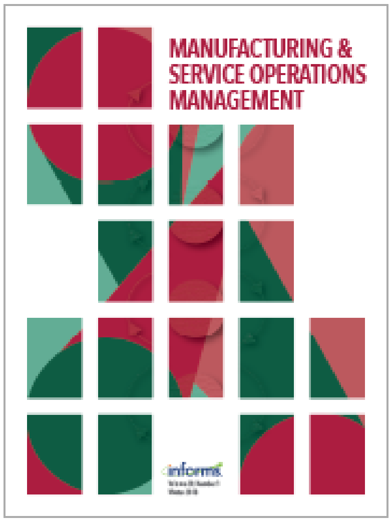
This article was downloaded by: [190.153.198.253] On: 03 May 2020, At: 16:42
Publisher: Institute for Operations Research and the Management Sciences (INFORMS)
INFORMS is located in Maryland, USA



Manufacturing & Service Operations Management

Publication details, including instructions for authors and subscription information:
http://pubsonline.informs.org

Why Empirical Research Is Good for Operations Management, and What Is Good Empirical Operations Management?

Marshall Fisher, Marcelo Olivares, Bradley R. Staats

To cite this article:
Marshall Fisher, Marcelo Olivares, Bradley R. Staats (2020) Why Empirical Research Is Good for Operations Management, and What Is Good Empirical Operations Management?. Manufacturing & Service Operations Management 22(1):170-178. https://doi.org/10.1287/msom.2019.0812

Full terms and conditions of use: https://pubsonline.informs.org/Publications/Librarians-Portal/PubsOnLine-Terms-and-Conditions

This article may be used only for the purposes of research, teaching, and/or private study. Commercial use or systematic downloading (by robots or other automatic processes) is prohibited without explicit Publisher approval, unless otherwise noted. For more information, contact permissions@informs.org.

The Publisher does not warrant or guarantee the article's accuracy, completeness, merchantability, fitness for a particular purpose, or non-infringement. Descriptions of, or references to, products or publications, or inclusion of an advertisement in this article, neither constitutes nor implies a guarantee, endorsement, or support of claims made of that product, publication, or service.

Copyright © 2019, INFORMS

Please scroll down for article—it is on subsequent pages

With 12,500 members from nearly 90 countries, INFORMS is the largest international association of operations research (O.R.) and analytics professionals and students. INFORMS provides unique networking and learning opportunities for individual professionals, and organizations of all types and sizes, to better understand and use O.R. and analytics tools and methods to transform strategic visions and achieve better outcomes.
For more information on INFORMS, its publications, membership, or meetings visit http://www.informs.org

informs
http://pubsonline.informs.org/journal/msom

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 22, No. 1, January–February 2020, pp. 170–178
ISSN 1523-4614 (print), ISSN 1526-5498 (online)

20th Anniversary Invited Article

Why Empirical Research Is Good for Operations Management, and What Is Good Empirical Operations Management?

Marshall Fisher,[a] Marcelo Olivares,[b] Bradley R. Staats[c]

[a] The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104; [b] Department of Industrial Engineering, University of Chile, Of 624 Santiago, Chile; [c] Kenan–Flagler Business School, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599

Contact: fisher@wharton.upenn.ed, http://orcid.org/0000-0002-5800-240X (MF); molivares@u.uchile.cl, http://orcid.org/0000-0003-4443-9261 (MO); bstaats@unc.edu, http://orcid.org/0000-0002-2674-5831 (BRS)

Received: March 29, 2019
Revised: April 24, 2019
Accepted: May 1, 2019
Published Online in Articles in Advance: November 8, 2019

https://doi.org/10.1287/msom.2019.0812

Copyright: © 2019 INFORMS

Abstract. Empirical research in operations management has increased steadily over the last 20 years. In this paper, we discuss why this is good for our field and offer some comments on the qualities we admire in an empirical operations management paper.

History: This paper has been accepted for the *Manufacturing & Service Operations Management* 20th Anniversary Special Issue.

Funding: This work was supported by the Complex Engineering Systems Institute [CONICYT PIAFB0816] and the Fishman Davidson Center for Service and Operations Management, the Wharton School, University of Pennsylvania.

Keywords: empirical research • causal inference • operations management

1. Introduction

Given that the field of operations management (OM) traces its roots to the empirical studies of Frederick Taylor and others of the scientific management movement (Smiddy and Naum 1954), it is not surprising that in its first year of publication in 1999, *Manufacturing & Service Operations Management* (*M&SOM*) published multiple empirical papers. But it was just barely multiple, sitting at two. At least it showed the journal's willingness to acknowledge a still early movement within modern OM. Over the ensuing eight years, those two publications represented one third of the total empirical operations papers published in *M&SOM* (Terwiesch et al. 2019). In response to this dearth of empirical research, one of us published a paper in *M&SOM* in 2007 noting that empirical research in OM lagged other disciplines and citing the advantages to our field of more empirical research (Fisher 2007). Recently, two of us coauthored a *M&SOM* paper noting that, in fact, much more empirical research has been generated. *M&SOM* now publishes five empirical papers per year, and the trend continues to increase (Terwiesch et al. 2019).

Although empirical OM has grown, in part, because of the support of *M&SOM*, much remains to be done. This article pursues three objectives in order to take stock of where empirical OM is and where it still can go. First, we explore the dual questions of what makes an OM paper empirical and when an empirical paper is considered OM. This permits us not only to define the universe of research that we explore in this paper but also to highlight key aspects of empirical operations research that distinguish it from empirical work in other fields, such as economics, which has been the paradigm on which recent empirical OM research has been built. Second, we seek to identify how empirical research has contributed to the academic field of OM as well as to real-world operations. We focus on two types of contributions: (1) prescriptions on how to improve operational decisions and (2) identification and description of a phenomenon observed in practice. Finally, we offer our thoughts, honed through numerous years as authors, reviewers, and editors, on how to evaluate the quality of empirical OM research.

