Management Review Quarterly (2024) 74:1809–1839
https://doi.org/10.1007/s11301-023-00348-2

Control variable use and reporting in operations management: a systematic literature review and revisit

Carla Curado[1] · Mírian Oliveira[1,2] · Dara G. Schniederjans[3] · Eduardo Kunzel Teixeira[4]

Received: 3 January 2023 / Accepted: 7 April 2023 / Published online: 30 May 2023

Abstract

The purpose of this paper is to identify (1) to what extent do researchers follow standards on the use and reporting of control variables? (2) what are the main weaknesses presented in current as they relate to control variable inclusion/exclusion? And (3) what are the recommendations based on previous standards when it comes to the selection and inclusion of control variables? We do this through the lens of Operations Management literature (inclusive of Supply Chain Management, Information Systems and Knowledge Management) to control for field specific practices and apply management theory as an initial lens to examine best practices. An extensive systematic literature review and content analysis is conducted on literature spanning 2010–2020. Control variable analyses are conducted and organized into interdisciplinary domains and DVs, providing researchers with insights in use of specific control variables from a micro-level perspective. Next, we identify strengths and weaknesses in current control variable use among a ten-year span from a macro perspective. We also provide trends across time on control variable use and inclusion.

Keywords Control variable use · Operations management · Management theory · Systematic literature review

JEL Classification M110 · C140 · C110

1 Introduction

A central pursuit of panel data and structural estimation is describing the relationships among variables. Pivotal for this are the identification and isolation of variables that explain or predict the dependent variable (DV) while simultaneously controlling for variables that extraneously impact the relationship. In other business areas, various studies have discussed and enacted models in the determination and

Extended author information available on the last page of the article use of control variables (CVs) in empirical research (Atinc et al. 2012; Nielsen and Raswant 2018; Spector and Brannick 2011), yet little work has been done in model development toward understanding how and when to utilize CVs. This lack of literature is prevalent in field-specific domains despite numerous insights into arbitrary inclusion of CVs in research (Miller et al. 2021). In this study, we focus our efforts on the field of operations management (OM)[1] inclusive of related literature about supply chain management (SCM) as well as information systems and knowledge management (IS/KM) characterized in terms of proactivity via hands-on techniques and best practices (Curry 2019). We use the OM context as a case for theory elaboration with current theory in management via Bernerth and Aguinis (2016) model of CV use.

We found several weaknesses related to OM's use and reporting of CVs based on existing management theory. This has led to a reduction in the explanatory power of research models (Atinc et al. 2012) over time and potentially created false results (Becker 2005). While largely unnoticed, there is some OM literature that has identified inappropriate CV inclusion/exclusion practices. With a focus on SCM, Helmuth et al. (2015) provide an assessment of empirical research detailing past and present recommendations for the future regarding effect size, statistical power, reliability, and controls. Lavenberg and Welch (1981) provide a comprehensive breakdown of the application of CVs for the purpose of increasing the efficiency of discrete event simulation. In more recent literature, Fisher et al. (2020) explicate one issue with identification of proper controls utilizing the example of measuring the impact of inventory and product variety on sales in a retail store. While each of these studies provides valuable insights into CV use particular to OM, SCM, and management science, apart from Lavenberg and Welch's (1981) study, they do not provide a model depicting decision criteria specific to CV use and reporting. Lavenberg and Welch (1981) focus was constrained to discrete event simulations.

For these reasons, we borrowed from the insights of management literature regarding the identification of standards in CV use and reporting. From current models, we utilized a critical review and content analysis of CV use and reporting in 10 years of OM literature. We then addressed the following questions:

(1) To what extent do researchers follow standards on the use and reporting of CVs?
(2) What are the main weaknesses presented in the current literature?
(3) What are the recommendations based on previous standards for researchers when it comes to selecting and including CVs in their research?

In addressing these questions, our findings provide several insights into the weaknesses of current literature as it relates to CV use and reporting. We thus developed a revised "Decision Making Tree Summarizing Sequential Steps in the Process of Selecting CVs" specific to the bottlenecks presented in our analyses. These results provide useful implications for causal identification, graph theory, and replication studies. Because the need for replication studies has been stressed in various fields

[1] OM: operations management, SCM: supply chain management, CV: control variable, DV: dependent variable, IV: independent variable, IS/KM: information systems and knowledge management.

not only within management research (Block et al. 2022) but outside the boundaries of business and management including ecology (Fraser et al. 2020), mathematics (Aguilar 2020), and psychology (Bonett 2012), new methodological and theoretical models can be leveraged to provide a basis for constructive or confounded replications.

Other contributions include an understanding of how CV use and reporting practices differ in various micro-level contexts (e.g., by interdisciplinary domain and by identification of three commonly used DVs based on research questions). In delineating the differences of CV use and reporting practices at a micro-level, we followed previous research that examined standards based not only on an area level but also on an interdisciplinary domain level and further adding a DV level. We did this to offer a more detailed and fine-grained analysis of standard practices and, more specifically, weaknesses on micro-levels for those researchers and journals interested in a particular context.

Identification of strengths and weaknesses and their trends across micro- and macro-level research domains will allow for the elucidation of current progress of field-specific research in CV use and reporting. This will provide the groundwork for identifying prescriptive and future-oriented standards that researchers, journal editors, reviewers, and scientific readership can use.

2 Concept and background

2.1 Control variable use

CVs can be used to argue for cause-and-effect relationships through experimental or quasi-experimental designs (Atinc et al. 2012; Bernerth et al. 2018). In experimental design, researchers prevent the introduction of extraneous influences through the constraint of experimental conditions, which may independently influence outcomes. Experimental design often utilizes laboratory settings for this purpose.

When researchers decide to examine hypotheses utilizing regression, the researcher will typically account for confounding effects with the use of CVs, which represent components of the main or interaction effects and have the potential to relate to the DV. CVs are included in the analysis to address omitted variable bias (Hill et al. 2020). In hierarchical regression, they are typically entered before other IVs to determine their power exclusive of the IVs or to understand how much variance can be explained by controls versus IVs (Tabachnik and Fidell 2000). With statistical control, the control of extraneous influences occurs after the fact and involves identifying sources of influence through study design. Control is imposed by mathematically partialling out the variance associated with CVs in specifying the relationships with other variables (Carlson and Wu 2012).

Previous scholars have identified the remedies we refer to as “standards” in this paper which have been identified by previous scholars as viable solutions to common pitfalls in CV use and reporting (e.g., Bernerth and Aguinis 2016).

2.2 Control variable use and reporting pitfalls and solutions

Carlson and Wu (2012) highlight various pitfalls attributed to CV use and reporting. These are based on three purposes for which CVs are included in a model. The first is the "purification principle" (Spector and Brannick 2011), which suggests that a researcher can remove any distortion associated with extraneous variables and thus purify the results by exposing the "true relationships in a model" (Atinc et al. 2012; Bernerth et al. 2018). The purification principle is so widespread that researchers generally do not question its validity. The primary challenge is finding an appropriate CV.

Candidate CVs will sometimes capture only part of the variance, and partialling removes a portion of the variance. Similarly, the IV and CV may share combinations of meaningful and contaminant variance, which can be especially problematic in multiple regression analysis. Several issues have resulted in what Carlson and Wu (2012) refer to as the consistent following of academic researchers in the "universal CV playbook." These practices have led to two overarching problems: confounded outcomes and partial alignment with the requirements of intended purposes. These issues and solutions are outlined below.

2.2.1 Selecting control variables based on their association with the dependent variable

In the decision to include CVs in one's research model, researchers often consult prior literature utilizing similar DVs and IVs. However, this over-simplification has led to various issues related to the relevance of the research question. If a CV is used in a different contextual setting from previous studies, its relevance, appropriateness, and validity should be questioned (Nielsen and Raswant 2018). Inappropriate inclusion can also confound attempts to purify a relationship (Carlson and Wu 2012).

The practice of including CVs just based on previous work solidifying the relationship with the DV is not appropriate with multiple regression analysis, as it will control the IV for the CV. However, it may be appropriate if the covariance of the CV and DV reflect a contaminant in the DV. Researchers must also consider that the intended control effect should be clearly specified because mathematically partialling out the shared variance between CVs and IVs can cause regression coefficients to have ambiguous interpretations.

Another problem associated with utilizing CVs based on their associations with IVs and DVs is collider variables where an IV and a DV can have a causal influence on a CV but the IV and DV are not necessarily associated. The collider will then block the association between the variables that influence it and thus undermine attempts to test causal theory (Greenland and Pearl 2011). CVs can also be misconstrued as not capturing confounding effects but rather capturing contingencies, mediators, or moderators that ought to be adopted for the purpose of isolating the underlying causal mechanisms of the causal effect.

