that it is timely for the IMM research community to take the issue of endogeneity seriously. Hence the objective of our paper is to *sensitize researchers and introduce an outline for diagnosing and correcting for potential endogeneity bias in marketing research.* We discuss potential sources of endogeneity and provide a brief overview of techniques available to account for it, followed by an assessment of their robustness. These considerations ought to provide suggestions for researchers in the field of marketing and supply chain management, and especially for future publications in IMM that examine cause and effect relationships.

Our paper thus contributes to the existing knowledge on endogeneity in two ways. First, we clarify the notion of endogeneity and its sources using marketing-related examples. Second, we emphasize the importance of accounting for endogeneity in marketing studies and provide an overview of remedies available to treat endogeneity bias. Overall, we aim at sensitizing researchers who aim at publishing in IMM to the hitherto somewhat neglected topic of endogeneity bias.

2. Sources of endogeneity

Literature emphasizes three primary instances where the condition of exogeneity becomes violated and therefore endogeneity occurs: omission of variables, errors-in-variables, and simultaneous causality (Wooldridge, 2002). The following subsections briefly outline the problems associated with each of these sources of endogeneity.

2.1. Omission of variables

Endogeneity may occur due to the omission of variables in a model. Omission of variables is usually attributable to data unavailability and can result in a violation of the exogeneity assumption if the omitted variable that is associated with the dependent variable is also correlated with any of the independent variables under investigation (Kennedy, 2008; Wooldridge, 2002). In such a situation, the error term will be correlated and the coefficient estimator of the independent variables will be biased. For instance, in investigating the effect of firm resources on foreign market entry modes, other variables that may affect both firms' resource slack and foreign market entry mode include managerial experience and market characteristics. If such variables are omitted from the model and thus not considered in the analysis, the variations caused by them will be captured by the error term in the model, thus producing endogeneity problems.

A common form of omitted-variable-based endogeneity is omitting selection (e.g. Antonakis et al., 2010; Clougherty, Duso, & Muck, 2016; Wooldridge, 2002). This problem arises when respondents self-select into treatment and non-treatment groups based on unobserved factors that correlate with the dependent and the independent variables under investigation (this is also called the 'choice problem'), which leads to a situation in which the dependent variable is observable for different parts of the sample on a nonrandom basis (Clougherty et al., 2016). Prior work shows that many business phenomena are subject to such self-selection-based endogeneity as they involve organizational choices that are endogenous and self-selected (Hamilton & Nickerson, 2003; Shaver, 1998). For example, firms may select a particular relationship governance mechanism (e.g. formal vs. informal) to achieve a high relationship performance with partner firms based on factors that are unobserved. These factors may, for example, include the level of trust in the partner or the relationship phase. An analysis that tests the effect of relationship governance mechanism on relationship performance will most likely yield biased coefficient estimates unless self-selection is controlled for.

2.2. Errors-in-variables

Besides omission of variables, a further source of endogeneity is errors-in-variables, which refers to problems that arise when variables are imperfectly measured and their true values remain unobserved (Wooldridge, 2002). Measurement errors result from the use of inadequate measurement instruments to capture concepts of interest, or non-comprehensiveness of the data collection method (Kennedy, 2008). Typical examples include scale items being improperly adapted to the research context, wrong aggregation of constructs, failures in survey translation, or non-reliable construct measures. In addition, missing data can be considered as a form of measurement error (Kennedy, 2008). Errors-in-variables constitute an issue when the variables on which data can be collected differ from the variables that influence decisions of relevant actors (Wooldridge, 2002). Measurement error in the dependent variable can cause biases if it is systematically related to one or more of the independent variables under investigation; however, it will play a subordinate role if it is uncorrelated with the independent variables and it is usually of minor relevance as it is captured by the error term of the model. Measurement error in independent variables is considered as important and the properties of the OLS estimation depend on particular assumptions about the measurement error (Wooldridge, 2002). The first assumption is that the measurement error and the observed independent variable are uncorrelated, and that the error term of the model is uncorrelated with the actual (unobserved) and the observed independent variable. In this case, estimation yields consistent coefficients. The second assumption, which is referred to as the 'classical errors-in-variables (CEV) assumption', is that the measurement error is uncorrelated with the actual (unobserved) independent variable, and that the error term of the model is uncorrelated with the actual and the observed independent variable. In this case, the observed independent variable and the measurement error are correlated and the estimation yields inconsistent coefficient estimates: the coefficient estimate will be biased towards zero ('attenuation bias') and the size of this bias depends on the variance of the actual independent variable relative to the variance in the measurement error.

2.3. Simultaneous causality

Endogeneity bias may also be caused by simultaneous causality, which occurs when one (or more) independent variable is jointly determined with the dependent variable, i.e. when independent variables and dependent variables simultaneously cause each other and causal effects run reciprocally (Wooldridge, 2002). Because the error term of the model contains all unobserved factors that influence the dependent variable and, in the presence of simultaneity, the dependent variable influences the independent variable, the error term is also correlated with the independent variable, thus leading to endogeneity