2. What Makes an OM Paper Empirical? When Is an Empirical Paper Considered OM?

In our view, a minimum requirement for a paper to be empirical is the use of real data. These data are often extracted from the regular operations of an organization—sales, inventory, labor costs, prices, orders, and so forth—which we refer to as *field data*. These data can be obtained directly from an organization on topics such as manufacturing lines (Fisher and Ittner 1999), software projects (Narayanan et al. 2009, Huckman and Staats 2011), hospitals and doctors (Olivares et al. 2008, Kc and Staats 2012), or retail stores (Kesavan et al. 2014). Field data can also be obtained indirectly from third parties that collect industry data (e.g., Compustat, WardsAuto; see Gaur et al. 2005 and Moreno and Terwiesch 2015) or collected independently by the researchers, for example, web-crawling transaction information from a market platform (Li et al. 2015, Fisher et al. 2017a) or processing images to measure lengths of queues (Lu et al. 2013). Data can also be collected via survey instruments, such as surveying individuals about their feelings and beliefs on knowledge sharing and process improvement (Tucker 2007, Siemsen et al. 2009) or project teams and customers to understand software team performance (Narayanan et al. 2011).

The next question to consider is how much data are required for a paper to be empirical. It is easiest to answer this question by providing examples from existing research. To do this, we first highlight that we believe empirical research has enhanced OM in two important ways:

1. Providing direct prescriptions on how to improve operational decisions
2. Identifying evidence of a phenomenon that happens in practice

Considering the first, we note that OM has a long history of impact research. These projects typically involve the careful analysis of an operational problem and then deployment of an algorithmic solution. Empirical analysis impacts both the initial analysis and algorithmic deployment, such as through the estimation of parameters and predictive models or analysis of the effort's impact. For example, Fisher and Jaikumar (1981) develop a heuristic to assign customers to vehicle routes. If the authors had stopped there, it would not have been an empirical paper. However, they evaluate their heuristic's performance using data on distances and travel times from real problem instances. Or consider assortment planning research (Kök and Fisher 2007, Fisher and Vaidyanathan 2014, Rusmevichientong and Shmoys 2014), in which the authors develop optimization models that use as an input a demand model that accounts for substitution patterns across product variants. Implementing this approach requires calibrating the demand model with empirical research to infer the substitution patterns from actual purchase data.

In vehicle routing, the estimation of travel times is about 10% of the work, whereas the rest is devoted to the optimization. In assortment planning, estimation of the demand and substitution patterns is usually 90% of the work. From our perspective, the percentage is less vital to determining whether a paper is empirical, and instead, the data analysis should be instrumental in tackling the research question. In other words, rather than using a strict categorization of OM work as either analytical or empirical, a reader's judgment is necessary.

We note that unlike most empirical research these days, neither of these examples apply statistical tools to a large data set, so certainly "Stata-/R-free" empirical research is possible. But many examples of prescriptive empirical research involve statistical analysis. For example, Fisher et al. (2017b) analyze store sales and staffing data from a retail chain to identify overstaffed and understaffed stores.

The bullwhip effect is probably the best-known example of the second type of empirical research, investigating a phenomenon that happens in practice. The bullwhip effect is also a useful context to consider whether the primary contribution of a paper is empirical or analytical. As noted earlier, in empirical research, the data analysis should be instrumental in tackling the research question. Consider the seminal paper on the bullwhip effect (Lee et al. 1997). Although the paper uses some data as a case study, the main contributions of the paper are based on analytical models that provide possible mechanisms to explain the observed phenomenon. Take away the data, and the contribution of the paper is largely the same. In contrast, Cachon et al. (2007) study the bullwhip effect with data across industries, using statistical methods to validate the extent to which the bullwhip effect generalizes across industries and supply chain levels. The central contribution of the paper comes from the data. Finally, Bray and Mendelson (2015) use detailed data from a single industry (automobile), using structural estimation to uncover multiple mechanisms that enhance or mitigate the bullwhip effect. Here the contribution arises from both the data and the estimation methodology. These three papers help us understand what make a paper empirical and also illustrate the value of a diversified mix of empirical and analytical research on a particular operational issue.

Although meaningful use of real data is necessary for a paper to be empirical, what characterizes an empirical study as OM? This is a difficult question. Some traditional topics, such as inventory, supply chain, capacity management, process design, labor productivity, and service operations, are obviously OM. But how about price optimization? Is it OM or marketing? In our view, debating this is pointless. Instead, we think that revealed preference is the best approach to answering this question. If the work is conducted by a researcher in the field of operations or is published or aimed to be published in an OM journal, then we'd regard it as OM. We note that this is also the approach taken by Terwiesch et al. (2019). Although it is true that the line gets blurry for flagship journals such as *Management Science* or *Operations Research*, which aim at crossing multiple disciplines, the key deciding factor is whether the work is interesting for researchers in the OM field.

In summary, we would say that research is in the domain of "empirical OM" if (1) the main contribution of the paper arises from the use of real operations data and (2) the topic of the research question is of interest to OM researchers. We feel that this "big tent" view of empirical research in OM is important to permit the field to innovate. Modern-day work has evolved significantly from the manual labor contexts of Frederick Taylor and *scientific management*, and we believe and hope that it will continue to evolve.