Steinmetz and Block (2022) help to identify the various challenges in developing a causal identification strategy through a discussion of graph theory about deciding whether an auxiliary variable should be included in a model. In

Steinmetz and Block (2022), they discuss upward or downward bias of ignoring confounders, controlling for mediator variables, a variable that is cause by the outcome, and a collider, all of which introduce bias. As a minimum solution, Nielsen and Raswant (2018) suggest consideration of the theoretical implication of specific CVs in relation to key relationships prior to data collection. This includes identifying predicted directionality of CV-DV relationships. Explaining the DV in terms of its residual ensures theoretical relevance and provides evidence of boundary conditions regarding theory and predictions. Predictions of sign for a CV-DV relationship can also help reveal the potential impact of inclusion a priori (Becker 2005).

2.2.2 Including too many versus not enough control variables

In their review of SCM research over three decades, Helmuth et al. (2015) highlight a Pearson correlation coefficient of 0.25 ($p < 0.01$) indicating a strong, growing prevalence of CVs over time. While the growing prevalence of CVs seemingly indicating methodological advances in the area, there is an important warning against not including enough CVs or including too many.

CVs substantively change the meaning of the relationship between the IVs and the DV. The inclusion can also reduce available degrees of freedom and lower the statistical power. Moreover, their inclusion may diminish the explainable variance in outcomes attributed to specific predictors (Becker 2005; Bernerth et al. 2018; Carlson and Wu 2012). Another complication is the interpretation and replication of a study's results with the use of demographic proxies in data analytic models. For example, Bernerth et al. (2018) identify a study that controls for managers' age, gender, education, and organizational tenure. This study in essence investigates an ageless, gender-neutral individual with no education or prior organizational experience based on the parsed-out variance attributed to the four abovementioned factors and thus reflects an unrealistic context.

Indiscriminate identification and utilization of controls can cause both Type I and Type II errors through the partialling out of true variance from the relationships of interest (Spector et al. 2000). Including more control also does not equate to rigor or event conservatism in hypotheses tests (Carlson and Wu 2012; Spector and Brannick 2011). To reduce these dilemmas, some authors have called for an automatic exclusion with doubt (e.g., "When in doubt, leave them out"). Becker et al. (2016) note that whenever a researcher is uncertain about utilizing a particular CV, the best advice is to leave it out. That is, researchers should exclude CVs that lack a defensible purpose. The inclusion of CVs should not only have justification but should also provide a basis steeped in theory (Bernerth and Aguinis 2016). This also avoids the problematic reduction of degrees of freedom and the power of designs (Carlson and Wu 2012). Other researchers concur and specify that theory should be the basis on which controls are added or excluded (Nielsen and Raswant 2018).

2.2.3 Entering all control variables as a block of variables and the first step of hierarchical analyses

The entering of all CVs as a block of variables has advantages and disadvantages. Specifically, placing all predictors in a multiple regression analysis model can provide a severe test of the capacity of the focal IV to increment R^2. However, this is not aligned with purification and accounting for other variables because it specifies that all CVs need to be correlated only with contaminants impacting the IVs. Additionally, if CVs representing different control objectives are included, any CV–IV effects for these variables can be confounded by the inclusion of CVs for various purposes. Including more CVs can complicate the interpretations of the regression coefficients. This evokes the problem associated with the inclusion of too many CVs.

Another commonly employed technique is entering all CVs before the IVs in the hierarchical analysis. When CVs are reflective of contaminants in IVs, this can ensure that the contaminants are removed when estimating IV–DV relationships with little to no effect of the magnitude of the βs in each model. However, this also gives the CVs first priority to account for the variance in the DV and thus may cause the variance of the IVs to be understated.

Becker et al. (2016) suggest that standard descriptive statistics and correlations for all CVs should be reported. Other research suggests reporting descriptive statistics, correlations between CVs and all other variables, and VIFs for all CVs and IVs. This provides a better understanding of psychometric properties and enhances replication potential. Researchers can also leverage descriptive statistics and correlations to gauge the amount of residual variance in a predictor after common variance is associated with the CV (Bernerth et al. 2018). In terms of measuring and analyzing CVs, previous research has suggested that errors in measuring CVs can lead to relationships becoming positively or negatively biased (Edwards 2008). Becker et al. (2016) suggest providing reliability information such as tax-equivalent reliability P_T (e.g., Cronbach's alpha) when CVs are perception based. Researchers should also analyze and report study results with and without CVs, which help to illustrate the actual impact of the CVs. In hierarchical regression where CVs are added first, this can inadvertently place theoretically meaningful variables at an inappropriate disadvantage, especially when CVs are not theoretically meaningful. Moreover, the relationship between CVs, IVs, and DVs should be provided because the shared variance between these variables can increase the magnitude of a regression coefficient (Bernerth et al. 2018). Unlike surrogate or proxy variables, all CVs should be conceptually meaningful (Becker et al. 2016; Bernerth et al. 2018; Spector and Brannick 2011).

Delineating theoretical versus artifact CVs in hierarchical analysis can also be useful because theoretically meaningful CVs and IVs can be given the first opportunity to account for variance in the outcomes since they offer explanations. Artifact CVs can be entered last to give reasonable account for those IVs and CVs that have theory-driven explanations (Carlson and Wu 2012).

3 Research method

We seek to add to management theory on CV use and inclusion through a context-specific case. Bernerth and Aguinis (2016) provide the foundational elements of CV use that we applied to the field of OM to (1) deliver a baseline model and (2) add to the baseline model through a case representation and systematic literature review. The literature review allows a systematic identification and integration of studies to elaborate on or enact models that set the groundwork for future research (Block and Brändle 2022). As such, a literature review was appropriate for our initial data collection given our research questions. Next, we chose OM as the field for analyses. Continued growth in the number of empirical papers, specifically those that draw upon secondary data, has been observed in this field (Terwiesch et al. 2019). This growth is pervaded by work bridging concepts from psychology, management, organizational behavior, IS/KM, economics, etc. Given the often interdisciplinary context and the diverse questions that OM has addressed in previous research, we sought to view CV standards from three levels: OM, SCM, IS/KM closely related to theories and methods from marketing, management, and operations; Hochrein et al. 2015) and DVs. We identified the similarities in CV use among each level (OM, SCM, IS/KM) and provide a breakdown by most common CVs. Additionally, we examined the larger picture by analyzing trends of CV use and inclusion standards over time by identifying frequencies by CV over time. Given the large amount of data, we sought to organize findings on a macro- and micro-level which provides a more fine-grained depiction of weaknesses on each level. The findings specify the weaknesses for the purpose of researcher-preferred interest. The refined results provide more specific contributions for researchers who wish to contribute to a certain interdisciplinary domain or research area versus the model, which provides researchers with a generalized tool to leverage in contributing to a broader context.

Given these aims, we conducted content analysis on a collection of empirical OM research. Content analysis is a class of varied methods bridging both quantitative and qualitative analyses (Seuring and Gold 2012). To incorporate the varied methods of content analysis, Mayring (2008) defines content analysis as any methodological measurement applicable to text for social science purposes that is systematic, rule governed, and theory driven. Our data derives from previous OM literature in the use of CVs while our research questions require statistical analyses to determine weaknesses and standards regarding CV use and reporting; hence, content analyses were appropriate given the context of our study and data.

There are several types and classifications of content analysis. Krippendorf (2004) refers to three techniques as discussed by Janis (1965), namely (1) pragmatic, which consists of procedures classifying signs based on the probable causes or effects; (2) semantic, which classifies signs according to their meaning; and (3) sign-vehicle analysis, which classifies content according to psychophysical properties. More recent literature Hsieh and Shannon (2005) has refined the definition of approaches through inductive versus deductive reasoning. Specifically, they describe conventional approaches whose aim is to describe a phenomenon when existing theory or research is limited versus directed content analysis where existing theory or prior research exists and would benefit from further description.

Our analyses are based on previous work in management literature discussing the various standards of CV use and reporting. As such, we took a directed semantic approach to our analyses utilizing existing literature (e.g., Bernerth and Aguinis 2016; Carlson and Wu 2012, etc.); hence, we used directed semantic content analysis.

3.1 Research context and data collection

We conducted content analysis on articles published by reputable journals in OM from January 2010 to January 2020. We used this period for two reasons. First, a decade provides a large amount of data to understand how different CVs have been utilized to measure varying DVs. The researchers also found that similar DVs and CVs were utilized in many studies during this period, which equates to theoretical saturation. Second, the period from January 2010 to January 2020 provided the researchers with the most recent research.

