


OM Forum

Causal Inference Models in Operations Management

Teck-Hua Ho,^a Noah Lim,^{a,b} Sadat Reza,^c Xiaoyu Xia^d

^aNational University of Singapore, Singapore 119077; ^bUniversity of Wisconsin—Madison, Madison, Wisconsin 53706;

^cInstitute on Asian Consumer Insight, Nanyang Technological University, Singapore 639798; ^dDepartment of Decision Sciences and Managerial Economics, CUHK Business School, Chinese University of Hong Kong, Shatin, N.T., Hong Kong

Contact: dprhoth@nus.edu.sg,  <http://orcid.org/0000-0001-5210-4977> (T-HH); nlim@bus.wisc.edu (NL); sreza@ntu.edu.sg (SR); xiaoyuxia@baf.cuhk.edu.hk. (XX)

Received: March 20, 2017

Revised: June 4, 2017; June 20, 2017

Accepted: June 22, 2017

Published Online in Articles in Advance:
October 5, 2017

<https://doi.org/10.1287/msom.2017.0659>

Copyright: © 2017 INFORMS

Abstract. Operations management (OM) researchers have traditionally focused on developing normative mathematical models that prescribe what managers and firms should do. Recently, there has been increased interest in understanding what managers and firms actually do and the factors that drive these decisions. To advance this understanding, empirical investigation using causal inference models is critical. However, in many contexts, the ability to obtain causal inferences is fraught with the challenges of endogeneity and selection bias. This paper describes five empirical tools that have been widely used in economics to address these challenges and how they can be adopted by OM researchers. We also present an example that illustrates how the various attributes of big data—variety, velocity, and volume—can be useful in addressing the endogeneity bias.

Funding: Teck-Hua Ho's research was supported by the National Research Foundation, Singapore [Grant NRF-RSS2014-001] to the National University of Singapore.

Supplemental Material: The online appendix is available at <https://doi.org/10.1287/msom.2017.0659>.

Keywords: empirical research • econometric analysis • causal inference • big data

1. Introduction

Operations management (OM) is an applied field. Historically, OM researchers have primarily focused on developing normative mathematical models that prescribe what managers and firms *should* do, and placed relatively little emphasis on understanding what they *actually* do and why they do it. Recently, however, there has been increased interest in the latter. The understanding of the causal impact of managerial actions is often obtained through empirical analysis using observational data. As is well known, such investigation is fraught with challenges.

Two such challenges that commonly arise in OM contexts are endogeneity and self-selection. It is difficult to find empirical contexts where neither of these challenges is present. Hence, researchers need to explicitly address the potential manifestations of these challenges in their research contexts and apply appropriate tools that allow them to mitigate any associated problems.

In this paper, we discuss five empirical modeling tools that have been widely applied in economics to overcome the challenges posed by endogeneity and self-selection. These are (1) the instrumental variable (IV) estimator in cases where endogeneity is due to omitted variables, (2) the instrumental variable estimator with exclusion restrictions in cases where endogeneity is due to simultaneity, (3) the propensity score matching (PSM) estimator in cases where the selection mechanism is explained by observables, (4) the regression discontinuity design (RDD) in cases where

the selection mechanism is not explained by observables, and (5) the difference-in-differences (DID) estimator in cases where the selection mechanism is not explained by observables but researchers have access to data that span multiple periods.

A review of the empirical literature in OM reveals two interesting observations. First, while the number of empirical papers published has increased over time, empirical OM researchers still remain a minority. Second, explicit discussion of the previously mentioned challenges and applications of the tools is not widespread. Consider the three well-established journals in the field: *Management Science*, *Manufacturing & Service Operations Management*, and *Production and Operations Management*. From 2010 to 2015, 1,015 articles were published on OM topics. Noticeably, only 18% of the articles contained empirical analysis. Among these articles, 75% were on causal inference using observational data, with the rest being on either predictive modeling or inference using experimental data. Only 37% of the papers using observational data explicitly addressed endogeneity and/or self-selection biases. These statistics, reported in Table 1, show that there is much scope for wider application of the previously mentioned tools in empirical OM.

The availability of tools needs to be complemented with good quality data for robust empirical investigation. One could argue that the historical lack of empirical research in OM could be attributed to the

Table 1. Empirical Papers in Top OM Journals: 2010–2015

Journal	# of OM articles	# with empirical focus	# using observational data for causal inference	# addressing endogeneity and selection bias
MS	214	53	38	20
M&SOM	245	58	34	13
POM	556	76	69	19

Note. MS, Management Science; M&SOM, Manufacturing & Service Operations Management; POM, Production and Operations Management.

difficulty of obtaining data. Recently, wide-scale digitization and the availability of big data have rapidly changed the information landscape. However, big data analysis is not immune to the challenges of endogeneity or self-selection. Researchers still need to apply the appropriate econometric tools to draw robust causal inferences. Nonetheless, big data has three characteristics—variety, velocity, and volume—that provide the following advantages: (1) variety can increase the sources of instrumental variables, a key ingredient in addressing the previous challenges; (2) velocity, or the high frequency nature of big data, opens up new avenues for empirical research by allowing investigation of problems that would otherwise not be feasible; and (3) volume can enhance the quality of estimates by improving their precision.