3. Prescriptions on How to Improve Operational Decisions

In the next two sections, we consider what the two types of empirical research—prescribing decisions and providing evidence of phenomena that occur in practice—can contribute to OM. Figure 1 provides an organizing framework for these two sections. We begin by exploring the former topic.

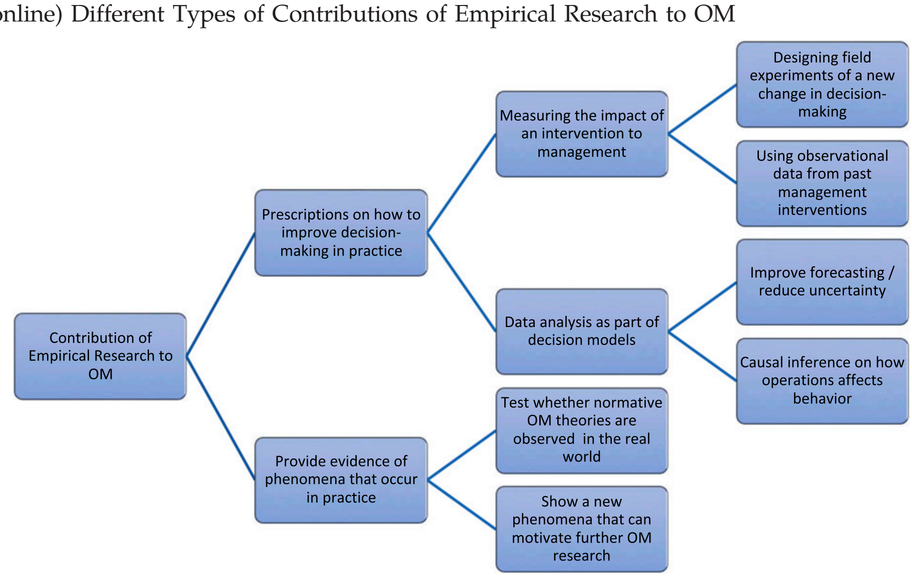
A primary objective of OM is to conduct rigorous research that is theoretically generalizable and can also be used by managers to improve their decision making. Empirical research has played a critical role in the development of the field in two ways: (1) providing valuable inputs to models that aid in making operations decisions and (2) evaluating the impact of innovative managerial practices. Many empirical papers do both of these: formulate and implement a decision model and then evaluate the impact of that intervention. In other cases (MacDuffie 1997), the innovation was developed by others, and the researcher's task is to answer the questions "Did it work?" and "Why did it work?"

3.1. Using Data Analysis to Implement Decision Models in Practice

Research in OM has a long-standing tradition of developing analytical models to help managers improve decisions. Inventory, revenue management, and labor staffing are classical examples in which these models have impacted practice. Data analysis is needed as input to implement these models. An inventory management model requires demand forecasts and a probability distribution of demand (Silver et al. 1998), revenue management requires a function that describes the sales response to changes in prices (Bitran and Caldentey 2003), and a staffing model for call centers requires statistical models for the incoming calls and service times (Gans et al. 2003). Traditionally, most of the research effort in developing this work went to the optimization and stochastic models, but over the last decade, the empirical portion has increased significantly. New data sources and the development of novel statistics/econometric methods permit more sophisticated modeling of the behavior of customers and employees, which are then incorporated to enrich the models and algorithms implemented in practice. We note that this broad class of models is now often referred to as *prescriptive analytics*. However, as the following examples show, prescriptive analytics often involves both the prediction of future events (predictive analytics) and then recommendations as a result.

First, we consider research in inventory management. In general, reducing demand uncertainty helps to lower inventory costs (Gerchak and Mossman 1992). Using real data from Sport Obermeyer, a ski apparel supplier, Fisher and Raman (1996) show that early season sales help to improve the forecast of total season sales. They develop an accurate response system in which an initial order for about half of production is placed before seeing any sales data, and then the remaining supply is ordered based on the improved forecast from initial sales. They create a model to make both of these supply decisions as well as a method to estimate the probability distribution of demand needed in the model. Note that in this example, the empirical analysis is focused on predictive analytics; causal inference is not needed to feed the decision model.

Figure 1. (Color online) Different Types of Contributions of Empirical Research to OM



When planning the inventory of multiple products in an assortment, a more sophisticated demand model is needed to account for how customers substitute across products when their preferred option is not available. Anupindi et al. (1998) approach this problem using data from vending machines and test their methodology for categories with six products. However, their method is difficult to scale because the estimated number of substitution patterns increases exponentially with the number of products.

Random utility models (RUMs) (see Ben-Akiva and Lerman 1985 and Train 2009) provide a general framework to capture substitution patterns based on microeconomic principles. Musalem et al. (2010) and Vulcano et al. (2012) develop methods to estimate these types of models when the demand data are censored because of product stock-outs. There is now extensive research developing decision models for assortment, inventory, and pricing optimization using different variations of RUMs (Davis et al. 2014, Rusmevichientong and Shmoys 2014), some of which have been applied in practice (Fisher and Vaidyanathan 2014; Harsha et al. 2019).

Labor staffing is another important research stream in which data analysis has been used to improve the implementation of decision models. One example is call centers, at which detailed records of incoming calls, waiting times, and service times allow modeling the stochastic behavior of callers and the service times needed to optimize server capacity (Gans et al. 2003, Brown et al. 2005). Although much of the work focuses on forecasting the variability of arrivals and processing times, setting appropriate service levels requires understanding how customers respond to waiting time, which, in contrast to the forecasting work, requires causal inference. A growing stream of empirical literature analyzes call center customer abandonment in response to waiting and delay announcements (Akşin et al. 2013, 2016).