We deployed a field-specific (OM) context for a few reasons. Research context is critical because it informs hypothesis development, measurement choice, data analysis, interpretation, and reportage of the findings (Johns 2006). Even trivial contextual stimuli are said to provide substantial impacts regarding study-to-study variation and thus should be defined accordingly (Johns 2006). Stremersch et al. (2023) have outlined several benefits to context specificity. First, context specificity provides more internal validity via greater accuracy and precision within versus across contexts. Second, we sought to create innovative ideas through an analysis specific to a context with general appeal. Like management, OM provides a context-specific field that leverages management theory and provides a context with interdisciplinary connections to other growing areas, namely SCM and IS/KM. This allowed us to identify patterns in a multidisciplinary context while at the same time breaking down analyses into several different sub-contexts.

Generalized analyses are also provided for various fields in the management domain. Our results depict both context-specific CVs (e.g., project complexity, which is typically focused on project management) and generalized CVs (firm size, age, industry), which are utilized in various management domains including organizational behavior, entrepreneurship, and marketing. Each field has selected theories, research questions, and common dependent variables that consistently appear within their domains (i.e., "willingness to pay" and "theory of planned behavior" in marketing; "technology adoption" and "technology acceptance model" in IS). Nevertheless, commonalities among different areas persist because many fields (including SCM) borrow from management theory since the field is in its infancy and currently developing theory specific to the area. The results of this study provide a unique and specific context for examining management theory in CV use while at the same time providing results (with similar CVs) that can be generalizable to other areas in the management domain.

Table 1 Variables and descriptions in data sets

Yield	Inclusion/Exclusion criteria
174,000	Search string development and refinement: "control variable" AND ("operations management" OR "production")
39,900	Time period: January 2010-January 2020
1031	ABDC inclusion; removal of non-reputable sources and articles not relevant to research

To ensure the journals' reputation, we used the Australian Business Dean's Council (ABDC) journal list. We decided to utilize this list for a few reasons. First, we sought to ensure that journal selection included both US and international journals. Second, the ABDC provides extensive reviews conducted by expert panels relating to various areas including those that have similar interdisciplinary DVs.

The next step was developing the search string. This occurred through an iterative process. First, we utilized an initial string, namely ("control variable" OR "controls") AND ("operations management" OR "production"). We initially developed this search string and gave it to one subject librarian (see Durach et al. 2017) and a Ph.D. student at one author's institution for further refinement. After a couple of iterations, a final search string was developed, namely "control variable" AND ("operations management" OR "production").

Using the work of Moher et al. (2009), we defined the number of records initially identified and screened by select eligibility guidelines. These are defined and included in Table 1. We used a Google Scholar search with a comparison to Academic Search Complete (EBSCOhost). Academic Search Complete yielded several more articles because of an inclusion of biology, chemistry, food science, etc. fields. The final search after identification of OM focus and database cleansing yielded approximately 174,000 articles. After focusing on the specified period, we identified 39,900 articles. After ABDC inclusion, we accessed a total of 22 journals including those from reputable publishers such as Elsevier, Wiley-Blackwell, Springer, the American Psychological Association, the Emerald Group, Taylor & Francis, and INFORMS. After removing articles that were not peer reviewed (e.g., introductions to special issues, author biographies,[2] etc.) we also excluded certain studies that did not fit within the confines of the study (e.g., could not be accurately applied to CV

[2] Some articles were published that referenced a biography of a particular author and/or an author who had recently passed away.

standards as set forth in previous literature; Bernerth and Aguinis 2016). We identified a total of 1031 articles from OM journals.[3,4]

Next, utilizing the cleansed data set, we sought to specify interdisciplinary domains. Doing so would allow for a fine-grained analysis of standards across varying interdisciplinary domains and thus allow us to identify both differences and weaknesses across these domains. Hence, we conducted a term frequency-inverse document frequency (TF-IDF). This technique is used when seeking to determine the relative frequency of words compared to the inverse proportion of a word in the corpus (Ramos 2003). This is used to determine how relevant a word is in a particular document and thus allowed us to identify domains by keyword.

We identified a total of 117 keywords as having significant weight (e.g., $>0.1\%$, occurring at least six times in 10 years). The keywords and their frequencies are depicted in Fig. 1 below.

We then conducted a Q-sort with a professor and two graduate students to derive a consensus-based assessment. Q-method provides a systematic study of a person's viewpoint through a topic called the Q-set. The method works best with limited numbers of informed respondents who are asked to order statements (or keywords) based on their individual perspective, preference, or judgement through a quasi-normality distribution (Exel and Graaf 2005). All three coders were provided with the list of keywords and were then asked to group the keywords into clusters based on their academic expertise. The researchers identified several clusters with agreement. All three coders agreed upon four clusters, namely (1) OM, (2) SCM, (3) research methods, and (4) theory.[5]

Two of the three coders agreed on separate codings including (1) IS/KM, (2) sustainability, (3) quality, (4) risk, and (5) project management. Any other coding was put aside. We attribute these differences in codings to the coders' experience. Specifically, the professor and the doctoral student who had done more coding also had more experience in research as well as field experience in OM and SCM. After two iterations, an inter-rater agreement of 80% was obtained within the four codings, and an inter-rater agreement of 85.25% was reached among the eight codings.

Kappa was calculated for each of the coder pairs. Kappa was computed as

$$\kappa = \frac{P(a) - P(e)}{1 - P(e)}$$

[3] To analyze the standards as set forth by control variable literature, reporting methods needed to include statistical data analysis. Most articles referenced archival data analyses and survey data analyses. As a result, our results are heavily biased toward these methods and should not be utilized to assess laboratory observational experimental design. However, we did not exclude survey-based experimental designs. Nevertheless, this design did not represent a large enough sample in our data to be analyzed separately.

[4] Control variable standard practices by method are provided in Table 8 in the Appendix.

[5] Research methods keywords included specific method keywords that can be utilized broadly. These keywords included "survey design," "case study," "action research," etc. Theory categorization was similar with specific theories referenced in the keywords utilized. Examples included "resource-based view," "systems theory," "agency theory," etc.

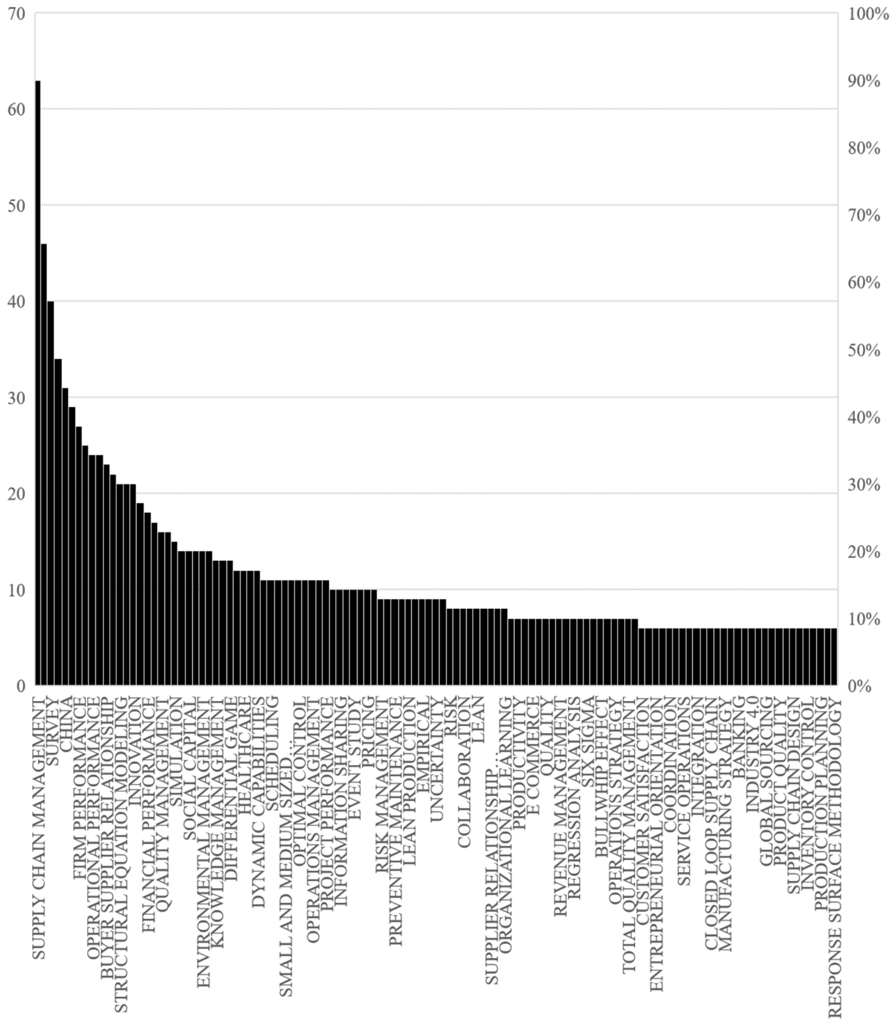

Fig. 1 Frequencies of keywords

where P(a) is the observed percentage of agreement and P(e) is the probability of expected agreement based on chance (Hallgren 2012). We then used the arithmetic mean of the pairwise estimates to provide an index of agreement (Light 1971).