The objective of this paper is to provide a concise introduction to the endogeneity and selection bias issues in empirical modeling for causal inference using nonexperimental observational data.¹ In particular, we focus on linear models and present several examples on how to address these issues in such models.² We hope that this paper will help OM researchers, particularly those who are less familiar with but interested in empirical modeling, gain a workable knowledge of a few useful tools for causal inference modeling.

We begin our discussion in Section 2 by laying out the requirements for consistent estimation using a linear regression model. In Section 3, we discuss the endogeneity bias and the tools to address it. In Section 4, we elaborate on the selection bias and the tools to address it. In Section 5, we present an example that illustrates how the various characteristics of big data can be useful in addressing the endogeneity bias. In Section 6, we conclude with a discussion of the research implications for OM.

2. Linear Regression—The Workhorse Model For Causal Inference

Causal inference models using observational data can be broadly categorized into two strands. The first strand of models imposes analytical structures on the data and is known as structural econometric modeling. These analytical structures are typically based on the optimality conditions derived from profit or utility

maximization models.³ Our paper focuses on the second and more widely used strand of causal inference models, known as reduced-form models. The reduced-form approach is useful when the analytical structure underlying the data is either not fully established or cannot be taken directly to the available data. Instead, one imposes a statistical relationship between an outcome variable and a set of explanatory variables that may potentially determine the outcome variable. The most well-known tool for reduced-form estimation is the linear regression model. In this paper, we focus primarily on addressing endogeneity and selection bias issues for this model.

Applications using linear regression models are ubiquitous in the empirical OM literature. Suppose the researcher observes the data (Y_i, \mathbf{X}_i) on $i = 1, \dots, N$ individuals or firms, where Y_i denotes some outcome variable of interest, and \mathbf{X}_i denotes an M -dimensional set of explanatory variables. The linear model assumes the following causal relationship:

$$Y_i = \beta' \mathbf{X}_i + \epsilon_i, \quad (1)$$

where ϵ_i is a zero-mean unobserved error term, and $\beta = (\beta_1 \beta_2 \dots \beta_M)'$ is an M -dimensional parameter vector. Thus, the conditional expectation $E[Y_i | \mathbf{X}_i]$ is $\beta' \mathbf{X}_i$. The objective is to estimate each parameter $\beta_m \in R$, which captures the causal effect of the component X_m of \mathbf{X} on Y , since $\beta_m = \partial E[Y_i | \mathbf{X}_i] / \partial X_{mi}$. Let \mathbf{Y} and $\boldsymbol{\epsilon}$ denote the N -dimensional vectors of the outcome variable and the unobserved variables, respectively, for the N individuals in the sample. The ordinary least squares (OLS) estimator is given by $\hat{\beta} = (\mathbf{X}'\mathbf{X})^{-1}(\mathbf{X}'\mathbf{Y})$. It can be shown that

$$\hat{\beta} = \beta + \left(\frac{\mathbf{X}'\mathbf{X}}{N} \right)^{-1} \frac{\mathbf{X}'\boldsymbol{\epsilon}}{N}. \quad (2)$$

One of the conditions necessary for the OLS estimator to be consistent (i.e., $\text{plim} \hat{\beta} = \beta$) is $\text{plim} \mathbf{X}'\boldsymbol{\epsilon}/N = 0$. In other words, \mathbf{X} should not be correlated with $\boldsymbol{\epsilon}$. When the previous condition is not met, causal inference based on the linear model is not valid. This is the fundamental concern in modeling causal relationships. In particular, we draw attention to endogeneity and self-selection problems, which are quite pervasive in OM contexts. If these problems are not properly addressed, then the validity of the empirical results would be questionable.

3. Endogeneity Bias

Endogeneity bias arises when the explanatory variables and the errors are correlated. Specifically, if any component X_m of \mathbf{X} is correlated with the error term, then it follows from (2) that $E[X_{mi}\epsilon_i] \neq 0$. Consequently, $\hat{\beta}_m$ does not converge to the true parameter β_m , even if the number of observations $N \rightarrow \infty$. This leads to the OLS estimator being inconsistent, and thus the correct correlation between the dependent and explanatory variables cannot be established. Two important reasons for the endogeneity bias are (1) omitted variables and (2) simultaneity.⁴ In the following sections we describe these issues and also discuss the econometric tools that can generate consistent estimates in the presence of these problems.