Labor staffing is also studied frequently in retail stores, at which the decision requires balancing staffing costs with the increase in revenue from reducing waiting times. Lu et al. (2013) measure how customer purchases are affected by waiting times in the context of supermarkets; Allon et al. (2011) analyze demand responses to waiting time in the fast-food industry. Both studies show that waiting can impact sales substantially and is a fundamental input for labor capacity planning.

Retail associates can also increase sales by assisting customers in their shopping. A standard approach used in the industry is to forecast future sales and plan staffing levels based on the staffing levels that have been used historically to support a comparable sales level. Because staffing levels are set in anticipation of forecasted demand, therefore, the high correlation between sales and staffing levels does not provide a valid measurement of the causal effect of staffing on sales. Causal inference is required in this case, and recent studies have used deviations between the planned and actual staffing levels at stores as an exogenous source of variability to measure the sales response to a change in staffing capacity (Fisher et al. 2009, 2017b). New data on incoming customer store traffic coupled with new methods to analyze these data have also become useful to measure the causal effect of staffing on sales (Perdikaki et al. 2012).

3.2. Measuring the Impact of a Management Practice Innovation

In the preceding section, we discussed how empirical research is important in improving decision making. In these cases, the same individuals who conduct the initial research also evaluate the impact it had on organizational performance. Some of the best examples of this come from the Franz Edelman Award, for which finalists have conducted research to develop new solutions and measured the impact of the research. A good illustration is the work by Caro et al. (2010), Franz Edelman Laureates (in 2009), which uses a field experiment based on a difference-in-difference approach to evaluate the impact of the adoption across different countries and product categories of a new inventory management system devised by them.

Field experiments are the “gold standard” in most scientific disciplines. The most noticeable example of this is the critical role of experiments in the regulatory approval of new pharmaceuticals. Field experiments in OM are generally based on a research design in which a new managerial practice is applied to a (ideally randomly) selected “treatment” group, which is compared with a control group in which the standard management practice is used. Zhang et al. (2017) use a field experiment to measure the impact of introducing tools that enhance social interactions in a massive open online education class in order to generate better learning outcomes for students. Fisher et al. (2017a) devise a competition-based dynamic pricing algorithm for an internet retailer and then show its effectiveness by comparing performance for a “treated” set of products and a matched set of products priced under the legacy method. Fisher et al. (2017b) use retail store associate absenteeism to identify overstaffed and understaffed stores in a retail chain and then use a field experiment whereby they manipulate labor budgets in the understaffed stores to measure the increase in sales.

From our experience in working with companies and other organizations, it is not always feasible to conduct an ideal field study, especially when it involves testing a negative result. In the last example, an ideal test would have also included increasing staffing in overstaffed stores to show this produced no benefit, but it is not surprising that the retailer was unwilling to do this.

In all these cases, researchers were evaluating the impact of an intervention they created. In other cases, the researcher's task is to evaluate the impact of an innovation devised by a company. A classic example of this is the numerous studies that have explored how organizations have rolled out facets of the Toyota Production System, otherwise known as *lean production*, and the impact this has had (MacDuffie 1997, Shah and Ward 2003, Staats et al. 2011).

In cases such as these, it's unlikely that one has the results of an ideal randomized experiment. No automobile company is going choose a random subset of their plants within which to try lean production. Moreover, the adoption of new practices often involves more than one intervention, and it becomes difficult to measure their effect separately. Also, better firms may implement the new practice or firms may launch initiatives within better subunits, thus making causal inference hard.

Estimating the impact of the adoption of information technology (IT) in organizations is a good illustration of this first challenge; company investments in IT are often complemented with other types of resources, including more advanced human capital, and it becomes difficult to tease these apart (see Krueger 1993 and DiNardo and Pischke 1997 for an interesting debate on this topic). Researchers need creativity to design an adequate *identification strategy* that exploits exogenous interventions that are "as good as" a randomized field experiment for the purpose of identifying causality. Following on the example of the impact of IT, Parker et al. (2016) use a natural experiment, an unexpected 12-day government ban on text messages, to measure the impact of a text message service providing price information on agriculture markets in India. As an example in healthcare, Staats et al. (2018) use an exogenous negative announcement from the U.S. Food and Drug Administration to show that experienced physicians may be less likely to change their behavior than their less experienced colleagues.

One cannot always find a clean natural experiment, so a viable alternative is to use a quasi-experimental design that involves adequate econometric techniques to identify the causal effect of an intervention from observational data. One of the most common techniques used is combining cross-sectional and longitudinal data, comparing outcomes before and after the intervention relative to units that did not receive the treatment (called *difference-in-difference*). Gallino and Moreno (2014) use this approach with a furniture retail chain to measure the effect of a new fulfillment option in their online channel (customers could buy online and pick up at a store). This initiative was implemented simultaneously nationwide. To identify a causal effect, the study compared market areas that were far from any store and therefore could not use the new fulfillment channel with markets near to stores, at which the new channel thus was available. The study revealed that online sales *decreased* in markets where the fulfillment option could be used, contrary to what was expected by management. The researchers were able to identify the underlying mechanism of this sales reduction: customers were using the online channel to search products to ensure in-store availability prior to visiting the store where they could touch and feel the product before finalizing the purchase. The increase in sales within stores compensated for the online sales reduction, so overall sales increased. This unexpected result and its underlying mechanism would have been difficult to observe by simply comparing sales before and after the implementation of this fulfillment option, and therefore, the empirical research was key to convincing managers of the program's effectiveness.