Given that we sought to identify inter-disciplinary domains, we focused our analyses on both OM-specific articles (e.g., OM categorization: 28 keywords) and two of the most common interdisciplinary contexts (SCM: 17 keywords and IS/KM: 5 keywords). Both categorizations were considered distinct[6] and had significant enough

[6] While not within the confines of the article to distinguish between operations management and supply chain management, given the various similarities, interdependencies, and more established corporate reputation for OM (e.g., chief operations officer overseeing supply chain management), we denoted a

Table 2 Keywords utilized in analysis and number of articles per keyword

Operations keywords	Supply chain keywords	IS/KM keywords
Behavioral Operations (14)	Bullwhip Effect (7)	E Commerce (7)
Inventory (9)	Buyer Supplier Relationship (23)	Industry 4.0 (6)
Inventory Control (6)	Closed Loop Supply Chain (6)	Information Technology (6)
Inventory Management (13)	Global Sourcing (6)	Knowledge Management (13)
Lean (8)	Green Supply Chain (11)	Manufacturing Systems (8)
Lean Manufacturing (10)	Logistics (7)	
Lean Production (9)	Offshoring (10)	
Manufacturing (12)	Outsourcing (19)	
Manufacturing Industries (6)	Supplier Integration (8)	
Mass Customization (6)	Supplier Involvement (7)	
New Product Development (22)	Supplier Relationship Management (8)	
Operational Performance (24)	Supply Chain (34)	
Operations Management (11)	Supply Chain Design (6)	
Operations Strategy (7)	Supply Chain Integration (16)	
Preventive Maintenance (9)	Supply Chain Management (63)	
Process Innovation (7)	Supply Chain Performance (6)	
Product Quality (6)	Supply Chain Risk Management (7)	
Production Planning (6)		
Project Management (25)		
Project Performance (11)		
Project Success (7)		
Quality (7)		
Quality Management (16)		
Scheduling (11)		
Service Operations (6)		
Six Sigma (7)		
Sustainable Operations (6)		
Total Quality Management (7)		

data to distinguish differences and weaknesses among standards.[7] Table 2 depicts the codings of the keywords.

The next step involved individually accessing each article to identify CVs and DVs. The excluded studies were either mathematical modeling or papers involving

Footnote 6 (continued)

delineation particularly from the standpoint of supply chain management referencing external activities (e.g., supplier relationship management, procurement/purchasing, transportation) versus operations management referencing internal activities (e.g. production, manufacturing, internal operational control, and monitor).

[7] Based on sample size calculation, the difference between samples testing for equivalence (with $\alpha = 5\%$, group number = 11), the minimum n per group = 11. Assuming a non-normal distribution, n per group = 13 (Lehmann 2012).

Table 3 Variables and descriptions in data sets

Variable	Description
Journal	Journal title the academic article was published in
Title	Title of the academic article
Keyword	Keywords provided by the author(s)
Control	Control variables utilized
Dependent Variable	Dependent variable study assessed

Table 4 Control variables and justifications in data sets

Criteria	Code
Provides a citation for inclusion	0: no citation; 1: citation
Previously used as a control	0: did not mention previous research as a reason; 1: did mention previous research as a reason
Anticipated relationships	0: did not cite a potential relationship with a variable as a reason; 1: did mention a potential relationship with a variable as a reason
Previously found relationships	0: did not cite a previously found relationship as a reason; 1: did cite a previously found relationship as a reason
Incremental/Discriminant validity	0: did not mention incremental/discriminant validity as a reason; 1: did mention incremental/discriminant validity as a reason
Eliminate alternative explanations	0: did not mention the elimination of alternative explanation; 1: mentions the elimination of alternative explanation
Theoretical explanation	0: did not attempt to explain how and why the control variable relates to other variables; 1: did attempt to explain how and why the control variable relates to other variables
Two-step process	0: did not describe a two-step process for control variable inclusion; 1: did describe a two-step process for control variable inclusion
Total number of justifications	Total number of justifications (1…n) for each control variable

Justification (0: no justification; 1: some justification); Multiple: article provides two or more justifications per variable (1 = yes; 0 = no)

the main purpose of developing a revised or new methodology. While some methodology papers employed an empirical test, the main research questions were not focused on an empirical research question. Tables 3 and 4 (adopted from Bernerth and Aguinis 2016) present the data collected for each article. OM, SCM, and IS/KM were each contained in a separate data set. Following this, frequencies were identified.

3.2 Control variable justification categories

Frequency procedures were performed on each of the levels (i.e., OM, SCM, IS/KM), and then the data from each domain were consolidated. The tables are too large (over 3000 pages each) to provide in the manuscript but are available upon request.

The consolidation table contains 827 controls and 251 DVs. The 10 most frequently used DVs are as follows: financial performance (144 counts, 6.70%), operational performance (140 counts, 6.52%), firm performance[8] (118 counts, 5.49%), project performance (116 counts, 5.40%), project success (38 counts, 1.77%), supplier engagement in pro-environmental practices (36 counts, 1.68%), quality performance (34 counts, 1.58%), innovation (28 counts, 1.30%), stock return (28 counts, 1.30%), trust (27 counts, 1.26%), and buyer continuity intention (26 counts, 1.21%[9]). All three of the top DVs were utilized in all three interdisciplinary domains. Hence, further analyses were focused on the varied interdisciplinary domains as well as these DVs and their associated controls.

3.3 Use of control variables

Table 5 represents the results regarding the justifications utilized for including CVs in the OM, SCM, and IS/KM domains. There are distinctions amongst each domain. in CV use practices. For the most part, domains are on relatively equal footing in providing justifications, describing current literature, providing incremental and discriminant validity tests and theoretical support, and conducting appropriate analyses.

The IS/KM domain exceeds both the OM and SCM domains when it comes to justifying the inclusion of each CV (56%; OM: 51%; SCM: 50%), presenting the background literature of the CV in terms of anticipated relationships (22%; OM: 18%; SCM: 20%), presenting discriminant and incremental validity (18%; OM: 8%; SCM: 11%), discussing previously found relationships (78%, OM: 25%; SCM: 21%), providing theoretical support (13%; OM: 8%; SCM: 11%), and including the CV in the data analysis (94%; OM: 82%; SCM: 61%). The SCM domain exceeded other domains in providing a citation for inclusion (50%; OM: 45%; IS/KM: 43%). Finally, the OM domain exceeded the other domains in utilizing at least a two-step process for CV inclusion/exclusion (57%; SCM: 16%; IS/KM: 94%).

In general, the IS/KM domain seems to exceed the other two domains according to a variety of standards. Notably, the weaker standards include providing a theoretical basis for the inclusion/exclusion of CVs, adding incremental and discriminant validity tests, and eliminating CVs when they add no value. Moreover, most studies do not appear to present the anticipated relationships of CVs. This suggests a trend like that of management research. OM researchers are utilizing previous research to

[8] Firm performance refers to dependent variables not offering specific identification of performance (e.g., "organizational performance," "performance") but rather as an aggregate construct theoretically specified as a composite of dimensions, and at times, those dimensions are not specified. While it is not within the confines of this article to debate the use of "firm performance" as a variable, one should note the prevalence of the performance variable in OM literature as well as current perspectives regarding its use as an aggregate measure (e.g., Miller et al. 2012).

[9] True to the OM literature, most control variables were firm level (e.g., firm age, financial performance, market performance, etc.), and few were demographic oriented (e.g., top management gender, age, etc.). This is understandable given that in most cases, the unit of analysis stays within a manufacturing organization. However, there was a lack of insight into interorganizational relationships with performance-dependent variables. With the growing importance of SCM, the inclusion of such control variables in OM and IS/KM domains may be necessary for assuming theoretical justification.

Table 5 Research domains and control variable standard practices

Research domain	Justification (%)	Citation (%)	Previous (%)	Anticipated (%)	Found (%)	Incremental (%)	Eliminate (%)	Theoretical (%)	Analysis (%)	Process (%)	Multiple (%)
OM	51	45	25	18	17	8	1	9	82	57	9
SCM	50	50	21	20	27	11	6	7	61	16	7
IS/KM	56	43	78	22	26	18	11	13	94	55	4

*When analyzing the differences amongst the use of differing practices between fields (i.e., Operations Management, Supply Chain Management, Information Systems/Knowledge Management), we find no significant differences when utilizing a Kruskal Wallis Test. We do, however, find significant differences between the varied practices (i.e., justification, citation, theory, etc.) (Means: Justification: 0.5233; Citation: 0.46; Previous: 0.4133; Anticipated: 0.2; Foundational: 0.2333; Incremental: 0.1233; Eliminate: 0.06; Theory: 0.0967; Analyze: 0.79000; Process: 0.4267; Multiple: 0.0667; p-value < 0.0001)

validate the inclusion of a CV as opposed to theoretically justifying its use and considering the potential relationships the CV has with either the DVs or IVs. Moreover, while optimistic that 50% or fewer studies provide justification and citations for CV use, these are also areas to improve.

