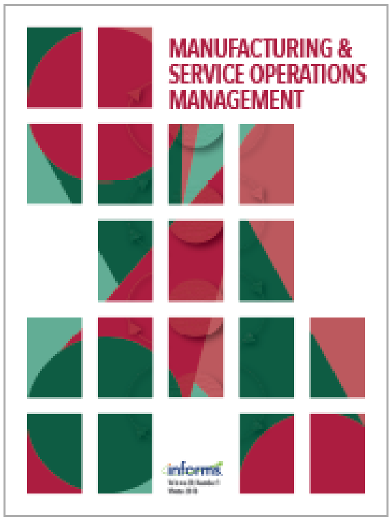


Manufacturing & Service Operations Management

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

OM Forum—Empirical Research in Operations Management: From Field Studies to Analyzing Digital Exhaust

Christian Terwiesch

To cite this article:
Christian Terwiesch (2019) OM Forum—Empirical Research in Operations Management: From Field Studies to Analyzing Digital Exhaust. Manufacturing & Service Operations Management 21(4):713-722. https://doi.org/10.1287/msom.2018.0723

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

MANUFACTURING & SERVICE OPERATIONS MANAGEMENT
Vol. 21, No. 4, Fall 2019, pp. 713–722
ISSN 1523-4614 (print), ISSN 1526-5498 (online)

OM Forum

Empirical Research in Operations Management: From Field Studies to Analyzing Digital Exhaust

Christian Terwiesch[a]

[a] The Wharton School, University of Pennsylvania, Philadelphia, Pennsylvania 19104-6366
Contact: terwiesch@wharton.upenn.edu, http://orcid.org/0000-0002-4050-4348 (CT)

Received: August 28, 2017
Revised: February 15, 2018
Accepted: February 26, 2018
Published Online in Articles in Advance: May 15, 2019

https://doi.org/10.1287/msom.2018.0723

Abstract. The amount of empirical research in operations management that has been published in this journal and in *Management Science* has recently witnessed a significant increase. Beyond this increase in numbers, research questions and empirical methodologies have also changed substantially. Whereas empirical research in the 1990s was primarily carried out at the firm, plant, or project level, recent research focuses much more on microlevel phenomena. This shift in the unit of analysis has been made possible by advances in data generation and collection technologies. This trend is likely to continue as the digitization of modern work creates massive data trails, which archive the state of an operation at a very fine level of granularity. I will refer to such data, which result as a by-product of the digital planning and control of operations, as digital exhaust. As such microlevel data become available abundantly, new opportunities for research arise. However, such data also pose new challenges, including the need for careful theory development and a deep contextual understanding of the underlying data generation process.

Keywords: data collection • econometric analysis • empirical research • primary data

1. Preface

Let me preface this article by expressing how greatly I appreciate being named a distinguished fellow of the Manufacturing and Service Operations Management Society (MSOM). MSOM and INFORMS (Institute for Operations Research and the Management Sciences) have been my scholarly community from the day I joined the INSEAD PhD program a quarter of a century ago. I have repeatedly benefitted from the peer review and editorial leadership of *M&SOM* and *Management Science*. I am particularly grateful to Hau Lee, former editor-in-chief of *Management Science*, who was willing to support my work, despite reviewer comments such as "this is not operations management" or "your R^2 of 0.44 is too low."

I also want to acknowledge how central my time in the INSEAD PhD program has been for my career and my life. PhD classes and seminars at INSEAD were engaging, at times even entertaining, with a healthy tension among the world views of the faculty, including Luk Van Wassenhove, Xavier de Groote, Enver Yücesan, Ludo Van der Heyden, Arnoud De Meyer, and, of course, my dissertation advisor, mentor, and friend, Christoph Loch. With Michael Lapré, I was lucky to have a classmate who was a great econometrician and deeply committed to empirical work.

I had barely started my doctoral training at INSEAD when fellow PhD student Charles Corbett and Luk Van Wassenhove published an article about the state of the operations management field entitled "Crisis? What Crisis?" I also distinctively remember INSEAD seminars featuring Uday Karmarkar, Morris Cohen, and Hau Lee, all of them articulating a crisis in our field and the need for operations management to become more empirical. From my view as a first-year PhD student, the only crisis on my radar was getting through my qualifying exam.

2. Introduction

The call for more empirical research in operations management is not new. In fact, my colleague, Marshall Fisher, dedicated his MSOM Fellows article to this very topic (Fisher 2007) and convincingly argued that empirical work is a necessity for any academic discipline. By drawing comparisons to the fields of physics and medicine, Marshall described the need for empirical work better than I could ever describe it myself. So, I will not comment on why we need more empirical work. Instead, I have three objectives that hopefully help (1) to improve our understanding of the present state of empirical operations management, (2) to document current trends related to what cutting-edge empirical research in operations management looks like, and (3) to provide a vision for the future.

Let me elaborate further on each of these objectives and preview my arguments of this article. My first objective is to demonstrate that the many calls for more empirical research have finally been answered: Our top journals, *Management Science*, *Production and Operations Management* (*POM*), and *M&SOM*, are now increasingly publishing empirical papers. In Section 3, I will carefully document this increase in empirical work and compare it with what has happened in the field of economics.