Sometimes identifying proper control groups is not possible. Consider measuring the impact of inventory and product variety on sales in a retail store. A naive approach would be to compare store sales at different levels of inventory and variety. However, because managers set their inventories in anticipation of demand, inventory and sales are highly correlated, but this correlation is not attributable to a causal effect. Cachon et al. (2019) study this problem using data from automobile dealerships. They use random supply shocks generated by extreme weather at auto assembly plants as an exogenous source of variation, which can be considered as good as a randomized, controlled trial. Using these weather-induced shocks as an instrumental variable, the authors show a scarcity effect by which lowering inventory (while keeping variety constant) increases demand. The authors then simulate an optimized inventory allocation policy in the dealership network, using the empirical results, that seeks to maximize variety with minimum inventory, which yields a significant sales increase with the same amount of inventory.

4. Describing a Phenomenon Observed in Practice

Some empirical research is not directly aimed at providing prescriptions, instead studying phenomena that occur in practice. This area can be divided into two streams: (1) examining whether normative OM models are consistent with real-world data and (2) discovering new phenomena, which can inspire new research.

4.1. Testing Normative OM Models

The example of the bullwhip effect mentioned earlier highlights how analytical and empirical research may fruitfully interact. Lee et al. (1997) focus on building models to show how and when the bullwhip effect occurred. Although there were thousands of analytical papers exploring the bullwhip effect, it took 10 years for the first empirical paper to be published in *M&SOM* using industry data. Cachon et al. (2007) find that the bullwhip was largely attenuated in industry data when seasonality was considered. However, the authors did find some bullwhip effect in their deseasonalized data and were left to ponder whether variability in demand or uncertainty in demand was driving this effect. To understand this question, Bray and Mendelson (2015) build and estimate a structural model to show that once one incorporates firm-level data, it is clear that the bullwhip effect does exist and that it is driven by both variability and uncertainty in demand. Subsequent work has continued to unpack assumptions within the bullwhip models to test normative theory and in so doing continues to shape subsequent theory (Bray and Mendelson 2015, Mackelprang and Malhotra 2015, Baron et al. 2018, Bray et al. 2019).

As another example, a fruitful interaction between analytical and empirical research has evolved in answering the question, "Do companies increase inventory to offer higher service levels when faced with competitors, or alternatively, do they cut inventory in order to protect themselves from negative effects resulting from overage costs?" (Lippman and McCardle 1997, Bernstein and Federgruen 2005). Olivares and Cachon (2009) test this using cross-sectional data from automobile dealerships in the United States and find that on market entry of competitors, dealerships are likely to hold more inventory in order to compete on service.

A large body of analytical work explores how firms should write contracts to incentivize performance (Cachon 2003). As scholars created increasingly sophisticated models to deal with complex performance demands, there was little verification of these models. Guajardo et al. (2012) investigate the impact of performance-based contracts on aircraft engine reliability. In the analytical literature, the impact of performance-based contracts on reliability was unclear. Using proprietary data, Guajardo et al. (2012) find that performance-based contracts yield significantly better reliability than time-and-materials contracts.

As these studies highlight, empirical operations improve our understanding of firms' operational decisions when partnered with analytical models. Models create a rigorous structure for decision evaluation. The model necessarily requires assumptions about how the world functions. Empirical operations can test both the assumptions and the predictions of models. For years, the field of OM has built models, but comparatively few have been validated with data. Fortunately, as analytical and empirical research work together, it is possible to support some existing work and find gaps with others that require follow-up.

4.2. Exploring New Phenomena

In many contexts, empirical research discovers a new phenomenon, shines light on a managerial practice that has been unstudied, or identifies where conventional wisdom may be wrong. The study of inventory provides a classic example. The many different models built to manage inventory share at least one common thread: they assume that inventory records are accurate. In their seminal work, DeHoratius and Raman (2008) show that this assumption is wrong, introducing inventory record inaccuracy. Not only did this empirical work lead to significant new analytical research in inventory modeling, but it also changed practice. For many years, a chief executive officer of a Fortune 500 retail firm attended one of our classes and shared how this study single-handedly led him to change the way that his firm managed inventory in its stores.

Studying inventory yields other new areas of research. For example, Kesavan et al. (2010) show that inventory data can improve sales forecasts for U.S. retailers and identify an inefficiency in public stock markets. The study arose, in part, from the practices of an investor (Raman et al. 2006). Other anomalous behavior identified, in part, through empirical studies of operations include showing that service rates are not exogenous to load but rather endogenous (Kc and Terwiesch 2009), that learning is not just a matter of repetition of task but also repetition of interaction between individuals on a team (Huckman et al. 2009, Huckman and Staats 2011), and that dedicated queues can, in certain circumstances, lead to improved performance over pooled queues (Lu et al. 2013, Song et al. 2015).

5. How to Evaluate the Quality of Empirical OM Research

In this section, we offer thoughts on what we think makes for good empirical research in OM, drawing on our collective experience as authors, reviewers, and editors in the field. Our goal is not to be dogmatic with our propositions but rather to encourage active debate.

First, empirical research should investigate problems that matter to practice. Recently, one of us was talking to a distinguished colleague in another department at a different school, having been introduced as an operations professor. The other scholar's response was that he loved to work with operations' scholars because we cared about how things actually work. We see this point as a meaningful compliment and an important consideration in conducting research.