Finally, we report the results of the justifications given for specific CVs. Utilizing our analyses above, we depict the three most common DVs across all three domains and their associated CVs. This provides an overview of the most common DVs and how previous research has defined CV inclusion. We also include the top 10 CVs by frequency for each of the domains, allowing researchers to find the weaknesses of CV use in their domain of expertise. We utilized standards by year to provide a figure depicting CV standards and trends. These trends encompass all CVs to provide a larger picture of standards over time.

We conducted a Friedman test to determine the significant differences among DVs and their associated CVs. There appear to be significant differences among the varied CV best practices among the top three DVs (*p*-value: 0.0176).[10] Additionally, we found support for differences of CV use best practices among the top CVs utilized (*p*-value: 0.0306). Specifically, based on the data presented in Tables 6 and 7, a majority of studies are including justifications and citations for including specific CVs. Yet, there is still much work to do in these domains to enhance CV standards, particularly in providing incremental/discriminant validity checks, eliminating the variable when non-significant, providing theoretical justification, and conducting appropriate regression analysis procedures. For example, firm size is the most utilized CV; however, only 21% of the studies identified through our work found a significant relationship with the DVs or IVs in this research model. Furthermore, only 3% of the studies removed the variable after determining non-significant findings. While CVs like project complexity and, to an extent, R&D intensity inflated the relationships among the DVs and/or IVs, in general, significant relationships equated to approximately 34% of the studies identified. However, the elimination of insignificant CVs averaged approximately 2%. The inclusion of such CVs reduces degrees of freedom, lowers statistical power, and may diminish the explainable variance of the outcomes attributed to predictors (Bernerth et al. 2018). Based on this analysis, this appears to be a major issue. To a lesser extent, theoretical justification is also problematic as seen in both Tables 6 and 7. While studies examining financial performance were slightly better at providing theoretical justifications for including specific CVs, firm performance and innovation (which are increasingly being utilized as DVs) continue to lag. These tables also reference the differences between different studies. For example, in Table 6, for the most part, proper regression analyses were utilized for studies in financial performance but lagged in firm performance and innovation. These studies may have skewed the results in Table 6, where we see optimistic results in proper regression analyses.

Examination of these factors from both micro- and macro-perspectives portrays the importance of standards in CV inclusion because research models with specific DVs may have more problems than others. With innovation being not only a top

[10] Only control variables that were common among dependent variables were leveraged for statistical comparison tests.

Table 6 Top dependent variables in Operations Management, Supply Chain Management, and Information Systems/Knowledge Management, control variables and standard practices

DV	CV	Justification (%)	Citation (%)	Previous (%)	Anticipated (%)	Found (%)	Incremental (%)	Eliminate (%)	Theoretical (%)	Analysis (%)	Process (%)	Multiple (%)
Financial performance	Firm size	37	33	30	19	19	0	0	15	70	37	7
	Age	45	27	9	0	0	0	0	9	55	64	9
	Industry	69	69	56	38	50	0	6	25	75	56	0
	R&D	71	86	86	57	71	0	14	43	100	43	0
	Capital intensity	75	100	100	50	75	0	0	50	100	25	0
	Country	13	13	13	13	13	0	0	0	25	75	50
	Competitive intensity	100	100	25	25	25	0	0	25	75	50	0
	Sales/Revenue	57	57	29	14	29	0	14	14	86	43	14
	Debt	60	60	60	60	60	0	0	40	100	0	0
	Project size	100	100	100	50	100	0	50	50	100	50	0
	Inter-organizational communication	100	100	100	50	100	0	50	50	100	50	0
Firm performance	Firm size	58	55	16	32	16	6	6	6	84	52	6
	Industry	36	36	27	45	18	27	9	0	91	45	0
	Nature of business	67	67	0	0	33	0	0	0	67	33	0
	Number of customers	100	100	0	0	0	0	0	0	0	0	0
	Number of suppliers	100	100	0	0	0	0	0	0	0	0	0
	Product range	100	100	0	0	0	0	0	0	0	0	0

Table 6 (continued)

DV	CV	Justification (%)	Citation (%)	Previous (%)	Anticipated (%)	Found (%)	Incremental (%)	Eliminate (%)	Theoretical (%)	Analysis (%)	Process (%)	Multiple (%)
Innovation	Firm size	70	70	10	20	20	0	0	0	60	40	0
	Industry	50	50	0	0	13	0	0	0	75	0	0
	Age	43	43	14	0	14	0	0	0	43	43	14
	Revenue	100	100	0	0	50	0	0	0	50	0	0
	Mergers and Acquisitions	100	100	0	0	50	0	0	0	50	0	0
	Patent activity	100	100	0	0	50	0	0	0	50	0	0
	R&D	100	100	0	0	50	0	0	0	50	0	0
	Professional experience	0	0	50	0	0	0	0	0	50	50	50

Table 7 Top control variables and standard practices

CV	Justification (%)	Citation (%)	Previous (%)	Anticipated (%)	Found (%)	Incremental (%)	Eliminate (%)	Theoretical (%)	Analysis (%)	Process (%)	Multiple (%)
Firm size	58	51	29	29	21	7	3	11	82	48	8
Industry	56	56	41	28	31	11	4	11	81	49	6
Location	43	40	40	12	20	2	0	11	60	52	20
Age	51	34	32	23	17	8	2	13	75	55	9
R&D	77	63	54	46	46	3	9	11	94	43	6
Sales	59	38	16	16	28	6	6	16	78	38	22
Capital intensity	71	71	79	57	36	7	0	21	93%	14	0
Competitive intensity	93	86	43	29	43	0	0	21	79	64	0
Project complexity	22	89	78	11	78	0	0	11	100	78	0
Technology (novelty)	50	33	0	0	17	17	0	0	83	33	33

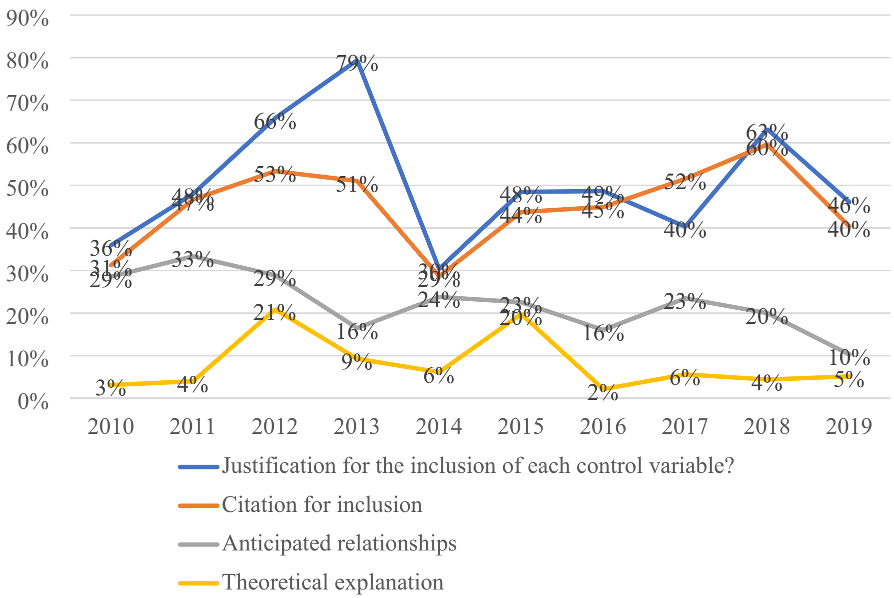

Fig. 2 Frequencies of best practices over time

DV but also an increasingly utilized DV in recent literature, it will be important for researchers to understand the weaknesses associated with inappropriate CV inclusion and use and reflect on new practices to enhance rigor in the research model.

Figure 2 presents the percentage of studies providing (1) justification for including each CV, (2) a citation for inclusion, (3) anticipated relationships, and (4) a theoretical explanation. We used percentages to depict the use of these CV practice standards throughout the entire collection of CVs and thus include data from the broader context of CV use versus micro-level analyses based on the top CVs or specific DV or interdisciplinary domains. Computations were on a CV level. It appears that between 2010 and 2013, there was growth in providing justification for including each CV as well as citations for inclusion. There was also slight growth in providing a theoretical explanation. In 2014, there was a rapid decline and subsequent growth, particularly in providing a citation for CV inclusion. There has been a decline in researchers presenting anticipated relationships for the CVs.

Overall, our analyses reflect varying degrees of weaknesses. It is important for researchers to understand these weaknesses on both a micro- and macro-level to account for potential contributions in methodological rigor. Trends over time also indicate projected directions, indicating where attention is needed by researchers.