Beyond documenting this increase in numbers, my second objective is to discuss how the type of empirical research in our field has been changing recently. As part of this second objective, I will discuss two trends (Section 4). The first trend is a shift in the unit of analysis that is used in theory development and empirical analysis. While empirical research in the 1990s was primarily carried out at the firm, plant, or project level, recent research focuses much more on microlevel transactions such as inventory levels and waiting times. I propose that this is a healthy development as it allows for a tighter connection between the mathematical models that our field has developed and the empirical work.

The increasing focus on the microlevel phenomena has been made possible by advances in data generation and collection technologies, which is the second trend I observe. Rather than actively collecting data through surveys or interviews, researchers now are often able simply to access archives of planning and control data and then retrospectively analyze the operational decisions a firm made. As an illustrative example for this approach, consider my work with Morris Cohen, Teck Ho, and Justin Ren on the semiconductor supply chain (Terwiesch et al. 2005). We explored how a buyer of semiconductor equipment shares information with suppliers and how this information sharing helps generate trust between the two parties. Following the traditional paradigm of survey-based research, we could have asked the buyer to what extent data shared with suppliers was truthful, and we could have asked the suppliers to what extent they trusted the forecasts provided by the buyer. Instead, we followed a different approach. The information exchange between buyer and supplier happened via a digital platform. We were granted access to the archives of this platform, which enabled us to retrospectively determine all forecast information provided by the buyer and how the data were then used in production and delivery by the suppliers. It was never the purpose of the platform to support research. We simply analyzed the data trails that were generated in the normal operations of this buyer–supplier relationship. I will refer to such data, which result as a by-product of the digital planning and control of operations, as *digital exhaust*.

Finally, my third objective is to discuss the implications of these two trends on future research in our field (Section 5). As such microlevel data become available abundantly, new opportunities for research arise. However, such data also pose new challenges. In particular, I will express my concerns that this new convenience of data collection might keep us as researchers from immersing ourselves in the complex reality of the operations that we attempt to study. As I will argue, such immersion is not only helpful in understanding the underlying data generation process, but it will also inform our choice of research question and help us turn our results into impact. In light of a growing enthusiasm for empirical research, fascinating new data sources reflecting the availability of digital exhaust, and an endless number of new research questions that will present themselves to those who are willing to immerse themselves in the empirical domains of operations, I conclude with a very positive outlook on our field (Section 6).

3. From an Empirical Discipline to Mathematical Modeling (and Back)

The term operations management has its roots in the Latin word *opus*, which means work. Thus, our field is about analyzing and improving how people work. Nobody in the history of our field is associated with this paradigm more than Frederick Winslow Taylor. Taylor studied bricklaying and pig iron shoveling. He and his colleagues, such as Henry Gantt, and Frank and Lillian Gilbreth, spent their time at the frontline, observing how people worked, generating hypotheses, and testing and refining them. Field experiments, a research method that has recently received a lot of enthusiasm spilling over from the field of economics, were commonly used by Taylor and his peers. In the observation or field experiments, whatever the method, this pioneering work was carried out empirically.

Empirical work in operations management has always been challenging. Unlike accounting or finance, where there exists an abundance of standardized data, operations management data are different. In our pursuit of analyzing how people work, we historically could not rely on archival corporate records (what worker would log in "waste" or "idle time" in his time chart?). Taylor had commented on this data collection challenge: "We can see and feel the waste of material things. Awkward, inefficient, or ill-directed movements of men, however, leave nothing visible or tangible behind" (Taylor 1919, p. 3).

Frustrated by the difficulties of data collection and attracted by the elegance of formal mathematical models, our field turned away from empirical work for much of the second half of the previous century.

Table 1. Empirical Research in Economics

Year	Theory	Empirical	Experiment
1963	52.2	47.8	0
1973	58.8	41.2	0
1983	61.6	37.6	0.8
1993	39.7	56.6	3.7
2003	40.0	56.3	3.7
2013	27.9	63.9	8.2

Source. Hammermesh (http://chrisblattman.com/2012/12/24/six-decades-of-economics-publishing/).
Note. Numbers show the percentages of publications in the top economics journals.