The field of operations is concerned with the study of work (Terwiesch 2019). Therefore, rooting research in how work is done is a way to increase relevance and, even more important, identify divergences in theory and practice. Those divergences may identify cases in which managers' actions are suboptimal, but often the findings may suggest just the opposite; when managers' actions are inconsistent with theory, it may be the theory that is incomplete (Bowman 1963). Companies' executives are not only critical gatekeepers to data, but they are often the providers of a spark that leads to an interesting research question. We encourage researchers to let these conversations motivate research questions rather than whatever news happens to be shared in the *Wall Street Journal* or *BusinessWeek*.

A second goal is to impact practice. By working with decision makers, we can push to implement the prescriptions of our research. To the extent that we believe that our findings can improve practice, we encourage researchers to try to do just that. Working with data offers an added benefit because it helps to give credibility to our prescriptions and then also can help to evaluate their implementation.

Third, and perhaps most important, good empirical research and sophisticated econometrics should not be seen as interchangeable terms. Terwiesch et al. (2019) find that, over time, the literature has used more sophisticated econometric techniques. To the extent that these techniques correct for challenges in data analysis, this is a step in the right direction. At the same time, this does not mean that a more complex technique is always better. There is a risk of an "arms race" in which individuals use increasingly complex approaches just because they can. One should keep in mind the Albert Einstein dictum that theory should be as simple as possible, but no simpler. Ideally, empirical data should be subjected to the "eyeball test": can you show a simple descriptive plot that reveals the key result? This may not always be possible because controlling for other variables and underlying relationships is often critical. However, keeping Einstein's advice in mind as an author and a reviewer is likely to yield papers that can better impact theory and practice.

Fourth, recent work highlights the importance of causal inference in empirical research (Ho et al. 2017). We agree that in papers that seek to draw causal conclusions, strategies to address endogeneity and other empirical concerns are mandatory. However, not all research has causal objectives. For example, research may seek to identify a relationship's existence. This is particularly true in early studies on a topic. Tucker (2004) shows that individuals engage in operational workarounds with negative performance consequences. Initial work, even without causal identification, was important to show that the effect existed (Tucker and Spear 2006). As understanding of the concept improved, causal studies were pursued (e.g., Tucker 2015).

Another example is data-driven prediction research. Techniques such as machine learning provide interesting opportunities to improve forecasts, often from unstructured data. Examples include Glaeser et al. (2019), in which the authors improve existing predictions using large-scale data to identify attractive retail locations. A running joke among empirical researchers with whom we talk is that when a paper is rejected, the reason is always "endogeneity." In many cases, this is a proper cause for rejection. In other cases, it is not. As authors, we must be mindful of what a lack of causality means and discuss this in our papers in a forthright manner. As reviewers, we must articulate why endogeneity is so problematic, if it is, and be open to the possibility that a lack of causality may be acceptable for publication, depending on the subject under study.

Finally, we wish to discuss *p*-values. We find it fascinating that until recently, a *p*-value of less than 5% flipped a switch whereby one no longer paid attention to the actual value. Alternatively, we have been equally interested in how, in industry, *p*-values receive very little attention. *p*-Values should be taken in context. They provide a useful piece of information, but a *p*-value of less than 0.05 is not inherently good, nor is one greater than 0.05 inherently bad. As a field, we have been lazy in adopting the standards of physical sciences, in which the cost of false adoption of a principle is much higher than the cost of false rejection. But if the purpose of an empirical study is to decide on whether to take an action in a company (e.g., whether to raise or lower a price), the cost of inaction may equal or exceed the cost of false action. This is analogous to legal proceedings, in which, in some contexts, "beyond a reasonable doubt" is the right standard, and in others, it's where is the greater "preponderance of evidence." Others have highlighted the harm such an approach can cause for research because scholars may be tempted to pursue less than ethical means to achieve their goals (Simonsohn et al. 2014). Here we call for a commonsense approach to *p*-values.

6. Conclusion

From modest beginnings, over the last 20 years, empirical research has increased dramatically. Given the rise of data available, the increase in scholars seeking to use these data, and the generation of new methods from econometrics, statistics, and machine learning, it takes little imagination to see how the empirical portion of our field is likely to grow. In this article, we have sought to establish how a growth in empirical operations is good and to offer thoughts on how that growth may proceed in a more impactful way. We hope that the field will find this essay valuable as individuals proceed with their own work.

References

Akşin Z, Ata B, Emadi SM, Su CL (2013) Structural estimation of callers delay sensitivity in call centers. *Management Sci.* 59(12): 2727–2746.

Akşin Z, Ata B, Emadi SM, Su CL (2016) Impact of delay announcements in call centers: An empirical approach. *Oper. Res.* 65(1): 242–265.

Allon G, Federgruen A, Pierson M (2011) How much is a reduction of your customers' wait worth? An empirical study of the fast-food drive-thru industry based on structural estimation methods. *Manufacturing Service Oper. Management* 13(4):489–507.

Anupindi R, Dada M, Gupta S (1998) Estimation of consumer demand with stock-out based substitution: An application to vending machine products. *Marketing Sci.* 17(4):406–423.

Baron O, Callen JL, Segal D (2018) Does the bullwhip matter economically? A cross-sectional firm-level analysis. Working paper, University of Toronto, Toronto.

Ben-Akiva ME, Lerman SR (1985) *Discrete Choice Analysis: Theory and Application to Travel Demand* (MIT Press, Cambridge, MA).

Bernstein F, Federgruen A (2005) Decentralized supply chains with competing retailers under demand uncertainty. *Management Sci.* 51(1):18–29.

Bitran G, Caldentey R (2003) An overview of pricing models for revenue management. *Manufacturing Service Oper. Management* 5(3):203–229.

Bowman EH (1963) Consistency and optimality in managerial decision making. *Management Sci.* 9(2):310–321.