4 Discussion

Our results provide evidence on the CV identification and utilization standard practices, strengths, and weaknesses through the utilization of management theory in CV use and application to OM research. Borrowing from the insights of previous studies in CV use and reporting (e.g., Bernerth and Aguinis 2016; Atinc et al. 2012; Carlson and Wu 2012, etc.), this study sought to answer several research questions. First, the results of the analysis addressed the extent to which researchers follow standards regarding the use and reporting of CVs. In this study, we analyzed this question through micro- and macro-perspectives specifying (1) common CVs utilized for the highest-frequency DVs; (2) the frequencies of standards by the three most frequently used DVs among the OM, SCM, and IS/KM domains; and (3) the 10 most frequently used CVs presented in our database across all domains.

We gleaned a few insights regarding the findings from this study. First, there appear to be problems with certain standards utilized. Specifically, we found substantially low percentages in finding significant relationships with CVs and subsequent removal of those CVs from the research model. As previously mentioned, this practice reduces degrees of freedom, lowers statistical power, and may diminish the explainable variance of the outcomes attributed to predictors (Bernerth et al. 2018).

While justification and citation scores are slightly higher practices in relation to other best practices, theoretical explanations are minimal. These findings together provide an interesting potential indication of isomorphism. Isomorphism is a constraining process forcing one unit to resemble other units that are subject to similar environmental conditions (DiMaggio and Powell 1983). This can be on an institutional level with a community of organizations or individuals that participate in sensemaking processes and value systems (Scott and Meyer 1994). Atinc et al. (2012) provide foundations for isomorphism in CV use through frequency identification, noting various controls that have been utilized repetitively over time. Our results also provide similar support for isomorphism specifying several commonalities in CVs utilized on macro- and micro-levels. In addition, we also note as potential evidence the commonalities of standards over time through a larger context of CV use. Specifically, from Fig. 2, we identify higher levels of justification and citation scores (e.g., the use of pre-existing literature for inclusion decisions) versus the lack of theoretical explanations from a micro-level CV-level data analysis. These findings coupled together are slightly concerning. Researchers utilize controls based on previous literature without understanding the reasoning behind utilizing the control to begin with. That is, we have minimal justification for including a CV except that the previous literature may have utilized the CV in the past. Without theoretical guidance, the use of CVs will likely change or even reverse conclusions, insinuating that flawed inputs lead to flawed outputs (Li 2021). Li (2021) provides further analysis using directed acyclic graphs suggesting that while using CVs can rule out alternative explanations, adding CVs without theoretical consideration can equally introduce overcontrol and endogenous selection biases.

As in other areas that rely on field research methods, our results depict researchers who tend to focus on establishing relationships with larger numbers of variables versus understanding the underlying processes involved (Spector and Brannick 2011). Like management scholars (Atinc et al. 2012), OM scholars appear to be subject to coercion, mimicry, and normative standards. Possible explanations may include an increasing need to strive to publish, and as such, scholars will look to earlier articles to guide decisions. Additionally, scholars can be coerced to leverage the same standards and often utilize the same variables as other scholars who came before them.

CV inclusion may also become normative in the review process based on expectations derived over time. This is commonplace and not necessarily problematic. The only time this becomes problematic is when scholars control for variables without a theoretical understanding of why they should be controlled for. While this analysis does not prove isomorphism is occurring, the numbers allude to a lack of insight on the theoretical justification of CVs. While other factors may contribute (e.g., limited word counts for journals), the lack of theoretical elaboration is prevalent in these findings, and thus, to avoid problems (e.g., reduction of available degrees of freedom, lower statistical power, interpretation, Type I and II errors), CV theoretical justification ought to be included, preferably within the manuscript or at the very least in an appendix available upon request.

4.1 Revised model for control variable identification and use

Based on these results, our research provides the following suggestions for scholars in the inclusion of CVs: as specified by Bernerth and Aguinis (2016), researchers should address the question of why they should use statistical controls. The results of this study specify that currently, OM researchers may be constrained by isomorphism, which means that previous research is utilized as the main mechanism to assess whether a CV should be utilized. This contradicts the management literature (e.g., Bernerth and Aguinis 2016), where the possible answer to this question is that the researchers believe that the CV relates to a variable included in the research. OM researchers appear to be behind other areas in business in this regard where even anticipated relationships are not well understood and there is a substantial lack of theoretical explanation.

Bernerth and Aguinis (2016) also highlight other answers to this question of inclusion, including (1) correlation between the CV and other variables, (2) presentation of an alternative explanation, (3) establishment of incremental and discriminant validity, and (4) contamination of results. While these represent possibilities for inclusion, none of them is a reason for inclusion unless the researcher shows that a theory has a possible role and justification. The decision to include controls, as in any empirical analysis, ought to be based on sound theoretical rationales about how the variables can impact the effect presented in the research model (Hansen et al. 2022). Once the theory is determined, the researcher should ask the following three questions: (1) Has this relationship been established in previous literature? (2) What

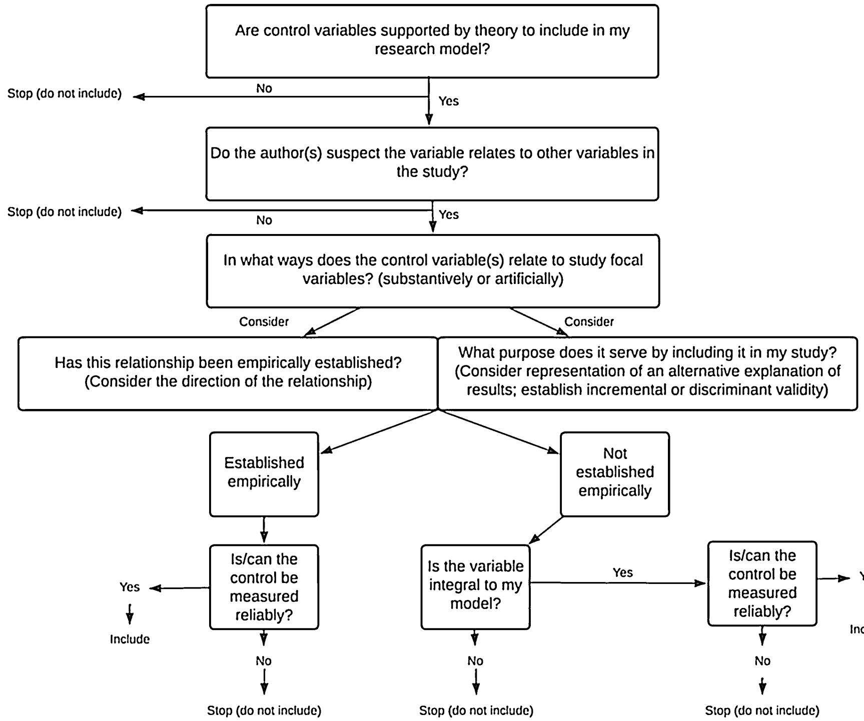

Fig. 3 Revised Bernerth and Aguinis (2016) model

purpose does this CV serve in my research model? (3) Can this CV be measured reliably? Based on our research, OM researchers place a special emphasis on Question 1 without considering the initial question with theory as well as Questions 2 and 3 (Fig. 3). In fact, Bernerth and Aguinis (2016) suggest that if this is the case, no CVs should be utilized in research models.

Researchers are encouraged to describe a process for inclusion and exclusion and feature a theoretical justification addressing the "what," "how," and "why" when assessing the relationships between CVs and IVs or DVs in the research model. To do this, researchers are encouraged to follow the "Decision Making Tree Summarizing Sequential Steps in the Process of Selecting CVs" (Bernerth and Aguinis 2016) with a slight adjustment toward a particular emphasis on theoretical rationale (see Fig. 3). In this sense, we would adjust this model as follows: instead of beginning with the question "Why do I want to use this CV?", researchers should focus on the theory utilized to develop the model and ask this question only when the theory supports an initial question. For Bernerth and Aguinis (2016), a theory is considered only after answering the following two questions: (1) "Why do I want to use this control?" and (2) "Is there any other reason why this control is being considered?"

Based on our results, the theory is suggested to be the starting point of this model because many researchers apparently include CVs that are commonly used with little to no theoretical justification. Thus, researchers are instead encouraged to start with the question, “What CVs are supported by theory to include in my research model?”.

4.1.1 Further recommendations

Our results depicting a general lack of theory use in justification is perhaps the most concerning practice because this coupled with the findings on utilization of previous research points to issues concerning isomorphism. To reduce the problems associated with lack of theoretical justification, the revised model brings this practice to the forefront of CV identification and use structure.

There are three warning signs of post-result theorizing that have been elaborated in previous literature, including contorting the theory to reflect what is defined in previous steps (e.g., cherry picking or choosing theory to fit within the parameters of preliminary steps), poorly defined constructs based on retrofitting results, construct measurement mismatch, and theory design mismatch (Shaw 2017).