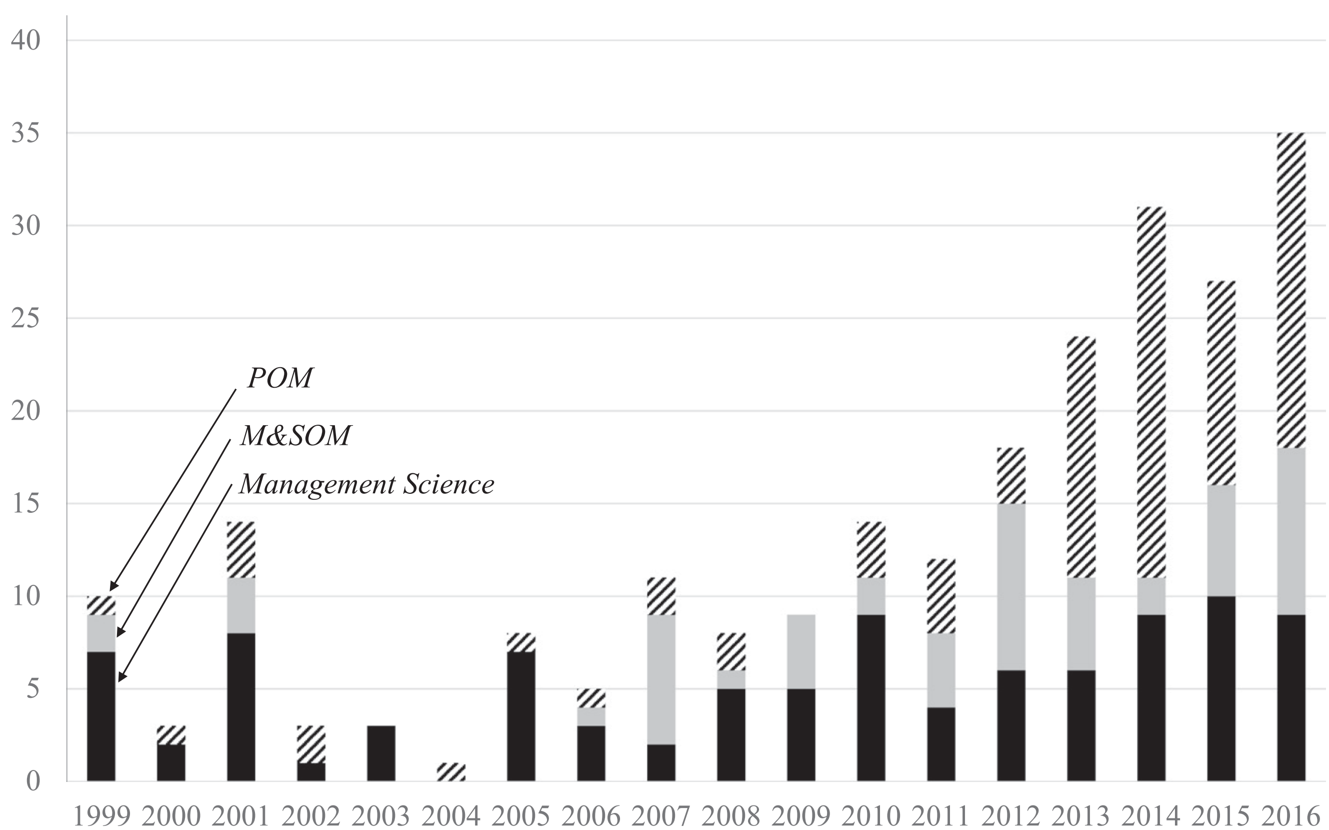
In a recent review of empirical operations management, my coauthors and I (Terwiesch et al. 2019) show that from among all operations articles published in the journals *M&SOM*, *POM*, and *Management Science* in 2004, only one article was empirical.

The drift away from data collection to mathematical modeling is also visible in other fields. Consider the data shown in Table 1, which summarizes the usage of empirical work in the top economics journals. Pure theory papers were on the rise in the 1960s and 1970s, a trend that peaked in 1980s. In contrast to our field, however, even in the most theory heavy days, more than 40% of the economics articles were empirical. Moreover, Table 1 clearly shows that empirical research is on the rise in economics journals. This revival of empirical research in economics begs the question: Can we observe a similar trend in operations management? The answer to this question is a resounding "Yes." Figure 1 shows the dramatic increase in the number of empirical articles published in the three journals mentioned above.

A description of empirical work in operations management would not be complete without pointing to the work of Aleda Roth, Arnoud De Meyer, Marshall Fisher, Roger Bohn, and the departments of Harvard Business School, INSEAD, Wharton, and the University of Minnesota. These authors and others conducted empirical research, and several articles were published in *Management Science* in the 1990s.

To put this data into perspective, I should note that 2004 was an exceptionally bad year for empirical work. In the fall of 2006, Taylor Randall, Vishal Gaur, and I launched a new annual conference for empirical research in operations management. The conference, which now also includes Brad Staats and Marcelo Olivares as cohosts, in many ways mirrors the data shown in Figure 1. In our first year of the event, we had to work hard to find the dozen or so papers that made

Figure 1. Rise of Empirical Operations Management



Source. Terwiesch et al. (2019).

up the conference program. In the 2017 event, we had 35 exceptionally strong submissions to choose from.

4. From Firm Level to Transaction Level and the Emergence of Digital Exhaust

Besides the trend of an increasing quantity of empirical work in operations management, I also see two more subtle trends within the empirical research that has been conducted. I want to describe these trends by comparing the different units of analysis that researchers use in their studies and the frequency with which data are collected.

First, consider the trend in shifting the unit of analysis. In their seminal 1994 *Management Science* paper, Aleda Roth and Jeffrey Miller presented a cluster analysis of 163 firms and their manufacturing strategies. Though the topic was clearly in the domain of operations management, the theoretical framework came from strategy. In fact, not a single *Management Science* paper was referenced in this article, a common pattern in early empirical operations management work. There simply existed a substantial divide between the small set of empirical papers, which were mostly addressing firm-level research questions, and the mainstream mathematical work, which focused primarily on transactional questions related to queueing and inventory models.