Bray RL, Mendelson H (2015) Production smoothing and the bullwhip effect. *Manufacturing Service Oper. Management* 17(2): 208–220.

Bray RL, Yao YO, Duan Y, Huo J (2019) Ration gaming and the bullwhip effect. *Oper. Res.* 67(2):453–467.

Brown L, Gans N, Mandelbaum A, Sakov A, Shen H, Zeltyn S, Zhao L (2005) Statistical analysis of a telephone call center: A queueing-science perspective. *J. Amer. Statist. Assoc.* 100(469):36–50.

Cachon GP (2003) Supply chain coordination with contracts. *Handbooks Oper. Res. Management Sci.* 11:227–339.

Cachon GP, Gallino S, Olivares M (2019) Does adding inventory increase sales? Evidence of a scarcity effect in U.S. automobile dealerships. *Management Sci.* 65(4):1469–1485.

Cachon GP, Randall T, Schmidt G (2007) In search of the bullwhip effect. *Manufacturing Service Oper. Management* 9(4):457–479.

Caro F, Gallien J, Díaz M, García J, Corredoira JM, Montes M, Ramos JA, Correa J (2010) Zara uses operations research to reengineer its global distribution process. *Interfaces* 40(1):71–84.

Davis J, Gallego G, Topaloglu H (2014) Assortment optimization under variants of the nested logit model. *Oper. Res.* 62(2):250–273.

DeHoratius N, Raman A (2008) Inventory record inaccuracy: An empirical analysis. *Management Sci.* 54(4):627–641.

DiNardo JE, Pischke JS (1997) The returns to computer use revisited: Have pencils changed the wage structure too? *Quart. J. Econom.* 112(1):291–303.

Fisher M (2007) Strengthening the empirical base of operations management. *Manufacturing Service Oper. Management* 9(4): 368–382.

Fisher M, Raman A (1996) Reducing the cost of demand uncertainty through accurate response to early sales. *Oper. Res.* 44(1):87–99.

Fisher M, Vaidyanathan R (2014) A demand estimation procedure for retail assortment optimization with results from implementations. *Management Sci.* 60(10):2401–2415.

Fisher M, Gallino S, Li J (2017a) Competition-based dynamic pricing in online retailing: A methodology validated with field experiments. *Management Sci.* 64(6):2496–2514.

Fisher M, Gallino S, Netessine S (2017b) Setting retail staffing levels: A methodology validated with implementation. Wharton Working Paper, University of Pennsylvania, Philadelphia.

Fisher ML, Ittner CD (1999) The impact of product variety on automobile assembly operations: Empirical evidence and simulation analysis. *Management Sci.* 45(6):771–786.

Fisher ML, Jaikumar R (1981) A generalized assignment heuristic for vehicle routing. *Networks* 11(2):109–124.

Fisher ML, Krishnan J, Netessine S (2009) Are your staffing levels correct? *Internat. Commerce Rev*. 8(2):110–115.

Gallino S, Moreno A (2014) Integration of online and offline channels in retail: The impact of sharing reliable inventory availability information. *Management Sci.* 60(6):1434–1451.

Gans N, Koole G, Mandelbaum A (2003) Telephone call centers: Tutorial, review, and research prospects. *Manufacturing Service Oper. Management* 5(2):79–141.

Gaur V, Fisher ML, Raman A (2005) An econometric analysis of inventory turnover performance in retail services. *Management Sci.* 51(2):181–194.

Gerchak Y, Mossman D (1992) On the effect of demand randomness on inventories and costs. *Oper. Res.* 40(4):804–807.

Glaeser CK, Fisher M, Su X (2019) Optimal retail location: Empirical methodology and application to practice. *Manufacturing & Service Oper. Management* 21(1):86–102.

Guajardo JA, Cohen MA, Kim SH, Netessine S (2012) Impact of performance-based contracting on product reliability: An empirical analysis. *Management Sci.* 58(5):961–979.

Harsha P, Subramanian S, Uichanco J (2019) Dynamic pricing of omnichannel inventories. *Manufacturing Service Oper. Management* 21(1):47–65.

Ho TH, Lim N, Reza S, Xia X (2017) OM forum—Causal inference models in operations management. *Manufacturing Service Oper. Management* 19(4):509–525.

Huckman RS, Staats BR (2011) Fluid tasks and fluid teams: The impact of diversity in experience and team familiarity on team performance. *Manufacturing Service Oper. Management* 13(3): 310–328.

Huckman RS, Staats BR, Upton DM (2009) Team familiarity, role experience, and performance: Evidence from Indian software services. *Management Sci.* 55(1):85–100.

Kc DS, Staats BR (2012) Accumulating a portfolio of experience: The effect of focal and related experience on surgeon performance. *Manufacturing Service Oper. Management* 14(4):618–633.

Kc DS, Terwiesch C (2009) Impact of workload on service time and patient safety: An econometric analysis of hospital operations. *Management Sci.* 55(9):1486–1498.

Kesavan S, Gaur V, Raman A (2010) Do inventory and gross margin data improve sales forecasts for U.S. public retailers? *Management Sci.* 56(9):1519–1533.

Kesavan S, Staats BR, Gilland WG (2014) Labor-mix and volume flexibility: Evidence from a retailer. *Management Sci.* 60(8): 1884–1906.

Kök A, Fisher M (2007) Demand estimation and assortment optimization under substitution: Methodology and application. *Oper. Res.* 55(6):1001–1021.

Krueger AB (1993) How computers have changed the wage structure: Evidence from microdata, 1984–1989. *Quart. J. Econom.* 108(1): 33–60.