Our identification of the problems presented in OM research led us to exploit the weakness in theoretical justification. In doing so, we borrowed insights from Cuervo-Cazurra et al. (2016), who provide a step-by-step breakdown for ensuring trustworthiness through theoretical development in controls. Several important insights are raised by this study that suggest ways in which authors can adjust for issues and relinquish theory-first advantages including greater theoretical contribution, developing alternative insights, and ensuring rigorous methodological results through a reduction of issues including multicollinearity and Type I or II errors (Shaw 2017).

First, Cuervo-Cazurra et al. (2016) suggest that a clear and explicit statement of the conditions under which the relationships hold is necessary provided that theoretical arguments cannot have universal applicability. Stated assumptions can thus enhance the discussion surrounding the research context. Once these are established, a second level practice involves theoretically accounting for alternative explanations, where the researcher not only explains the proposed relationship but is also subject to ruling out alternative accounts.

Two questions a researcher should address as outlined by Cuervo-Cazurra et al. (2016) include first under which conditions the arguments hold? Within this, question authors are encouraged to identify complementary factors at various levels (e.g., country, industry, individual) that a researcher assumes is the context of their insights. Additionally, authors should discuss how proposed arguments can apply to particular individuals or companies. The second question that needs to be addressed is how the relationships could be alternatively explained? Specifically, authors should identify alternative theories that may explain the proposed relationships, discuss how mechanisms proposed by alternative theories can differ from those proposed by the authors' arguments, and argue how the predictions driven by the theory are more applicable than the predictions driven by an alternative theory. In our analysis of the literature in OM, we were unfortunately not able to identify an article following Cuervo-Cazurra's (2016) suggestions above. As such, future research should promote theory development and elaboration on current and revised models prevalent in not one specific field but rather a broader base of management and its various sub-disciplines.

4.2 Future research opportunities

Like Bernerth and Aguinis (2016), the results of this paper reflect minimal progress on the use of CV best practices in research. Instead, the general insights reflect a tendency toward isomorphism and a lack of theoretical explanation. We believe that a renewed focus on theory will help address both issues simultaneously. However, this does not necessarily adjust for the statistical analysis issues such as the lack of inclusion of correlation tables or the lack of reliability as well as discriminant and incremental validity tests.

Additionally, our analyses do not currently reflect CV use by theory. It would be advantageous to assess and distinguish among theories. Future research may also incorporate a deeper analysis to show how CVs can be derived through theory and used as examples for other fields.

Future research may focus on experimental design controls. Additionally, our results are based on a depiction of the vast number of interdisciplinary domains as well as DVs. Future research may address a larger context of CV data. Finally, future research may consider utilizing a specified set of journals or a varied search string to replicate the results. This may include specifying journals from a different journal list or utilizing synonymous terms to further refine or broaden the search string. While researchers are being encouraged to adjust for all weaknesses, a renewed focus on the bottleneck issue is also being supported to adjust for the statistical power of research models.

In addition to the aforementioned possibilities, there are varied opportunities for additional research considering the model presented in this paper and its integration. We outline below possibilities for causal identification, theory development, and elaboration as well as the growing use of replication studies.

4.2.1 Causal identification and additional theoretical considerations

Auxiliary variables (represented as CVs) are fundamental because they represent a causal identification strategy allowing researchers to remove observed associations of spurious components (Steinmetz and Block 2022). Unfortunately, as demonstrated by the results of this study, the decision to include these variables may not be based on sound decision making (Steinmetz and Block 2022).

Graph theory has been offered as a potential theoretical framework to decide whether a variable ought to be included in a model (Steinmetz and Block 2022). Graph theory outlines a causal model as being represented by a graph with nodes as constructs and edges as potential causal effects. A causal model can be represented by a graph depicting both links and constraints representing exclusion restrictions. The model can provide implications for data in terms of covariances and conditional independencies (Steinmetz and Block 2022).

Steinmetz and Block (2022) leverage graph theory to help explain the challenges of developing a causal identification strategy as well as potential solutions. In using graph theory, they identify several types of auxiliary variables including confounders or surrogate confounders, instruments, selection factors (or colliders), and variables with several plausible causal roles. As specified in this research, it is important that a variable should be used as a CV or should not be controlled to avoid any possible bias of an unbiased effect. For example, mediators should not be used as controls because doing so will possibly give a biased impression of the impact of the IVs on the DV. While the model depicted in this study presents a useful starting point in the consideration of CVs, graph theory and the "decision process underlying the consideration of auxiliary variables" provided by Steinmetz and Block (2022, p. 614) can be integrated to determine CV inclusion.

4.2.2 Implications for replication studies

Replication studies serve an important purpose in empirical research by helping to build and establish knowledge regarding a particular phenomenon (Block et al. 2022). The foundation of empiricism rests on the idea that science never provides certainty because there are consistent threats to validity, and thus certainty can be enhanced through more studies that address a similar or the same research question or hypothesis. Additionally, with the passage of time, and particularly prevalent among economic implications, replication can foster greater certainty (or uncover uncertainty) regarding age-old theories.

There are several types of replication studies as outlined by Block et al. (2022), including literal replication (where the study mirrors the original), constructive replication (where a study enhances external or internal validity or both but maintains characteristics of the original study), quasi-random replication (where a study differs from without clearly enhancing the original study), confounded replication (where the study may have lower external validity but improved internal validity), and regressive replication (where a study is similar on all quality dimensions except for one; Block et al. 2022).

Applications utilizing the aforementioned developed research model can be conducted for a variety of different replication types. For example, in constructive replication, authors may decide to leverage previous studies and follow this model to detect discrepancies or adjustments of the study's hypotheses or results. Given the lack of exemplary CV use standards, there are a variety of studies that can be replicated that may enhance contributions utilizing constructive replication with the research model. The model may also be utilized for confounded replication, where the amount of data may be lacking but enhanced internal validity can be achieved by following the aforementioned CV standards. The model can also be tested by leveraging replication through regressive techniques, where a step is taken away (i.e., theory support or the use of isomorphism) to examine the results and compare to the original research results.

Replication studies have been and to some extent continue to be criticized for their "lack of theoretical contribution" (Block et al. 2022). Nevertheless, they remain critically important for empirical research and enhancing theoretical knowledge. The model presented in this research provides a unique opportunity to be leveraged as a tool for constructive and confounded replication.

5 Conclusion

The purpose of this paper was to (1) assess the extent to which researchers follow standards regarding the use and reporting of CVs; (2) identify the main weaknesses as they relate to CV use and reporting in literature; and (3) provide recommendations to researchers on the selection, inclusion, and use of CVs in their research. Contributing to the current literature on CV use in the management literature (e.g., Bernerth and Aguinis 2016), this study not only provides a domain-focused analysis but also a micro-perspective variable-level analysis by providing common CVs via DVs utilized in OM, SCM, and IS/KS domain research. Moreover, we identified the extent of weaknesses through a frequency analysis of 10 standards via the micro-perspective (e.g., the three most commonly utilized DVs in the OM, SCM, and IS/KM domains and their associated CVs) and the macro-perspective (e.g., the 10 most frequently utilized CVs across domains). These results indicate significant differences at micro- and macro-levels when it comes to different standards. Isomorphism and a lack of theoretical explanation were the two most common issues presented. As such, we presented a refined "Decision Making Tree Summarizing Sequential Steps in the Process of Selecting CVs" (Bernerth and Aguinis 2016).

Appendix

See Table 8.

Table 8 Control variable standards by method

Method	Justification (%)	Citation (%)	Previous (%)	Anticipated (%)	Found (%)	Incremental (%)	Eliminate (%)	Theoretical (%)	Analysis (%)	Process (%)	Multiple (%)
Case/Action research**	70	70	10	20	30	0	0	0	70	60	70
Event study**	40	40	40	0	0	0	0	0	100	0	40
Experiment	29	22	24	19	10	8	0	8	75	71	38
Survey	56	52	27	22	21	11	4	10	80	49	38
Archival	53	43	27	18	18	4	1	8	79	44	33
Multi-method	69	67	46	16	23	9	7	11	85	57	50

*Designations are not mutually exclusive

**Designations have significantly lower samples sizes and may not be indicative of actual control variable standard use

Author's contribution All authors contributed to the study conception and design equally. All authors read and approved the final manuscript.

Funding The authors declare that no funds, grants, or other support were received during the preparation of this manuscript.

Data availability The data set analyzed in the study are available from the corresponding author on reasonable request.

Declarations

Conflict of interest The authors have no relevant financial or non-financial interests to disclose.