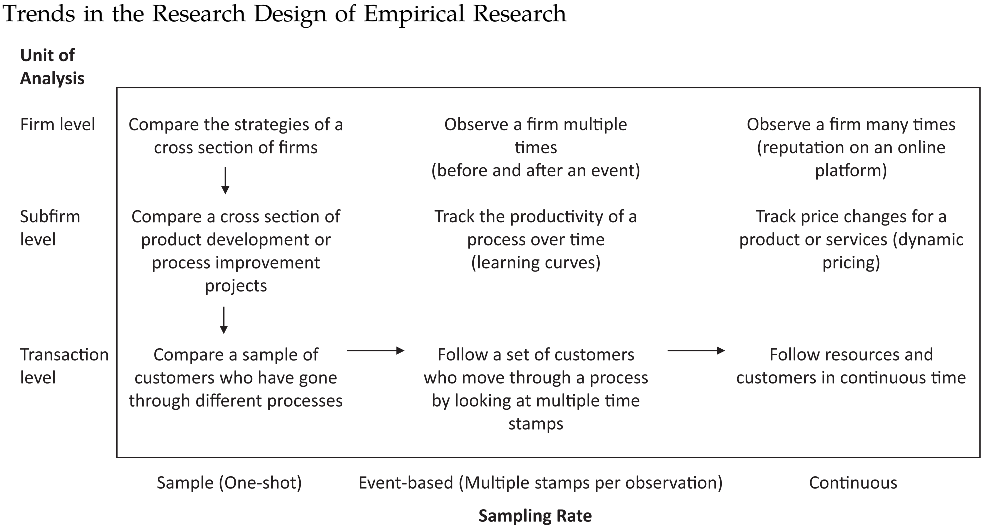
If we move down from the firm level to smaller units of analysis, a similar pattern persists. Examples for such research at the subfirm level include the empirical work on product development by scholars Kim Clark, Takahiro Fujimoto, Marco Iansiti, Stefan Thomke, Karl Ulrich, Kamalini Ramdas, and others, as well as my dissertation work with Christoph Loch, all of which were built on frameworks of Organizational Theory, in particular, Galbraith's information processing model.

As we continue zooming into the operations and take an even more microlevel perspective, the next unit of analysis I would like to feature is the level of individual transactions: calls in a call center, patients in a hospital, or orders in a supply chain. This transaction level has only been studied empirically much more recently. In the analysis of hospital operations, a number of articles look at a large sample (1,000 to more than 100,000) of patient encounters and try to explain some variance in patient outcomes (e.g., KC and Terwiesch 2011). Because it is known which patient was treated at which hospital, researchers can try to explain some of the variation in outcome through variation in structure or process. For example, studies with this research design have shown that hospitals with a higher degree of focus are able to provide better care, an argument that supports the pioneering work by Wickham Skinner.

Consider the framework illustrated in Figure 2. The previously discussed trend of moving from the firm level to the subunit level and from there to the transaction level is visualized along the vertical dimension of the figure. Now, let's turn to the second trend. The horizontal dimension in Figure 2 captures how often the researcher observed the unit of analysis. In most of the previously mentioned examples, researchers observed the unit of analysis only once: one measurement of firm performance, one data point for plant flexibility or project duration, or one length of stay per patient. No matter the unit of analysis, these research designs are inherently cross-sectional.

This contrasts with a "within-unit" research design that observes the unit of analysis multiple times. At the firm level, Vishal Gaur, Marshall Fisher, Kevin Hendricks, Vinod Singhal, and others used panel or event-study based research designs to demonstrate how particular practices or events caused a change in a dependent variable such as financial returns. Similarly, in their empirical analysis of learning curves, Paul Adler, Gary Pisano, Rob Huckman, Roger Bohn, Luk Van Wassenhove, and Michael Lapré used a particular production line as the unit of analysis and observed dependent variables such as productivity measures

Figure 2. Two Trends in the Research Design of Empirical Research



over time. Most recently, digital market places and online platforms allow us to track firm attributes, such as reputation, dynamically over time.

At the transaction level, tracking the unit of analysis over time allows researchers to connect with the mathematical models of process flows and queueing networks. Personally, a big revelation in terms of empirical research happened for me around 2005. My doctoral student, Diwas KC, and I were exploring joint research opportunities with the University of Pennsylvania Health System. The hospital had just created a new workflow for patient transporting. Each transporter was connected with a mobile phone, and jobs were assigned to transporters via text messages. More importantly for us as researchers, every transport assignment was time stamped and archived. Roughly at that same time, the hospital also implemented a patient-tracking system that supported the flow of patients through the hospital. Again, every patient movement was time stamped and archived, creating massive data trails for each patient. This archival data had two advantages to us as researchers: (1) it provided us with multiple time stamps for each patient, similar to a sample path in a queueing network; and (2) it allowed us to retrospectively determine the state of the hospital, including how patients were competing for scarce resources in the system. This enabled a test of some of the commonly held assumptions of queueing models, in particular, the question to what extent processing rates are exogenous with respect to load (KC and Terwiesch 2009).