Lee HL, Padmanabhan V, Whang S (1997) Information distortion in a supply chain: The bullwhip effect. *Management Sci.* 43(4): 546–558.

Li J, Moreno A, Zhang DJ (2015) Agent behavior in the sharing economy: Evidence from Airbnb. Working Paper Series 1298, Ross School of Business, University of Michigan, Ann Arbor.

Lippman SA, McCardle KF (1997) The competitive newsboy. *Oper. Res.* 45(1):54–65.

Lu Y, Musalem A, Olivares M, Schilkrut A (2013) Measuring the effect of queues on customer purchases. *Management Sci.* 59(8):1743–1763.

MacDuffie JP (1997) The road to "root cause": Shop-floor problem-solving at three auto assembly plants. *Management Sci.* 43(4): 479–502.

Mackelprang AW, Malhotra MK (2015) The impact of bullwhip on supply chains: Performance pathways, control mechanisms, and managerial levers. *J. Oper. Management* 36:15–32.

Moreno A, Terwiesch C (2015) Pricing and production flexibility: An empirical analysis of the U.S. automotive industry. *Manufacturing Service Oper. Management* 17(4):428–444.

Musalem A, Olivares M, Bradlow E, Terwiesch C, Corsten D (2010) Structural estimation of the effect of out-of-stocks. *Management Sci.* 56(7):1180–1197.

Narayanan S, Balasubramanian S, Swaminathan JM (2009) A matter of balance: Specialization, task variety, and individual learning in a software maintenance environment. *Management Sci.* 55(11): 1861–1876.

Narayanan S, Balasubramanian S, Swaminathan JM (2011) Managing outsourced software projects: An analysis of project performance and customer satisfaction. *Production Oper. Management* 20(4): 508–521.

Olivares M, Cachon G (2009) Competing retailers and inventory: An empirical investigation of General Motors' dealerships in isolated markets. *Management Sci.* 55(9):1586–1604.

Olivares M, Terwiesch C, Cassorla L (2008) Structural estimation of the newsvendor model: An application to reserving operating room time. *Management Sci.* 54(1):41–55.

Parker C, Ramdas K, Savva N (2016) Is IT enough? Evidence from a natural experiment in India's agriculture markets. *Management Sci.* 62(9):2481–2503.

Perdikaki O, Kesavan S, Swaminathan JM (2012) Effect of retail store traffic on conversion rate and sales. *Manufacturing Service Oper. Management* 14(1):145–162.

Raman A, Gaur V, Kesavan S (2006) David Berman. Harvard Business School Case 605-081, Harvard University, Boston.

Rusmevichientong P, Shmoys D (2014) Assortment optimization under the multinomial logit model with random choice parameters. *Production Oper. Management* 23(11):2023–2039.

Shah R, Ward PT (2003) Lean manufacturing: Context, practice bundles, and performance. *J. Oper. Management* 21(2):129–149.

Siemsen E, Roth AV, Balasubramanian S, Anand G (2009) The influence of psychological safety and confidence in knowledge on employee knowledge sharing. *Manufacturing Service Oper. Management* 11(3):429–447.

Silver EA, Pyke DF, Peterson R (1998) *Inventory Management and Production Planning and Scheduling* (Wiley, New York).

Simonsohn U, Nelson LD, Simmons JP (2014) P-curve: A key to the file-drawer. *J. Experiment. Psych. General* 143(2):534–547.

Smiddy H, Naum L (1954) Evolution of a "science of managing" in America. *Management Sci.* 1(1):1–31.

Song H, Tucker AL, Murrell KL (2015) The diseconomies of queue pooling: An empirical investigation of emergency department length of stay. *Management Sci.* 61(12):3032–3053.

Staats BR, Brunner DJ, Upton DM (2011) Lean principles, learning, and knowledge work: Evidence from a software services provider. *J. Oper. Management* 29(5):376–390.

Staats BR, KC D, Gino F (2018) Maintaining beliefs in the face of negative news: The moderating role of experience. *Management Sci.* 64(2):804–824.

Terwiesch C (2019) Empirical research in operations management: From field studies to analyzing digital exhaust. *Manufacturing Service Oper. Management*, ePub ahead of print May 15, https://doi.org/10.1287/msom.2018.0723.

Terwiesch C, Olivares M, Staats BR, Gaur V (2019) A review of empirical operations management over the last two decades. *Manufacturing Service Oper. Management*, ePub ahead of print June 19, https://doi.org/10.1287/msom.2018.0755.

Train KE (2009) *Discrete Choice Methods with Simulation* (Cambridge University Press, Cambridge, UK).

Tucker AL (2004) The impact of operational failures on hospital nurses and their patients. *J. Oper. Management* 22(2):151–169.

Tucker AL (2007) An empirical study of system improvement by frontline employees in hospital units. *Manufacturing Service Oper. Management* 9(4):492–505.

Tucker AL (2015) The impact of workaround difficulty on frontline employees' response to operational failures: A laboratory experiment on medication administration. *Management Sci.* 62(4): 1124–1144.

Tucker AL, Spear SJ (2006) Operational failures and interruptions in hospital nursing. *Health Services Res.* 41(3, part 1):643–662.

Vulcano G, Van Ryzin G, Ratliff R (2012) Estimating primary demand for substitutable products from sales transaction data. *Oper. Res.* 60(2):313–334.

Zhang DJ, Allon G, Van Mieghem JA (2017) Does social interaction improve learning outcomes? Evidence from field experiments on massive open online courses. *Manufacturing Service Oper. Management* 19(3):347–367