References

Aguilar MS (2020) Replication studies in mathematics education: what kind of questions would be productive to explore? Int J Sci Math Educ 18(1):37–50. https://doi.org/10.1007/s10763-020-10069-7

Atinc G, Simmering MJ, Kroll MJ (2012) Control variable use and reporting in macro and micromanagement research. Organ Res Methods 15(1):57–74. https://doi.org/10.1177/1094428110397773

Becker TE (2005) Potential problems in the statistical control of variables in organizational research: a qualitative analysis with recommendations. Organ Res Methods 8(3):274–289. https://doi.org/10.1177/1094428105278021

Becker TE, Atinc G, Breaugh JA, Carlson KD, Edwards JR, Spector PE (2016) Statistical control in correlational studies: 10 essential recommendations for organizational researchers. J Organ Behav 37(2):157–167. https://doi.org/10.1002/job.2053

Bernerth JB, Aguinis H (2016) A critical review and best-practice recommendations for control variable usage. Pers Psychol 69(1):229–283. https://doi.org/10.1111/peps.12103

Bernerth JB, Cole MS, Taylor EC, Walker HJ (2018) Control variables in leadership research: a qualitative and quantitative review. J Manag 44(1):131–160. https://doi.org/10.1177/0149206317690586

Block J, Brändle L (2022) Looking back to look ahead: MRQ's past five years of evidence-based management research in numbers. Manag Rev Q 72(4):917–926. https://doi.org/10.1007/s11301-022-00303-7

Block JH, Fisch C, Kanwal N, Lorenzen S, Schulze A (2022) Replication studies in top management journals: an empirical investigation of prevalence, types, outcomes, and impact. Manag Rev Q. https://doi.org/10.1007/s11301-022-00269-6

Bonett DG (2012) Replication-extension studies. Curr Dir Psychol Sci 21(6):409–412. https://doi.org/10.1177/0963721412459512

Carlson KD, Wu J (2012) The illusion of statistical control: control variable practice in management research. Organ Res Methods 15(3):413–435. https://doi.org/10.1177/1094428111428817

Cuervo-Cazurra A, Andersson U, Brannen MY, Nielsen BB, Rebecca Reuber A (2016) From the editors: can I trust your findings? Ruling out alternative explanations in international business research. J Int Bus Stud 47:881–897

Curry A (2019) Across the great divide: a literature review of management accounting and operations management at the shop floor. Manag Rev Q 69(1):75–119. https://doi.org/10.1007/s11301-018-0147-5

DiMaggio PJ, Powell WW (1983) The iron cage revisited: Institutional isomorphism and collective rationality in organizational fields. American Sociological Review, 147–160

Durach CF, Kembro J, Wieland A (2017) A new paradigm for systematic literature reviews in supply chain management. J Supply Chain Manag 53(4):67–85

Edwards JR (2008) To prosper, organizational psychology should… overcome methodological barriers to progress. J Organ Behav Int J Ind Occup Organ Psychol Behav 29(4):469–491

Fisher M, Olivares M, Staats BR (2020) Why empirical research is good for operations management, and what is good empirical operations management? Manuf Serv Oper Manag 22(1):170–178

Fraser H, Barnett A, Parker TH, Fidler F (2020) The role of replication studies in ecology. Ecol Evol 10(12):5197–5207. https://doi.org/10.1002/ece3.6330

Greenland S, Pearl J (2011) Adjustments and their consequences—collapsibility analysis using graphical models. Int Stat Rev 79(3):401–426
Hallgren KA (2012) Computing inter-rater reliability for observational data: an overview and tutorial. Tutor Quant Methods Psychol 8(1):23
Hansen C, Steinmetz H, Block J (2022) How to conduct a meta-analysis in eight steps: a practical guide. Manag Rev Q 72(1):1–19. https://doi.org/10.1007/s11301-021-00247-4
Helmuth CA, Craighead CW, Connelly BL, Collier DY, Hanna JB (2015) Supply chain management research: key elements of study design and statistical testing. J Oper Manag 36:178–186
Hill TD, Davis AP, Roos JM, French MT (2020) Limitations of fixed-effects models for panel data. Sociol Perspect 63(3):357–369
Hochrein S, Glock CH, Bogaschewsky R, Heider M (2015) Literature reviews in supply chain management: a tertiary study. Manag Rev Q 65(4):239–280. https://doi.org/10.1007/s11301-015-0113-4
Hsieh HF, Shannon SE (2005) Three approaches to qualitative content analysis. Qual Health Res 15(9):1277–1288
Janis I (1965) The problem of validating content analysis. The content analysis reader, 358–375
Johns G (2006) The essential impact of context on organizational behavior. Acad Manag Rev 31(2):386–408. https://doi.org/10.5465/amr.2006.20208687
Krippendorf K (2004) Content analysis: an introduction to its methodology, 2nd edn. Sage, Thousand Oaks, CA
Lavenberg SS, Welch PD (1981) A perspective on the use of control variables to increase the efficiency of Monte Carlo simulations. Manag Sci 27(3):322–335
Lehmann EL (2012) "Student" and small-sample theory. Selected works of EL Lehmann, 1001–1008
Light RJ (1971) Measures of response agreement for qualitative data: some generalizations and alternatives. Psychol Bull 76(5):365
Li M (2021) Uses and abuses of statistical control variables: Ruling out or creating alternative explanations? J Bus Res 126:472–488
Mayring P (2008) Qualitative inhaltanalyse—Grundlagen und Techniken (qualitative content analysis). Beltz Verlag, Weinheim
Miller J, Davis-Sramek B, Fugate BS, Pagell M, Flynn BB (2021) Editorial commentary: addressing confusion in the diffusion of archival data research. J Supply Chain Manag 57(3):130–146
Miller CC, Washburn NT, Glick WH (2012) PERSPECTIVE—The Myth of firm performance. Organ Sci 24(3):948–964
Moher D, Liberati A, Tetzlaff J, Altman DG, PRISMA Group* T (2009) Preferred reporting items for systematic reviews and meta-analyses: the PRISMA statement. Ann Intern Med 151(4):264–269
Nielsen BB, Raswant A (2018) The selection, use, and reporting of control variables in international business research: a review and recommendations. J World Bus 53(6):958–968
Ramos J (2003) Using TF-IDF to determine word relevance in document queries (Vol. 242). Proceedings of the first instructional conference on machine learning
Scott WR, Meyer JW (1994) Institutional environments and organizations: structural complexity and individualism. Sage, New York, NY
Seuring S, Gold S (2012) Conducting content-analysis based literature reviews in supply chain management. Supply Chain Manag Int J 17(5):544–555
Shaw JD (2017) Advantages of starting with theory. Acad Manag J 60(3):819–822
Spector PE, Brannick MT (2011) Methodological urban legends: the misuse of statistical control variables. Organ Res Methods 14(2):287–305. https://doi.org/10.1177/1094428110369842
Spector PE, Zapf D, Chen PY, Frese M (2000) Why negative affectivity should not be controlled in job stress research: don't throw out the baby with the bath water. J Organ Behav 21(1):79–95
Steinmetz H, Block J (2022) Meta-analytic structural equation modeling (MASEM): new tricks of the trade. Manag Rev Q 72(3):605–626. https://doi.org/10.1007/s11301-022-00293-6
Stremersch S, Gonzalez J, Valenti A, Villanueva J (2023) The value of context-specific studies for marketing. J Acad Mark Sci 51(1):50–65. https://doi.org/10.1007/s11747-022-00872-9
Tabachnick BG, Fidell LS (2000) Computer-assisted research design and analysis. Allyn & Bacon Inc, Boston, MA
Terwiesch C, Olivares M, Staats BR, Gaur V (2019) OM forum—A review of empirical operations management over the last two decades. Manuf Serv Oper Manag 22(4):656–668. https://doi.org/10.1287/msom.2018.0755

Van Exel J, De Graaf G (2005) Q methodology: a sneak preview. https://www.betterevaluation.org/sites/default/files/vanExel.pdf (Accessed 15 April 2023)

Authors and Affiliations

Carla Curado[1] · Mírian Oliveira[1,2] · Dara G. Schniederjans[3] · Eduardo Kunzel Teixeira[4]

✉ Dara G. Schniederjans
schniederjans@uri.edu

Carla Curado
ccurado@iseg.utl.pt

Mírian Oliveira
miriano@pucrs.br

Eduardo Kunzel Teixeira
eduardo.kunzel@unoesc.edu.br

1 ADVANCE/CSG, ISEG, Universidade de Lisboa, Rua Miguel Lupi, 20, 1200-725 Lisbon, Portugal

2 Escola de Negócios, Pontifícia Universidade Católica do Rio Grande do Sul, Av. Ipiranga, 6681-Prédio 50, Porto Alegre, RS 90619-900, Brazil

3 University of Rhode Island, Kingston, RI 02881, USA

4 Graduate School of Management, University of the West of Santa Catarina, Chapecó, SC 89813-000, Brazil