As more of the empirical operations management research shifted to the analysis of such time-stamped data, a new data problem appeared. As researchers, we only observed time stamps that related to the movement of customers (patients) through the process but not the actual work of the resources (care providers). Consider two patients in an emergency department. Both patients arrive at 8 a.m., both are moved to a treatment room at 9 a.m., and both are discharged by 1 p.m. One of the patients gets an hour of physician attention, imaging and laboratory work, as well some wound care by a nurse. However, the other patient gets five minutes with a physician and then waits several hours for a laboratory test only to be discharged by a nurse without further treatment. If all the data we have are the time stamps of arrival, entry to the treatment room, and discharge, our two patients create the same sample path. In complex processes such as an emergency department, we not only have one resource (a treatment room) but many resources (doctors, nurses, labs, radiology, space) with very complicated interactions among them. To really understand these systems, we also need to track the resources, not just the flow units.

When several years later my doctoral student, Robert Batt, and I worked in the emergency department and articulated this data limitation to the hospital leadership, we received a new data set that was beyond anything I had seen before. To be able to locate the members of the nursing staff quickly, the hospital recently had equipped all nurses with a location tracker. This tracker emitted a signal every couple of seconds that stored time and location in a central database (i.e., one nurse for one shift would leave several thousands of observations behind). Equipped with the architectural drawings of the hospital, we could compute how far and fast the nurses walked, which patients they visited, and how much time they spent at their base station doing computer work (Batt and Terwiesch 2015, Meng et al. 2018). Note that these data were not collected to support academic studies; they were an unobtrusive by-product of the digitization of work. The distinctive characteristic of digital exhaust, as I introduced the term in the introduction, is that it gets created even without the active involvement of the researcher.

A quick comment on the origins of the term digital exhaust. Though I have used such data extensively in my research over the last 15 years, I never used the term before. I first picked it up at a 2017 Big Data panel when a speaker referred to the book, "Digital Exhaust," by Neef (2014). Brad Staats then pointed me to the work by Webb et al. (1999) that coined the term "unobtrusive data collection," which has a long tradition in the social sciences predating digitization.

As illustrative examples of this trend toward analyzing digital exhaust, consider the recent work on handwashing compliance by Staats et al. (2017), which uses radio frequency identification (RFID) data to determine how much care providers adhere to hand hygiene procedures; by Olivares et al. (2013), who use video analysis to analyze the queueing dynamics at a deli counter; by Bavafa et al. (2018), who use time stamps of secure messages exchanged between patients and their doctors; and by Moreno and Terwiesch (2015), who use vehicle pricing and discount data from a large online platform. All these studies cleverly exploit advancements in digitization and sensor technologies, just like our nurse-tracking example described above.

This increase in sampling rate going from a single observation of the unit of analysis to multiple (potentially thousands of) data points per unit of analysis completes the second trend I see in operations management. Both of these trends in empirical research, I propose, are welcomed as they allows us to better connect the empirical work with the abundance of mathematical models that have been created in our field. In fact, Taylor himself was looking at work at the transaction level in continuous time (he observed the process, not just time stamps). So, analyzing digital exhaust moved us back to the roots of our field. However, now that we can automate the process observation, we can collect data at dramatically lower costs per observation, leading to much larger sample sizes.

5. Empirical Work in a World of Abundant Data

If we extrapolate these trends, it is easy to imagine that soon every truck, cab, and personal vehicle will likely be monitored continuously for factors such as adherence to speed limits and stop signs. In fact, companies like UPS, Uber, and Progressive Auto Insurance already have access to this type of data. In its recent developer conference, Microsoft showcased new video technologies to monitor employees and their behavior in the workplace and turn this stream of images into numerical data. Also, Amazon just filed a patent for tracking the hand movements of its warehouse operators.

One thing I find remarkable about these new technologies is that they produce massive data sets of how people work, without any effort from the researcher. This is a sharp contrast to the pains and efforts it took to conduct a careful time and motion study in the past. Finally, Taylor's frustration that "awkward, inefficient, or ill-directed movements of men, however, leave nothing visible or tangible behind" has been addressed (Taylor 1919, p. 3). From a technical perspective, every movement of a worker, every stop of a driver, every mouse click of a white-collar worker, everything we do now leaves something visible behind. Digital exhaust becomes the promised land in which data are available to us researchers at almost zero marginal costs, but what will good research look like under this new paradigm?

My colleague, Marshall Fisher, describes excellent research as an important question well answered (Fisher 2007). According to Marshall, our field fails more often at asking an important question than in the methodology of answering that question (Fisher 2007). Our field, it seems, struggles with the "what" and excels at the "how." There is no reason to believe that abundant data alone will address this struggle. In the remainder of this article, therefore, I articulate five propositions on how our field should and should not take advantage of this new type of data. I illustrate each of them with anecdotes from my own research career. The first three focus more on picking a research question whereas the last two focus more on the methodology. To state the obvious, these propositions are an expression of my personal opinion rather than constituting an empirical study.

5.1. Not All Empirical Research Is Interesting

"Imagine you have access to all the data in the world," I often tell my PhD students, "What is it that you really want to find out?" The challenge of picking a good research question is not going away because of an abundance of data. To the contrary, the only thing that goes away are the excuses why limited data availability drove the study design rather than the lack of author imagination. This makes the question of what is interesting focal to every researcher.

"That is interesting?" is the title of a legendary article written by Murray Davis and published in the *Philosophy of Social Sciences* (Davis 1971). The article was my first reading in my first PhD class at INSEAD, though it took me years to appreciate the power of the article. In the view of Davis, research that is interesting in the social sciences attacks a taken-for-granted result and challenges it. "An interesting proposition was always the negation of an accepted one," he wrote. Davis (1971) proposed that an interesting study follows a four-step logic:

- The author articulates a taken-for-granted assumption ("it has long been thought...").
- The author then adduces one or more propositions that deny what has traditionally been assumed ("but this is false...").
- The body of the work then uses various methodologies to show that the old propositions are wrong and the new ones are correct.
- The conclusion of the article discusses the practical consequences of the new propositions.

In his *M&SOM* article, my colleague, Gerard Cachon, discusses the implication of this logic for the field of operations management (Cachon 2012). He mostly agrees with the original logic proposed by Davis, though he adds the dimension of managerial importance as a second driver of impact.

I have tried to use Davis' logic in all my papers, from my dissertation work on concurrent engineering, where it has long been thought that working in parallel speeds things up, but might actually lead to rework, which more than offsets the benefit from working in parallel, to my most recent paper in *Management Science*, where it seems plausible that allowing patients to send emails to their doctors leads to fewer office visits, but actually increases the number of times a patient is seen, as emails remove the friction of contacting the doctor.

In summary, the abundance of data does not free us from asking interesting research questions. Despite my admiration of Davis' article and how much his logic has helped me in the past, I have two concerns with this being the guiding rule for publishing future research in our field, which I try to articulate in the following two propositions.

5.2. Management Is Not All About the β

In a 2015 article published by the *American Economic Review*, Jon Kleinberg and coauthors provide the following decision problems: "One policymaker facing a drought must decide whether to invest in a rain dance to increase the chance of rain. Another seeing clouds must decide whether to take an umbrella to work to avoid getting wet on the way home" (Kleinberg et al. 2015, p. 491). Both decisions, the authors argue, would benefit from an empirical study of rain. However, as the authors argue, two very different empirical studies are needed to help the policymakers.

When we as researchers build empirical models, we often rely on regression models of the form $Y = \beta X + \varepsilon$. Most of social science research as well as most of operations management has focused on estimating the β. Our research designs have gotten better at addressing endogeneity challenges (see Ho et al. 2017), and instrumental variable estimation is on the rise (see Terwiesch et al. 2019). However, mostly, our goal is to establish causality: We want to put the rain men out of business. In operations management, we pick a managerial practice as explanatory variable X (e.g., concurrent engineering and patients emailing physicians), and our aim is to show whether this practice is effective. Articles become interesting when it previously was believed that Y cannot be influenced and we show that X makes a difference or when it previously was believed that X works and we show that it does not.

Is estimating β enough? Though interesting and important, such β estimates do not tell us if we should bring an umbrella. To tackle such problems, we have to develop a good prediction model of rain and then use that to make a recommendation when the chance of rain is high enough to merit bringing an umbrella along (I very much enjoyed a talk by Jake Hofman of Microsoft Research, who distinguished between the prediction and the causal effect paradigms; see Hofman et al. 2017). Similarly, my own empirical research neither established the optimal level of parallel product development work nor did it establish a better interface for patients contacting their doctors.

Following the logic of Kleinberg et al. (2015), I propose that most of the empirical work in our field has focused on building and refining theories and somewhat neglected prediction problems. To see the importance of prediction and how it differs from explanation, consider the 2018 MSOM Data Driven Research Challenge in partnership with Cainiao (Alibaba's logistics arm), a great initiative pushing our field toward being more empirical. The challenge is organized by an MSOM committee chaired by Gad Allon and Chris Tang and a group of Cainiao executives. It is open to all MSOM members, who, upon registration will be granted access to this otherwise proprietary data set. More details can be found on the MSOM website (http://connect.informs.org/msom/home). Many of the questions posed to the researchers are about establishing an effect and thus about estimating a β (e.g., "How does the current reputation system impact consumer purchasing behavior?"). Others, however, are more about making accurate predictions ("Can Cainiao's platform be used to develop more accurate demand forecasts?"). Coming up with better prediction methods, in my view, requires more of an engineering approach to science than following the social science approach. In this paradigm, the value of a paper is not determined by how interesting its theoretical findings are but by whatever works to make the best prediction.

5.3. Replication of Empirical Studies Is What Moves Science

My second concern is related to Davis' desire for denying what was previously known. Imagine that our previously discussed operational practice X is beneficial in driving some outcome measure. For sake of argument, assume it works in 99% of the applications. Which empirical study is more likely to get published in the Davis paradigm, the one that confirms that X is pretty effective on average or the one highlighting the 1% of the cases when it is not?

If we reward authors for creating surprising results, most of the published articles highlight the 1% in which the practice fails. It is simply more interesting. In a world in which thousands of papers show the presence of man-made global warming, it is the one paper that does not find the effect that draws attention. The audience has finite absorptive capacity and concludes that "opinions are divided" and "X sometimes works and sometimes doesn't." If that is the case, we as researchers have done more harm than good.

When data become available at zero marginal costs and researchers are rewarded for surprising results, we are tempted to start fishing for them. One way of achieving this is by postponing the definition of a concise research question. Initially, we wanted to find the effect of X on Y controlling for A. However, because this model does not lead to a (surprising) statistically significant result, we instead control for B and stratify the sample by C. Data exists in abundance, so we always can come up with additional covariates and, 20 model specifications later, voilà, our result is significant at the 5% level. This practice is, first and foremost, violating good scientific conduct (Simonsohn and Simmons 2017), but it also gets reinforced through a publication process that looks for surprising results.

One way of addressing this concern is by replicating empirical studies. The reason why scientists and (serious) politicians are active in the fight to stop global warming is that its presence has been established not by one study but by hundreds. In most fields of science, progress is made through many small steps of replication, refinement, and extension not by creating headlines based on results that overthrow our current understanding of the world (e.g., "Global warming might not be man-made after all" or "Practice X will not work"). Such replication papers create a robust foundation of understanding the world, though each of these papers individually might be at risk for being rejected, not cited, or not recognized because they are not yielding surprising results.

Again, a personal example might help illustrate the point: Diwas KC and I recently ran a study quantifying the impact of surgical smoothing on hospital access and patient flow. A purely conceptual paper with no data or formal tests had argued that smoothing the surgical schedule would be a good way to improve patient flow. Diwas and I were the first to show this with a remarkable data set. We tested the arguably obvious hypothesis and measured the strength of the effect (i.e., is a 10% improvement in smoothing worth the equivalent of one extra bed or a 100 beds?). The reviewer response was that our empirical finding would be obvious and not surprising.

Good theory is commonly obvious, especially with the luxury of hindsight. I would argue that here lies an important difference between the playful economic models that have dominated our field over the last decades and a more empirical approach to operations management. Empirical work involving primary data collection, in contrast, is novel by its very nature (see Section 5.5), and a great empirical article, in my view, neither needs surprises nor 10 pages of hypotheses development.

5.4. Contextual Immersion Beats Armchair Research

As our field has moved to relying primarily on mathematical models, it has suffered from a growing divide between those coming up with new formal models and the actual work that generated the data. As an example of this, let me recall and confess how Christoph Loch and I designed a stochastic model of coordination challenges in a product development team—all from the comfort of our Fontainebleau offices. During later parts of my dissertation, Christoph and other committee members, Arnoud De Meyer and Takahiro Fujimoto, encouraged me to go out into the field and experience automotive product development firsthand. Christoph persuaded executives at BMW to let me observe their teams, and I spent three months onsite studying how engineers shared information with each other. Those three months provided me with a convenience sample of five components.

When I expressed my reluctance to spend more than two weeks for each data point, Takahiro explained that in his dissertation, he had spent three years traveling to automotive companies around the world only to get a total of 30 data points. To some, such painful investment in collecting primary data might be perceived as a waste of time. After all, a quick CompuStat query would have allowed us to empirically analyze the effect of R&D spending on profits for thousands of companies. With the luxury of the 20-year hindsight I now have on my dissertation, however, I am convinced that no PhD course, no book or article, and no conference better prepared me for my research career as those three months in the field. Such immersion prepares the researcher for asking the right research questions while also helping to build face validity of the work.

Earlier this year, my oldest son, an economics major focusing on econometrics, had the opportunity to interact with Harvard economist Roland Fryer. Fryer shared the story of his career and described how his research on the use of force by police officers was guided by him spending countless hours shadowing officers on the job. This allowed him to more fully understand the context in which officers operated and the limitations of prior work. It resulted in the collection of new primary data, a much better interpretation of existing data, and a more convincing and empathetic communication of the research findings. If economists can now go out into the field and observe how people work, we should not stay chained to our office chairs!

Being in the field allows us as researchers to understand the data generation process and to learn about the true management challenges inherent in the work. Observing the process firsthand also allows for a better understanding of how generalizable a particular operational setting is. When you are testing a theory (looking for a β), outliers in your sample are your enemy as they most likely will hurt your p values. However, if you are willing to get into the field to find out what explains the outliers, you likely will be in a better position to know when your theory works and when it does not. In fact, you might even find a new, surprising twist of what has long been thought.

5.5. The Beauty of Raw Data

Starting in the 1980s, researchers at Massachusetts Institute of Technology and Harvard came together to form the International Motor Vehicle Program (IMVP). They studied different operational practices of automakers around the world and convinced scholars and practitioners alike that quality management systems could boost productivity, that inventory turns could be greatly improved by following just in time principles, that product development worked best in cross-functional teams, and that suppliers should be managed as partners rather than as enemies. Academic researchers previously could have long arguments about how Kanban systems work, how many Kanban cards to have, and how variability affected replenishments. However, when the IMVP reports showed that the General Motors Framingham plant held two weeks of inventory while the Toyota Takaoka plant could run with two hours of inventory (Womack et al. 1991), practitioners and academics alike were surprised. Sometimes, you do not even need to have a hypothesis to generate a surprise. The empirical methodology of these studies resembled similar data illustrations seen in *New York Times* articles and in *Econometrica*. Tables of raw data, comparisons of subsample means, and simple ordinary least squares regression were all it took to reshape our understanding of the automotive industry and beyond. The power of a great empirical study lies in the data, not in its methodology.

As another example of the beauty of raw data, consider Anita Tucker's research shadowing nurses. She showed that nurses worked around the problems they encountered while on the job, as opposed to fixing the root cause, in 93% of the cases. Tucker quoted one of the nurses as saying, "Working around problems is just part of my job. Being able to get IV bags or whatever else I need, it enables me to do my job and to have a positive impact on a person's life" (Tucker 2004, p. 151).

Just as good, fresh ingredients make for a great meal without much cooking, there exists true beauty in good data. Empirical studies in which the estimation is just one opaque black box, however, pose the same risks as synthetic foods full of artificial flavors and chemicals. Interesting questions with good primary data should be a recipe for success, even with simple econometric methodology. Researchers should be proud of their raw data and work on communicating it to the reader rather than hiding behind complicated statistical models.

6. Conclusions

What will be the impact of this new type of data on our field? Looking at another field might be instructive to assess how revolutionary this change might be. In 1983, Peter Guadagni and John Little published a paper in *Marketing Science*. Their data set was the first to use scanner data in supermarkets (Guadagni and Little 1983). No longer did researchers have to rely on a small sample of customers subjectively keeping a diary about their purchases. Instead, they had access to microlevel data. In the words of the authors, "Compared to diary panels, scanner measurement is relatively unobtrusive, bias-free, and complete across products" (p. 205). Scanner data allowed empirical research in marketing to catch up with the advances in the theory of choice models, leading to numerous future research topics.

In my view, this is our scanner data moment. Because of new data collection technologies (digital exhaust), empirical research in our field now has the luxury to investigate operations at the microlevel and do so almost continuously. Just like in marketing, our empirical models can finally catch up to our theory.

Operations means work, and we as scholars of operations management study how people work. In a world of globalization, automation, and scarcity of resources, I cannot think of a more fascinating area of academic inquiry. If we take advantage of this new type of data and are willing to immerse ourselves into the context of where the actual work is done, empirical research is not just the past of our field but also its future.

Acknowledgments

The author is grateful to Chris Tang for the invitation of authoring this article. The author acknowledges the helpful comments by Jan Van Mieghem, Gerard Cachon, Marshall Fisher, Toni Moreno, Diwas KC, Robert Batt, Christophe Van den Bulte, Hessam Bavafa, Senthil Veeraraghavan, and Brad Staats on earlier versions of this document. Jake Hofman pointed the author toward articles on the importance of predictions, for which he is very thankful, as well. Finally, two anonymous reviewers helped with a number of very constructive comments.

References

Batt R, Terwiesch C (2015) Transparency in waiting lines: An empirical analysis of patient abandonment behavior in the emergency department. *Management Sci.* 61(1):39–59.

Bavafa H, Hitt L, Terwiesch C (2018) The impact of e-visits on visit frequencies and patient health: Evidence from primary care. *Management Sci.* 64(12):5461–5480.

Cachon G (2012) What is interesting in operations management? *Manufacturing Service Oper. Management* 14(2):166–169.

Davis M (1971) That's interesting. *Philos. Soc. Sci.* 1:309–344.

Fisher, M (2007) Strengthening the empirical base of operations management. *Manufacturing Service Oper. Management* 7(4):368–382.

Guadagni P, Little J (1983) A logit model of brand choice calibrated on scanner data. *Marketing Sci.* 2(3):203–238.

Hofman J, Sharma A, Watts D (2017) Prediction and explanation in social systems. *Science* 355(6324):486–488.

KC D, Terwiesch C (2009) Impact of workload on productivity: An econometric study of hospital operations. *Management Sci.* 55(9): 1486–1498.

KC D, Terwiesch C (2011) The effects of focus in operational performance: Evidence from California hospitals. *Management Sci.* 57(11):1897–1912.

Kleinberg J, Ludwig J, Mullainathan S, Obermeyer Z (2015) Prediction policy problems. *Am. Econ. Rev.* 105(5):491–495.

Meng L, Batt R, Terwiesch C (2018) The impact of facility layout on the provision of care: A study of nurse movements in the emergency department. Working paper, Wharton School of Business, Philadelphia.

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.

Neef D (2014) *Digital Exhaust: What Everyone Should Know About Big Data, Digitization and Digitally Driven Innovation* (FT Press Analytics, Upper Saddle River, NJ).

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

Simmons J, Simonsohn U (2017) Power posing: P-curve evidence. *Psych. Sci.* 28(5):687–693.

Staats BR, Dai H, Hofmann D, Milkman KL (2017) Individual electronic monitoring: An empirical examination of hand hygiene in healthcare. *Management Sci.* 63(5):1563–1585.

Taylor FW (1919) *The Principles of Scientific Managemen*t (Harper, New York).

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

Terwiesch C, Loch CH (1999) Measuring the effectiveness of overlapping development activities. *Management Sci*. 45(4): 455–465.

Terwiesch C, Ren JZ, Ho TH, Cohen M (2005) An empirical analysis of forecast sharing in the semiconductor equipment supply chain. *Management Sci*. 51(2):208–220.

Terwiesch C, Olivares M, Staats BR, Gaur V (2019) A review of empirical operations management over the last two decades. *Manufacturing Service Oper. Management*. Forthcoming.

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

Webb E, Campbell D, Schwartz R, Sechrest L (1999) *Unobtrusive Measures*, revised edition (Sage Publications, Thousand Oaks, CA).

Womack J, Jones DT, Roos D (1991) *The Machine that Changed the World: The Story of Lean Production* (Free Press, New York).