3.1. Omitted Variable Bias

Omitted variable bias occurs when one ignores the possibility of some unobserved variable affecting both the outcome and key explanatory variables in the estimation models. We demonstrate this using a service management example where the research question is how staffing levels affect store sales. Perdikaki et al. (2012) and Mani et al. (2015) investigated this question. Suppose the data we observe consist of average daily sales (in thousands of dollars) and average daily staffing levels (number of staff) for $i = 1, \dots, N$ stores located in various markets for a particular week. Let Y_i and X_i denote store sales and staffing level, respectively, for store i .

A standard regression model to estimate the effect of the staffing levels on store sales would be as follows:

$$Y_i = \alpha_0 + \alpha_1 X_i + \mu_i, \quad (3)$$

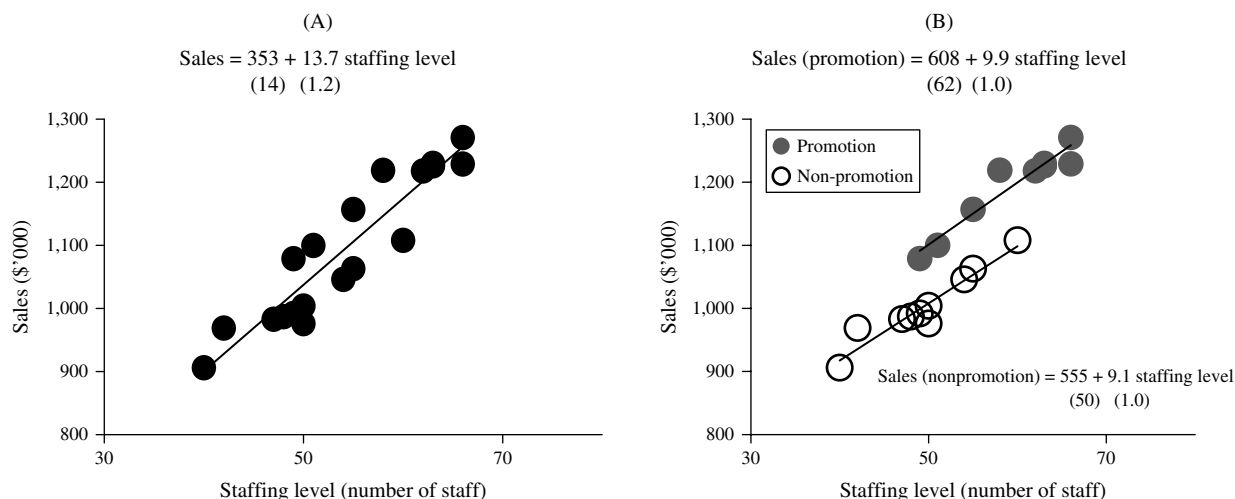
where μ_i is a random error term. However, if there exists some variable Z_i , such as store promotions conducted

during the observation week, that is correlated with both store sales and the staffing levels, then a more accurate model for the data generating process is possibly given by $Y_i = \beta_0 + \beta_1 X_i + \beta_2 Z_i + \epsilon_i$, $\text{Cov}(X_i, Z_i) \neq 0$. In this situation, if we regress Y_i only on X_i , as in (3), then the estimated coefficient of X_i , $\hat{\alpha}_1$, will not be a consistent estimate of β_1 , the true causal effect. Since Z_i is unobserved and omitted from the regression model, X_i and μ_i are correlated in (3), which makes X_i endogenous.

We illustrate the extent of the omitted variable bias using simulated data generated based on the previous example with $\beta_1 = 10$.⁵ In Figure 1(A), we draw a scatter plot of \mathbf{Y} and \mathbf{X} for a few observations. We do not distinguish between promotional and nonpromotional sales, since these data are unobserved. As shown in the graph, the estimated coefficient of the staffing level based on regression model (3) is 13.7 (standard error (s.e.), 1.2), which is significantly larger than the true parameter value. However, if we also observe which stores offered promotions (Z_i), we can group the observations on the basis of whether the store offered a promotion and reestimate regression model (3) for these two groups. We draw the predicted lines based on the two regressions in Figure 1(B). The estimates of the coefficient of X_i for both the regressions are not significantly different from 10, the true parameter value.

3.1.1. Solution. One way to correct for the omitted variable bias is to introduce an instrumental variable. Intuitively, the IV filters out the influence of the omitted variable in a regression model. Let V_i be such an IV. Then two conditions must be satisfied: (1) V_i must be correlated with X_i , i.e., $\text{Cov}(V_i, X_i) \neq 0$, and (2) V_i must not be correlated with the omitted variable Z_i , i.e., $\text{Cov}(V_i, Z_i) = 0$. These conditions allow us to generate

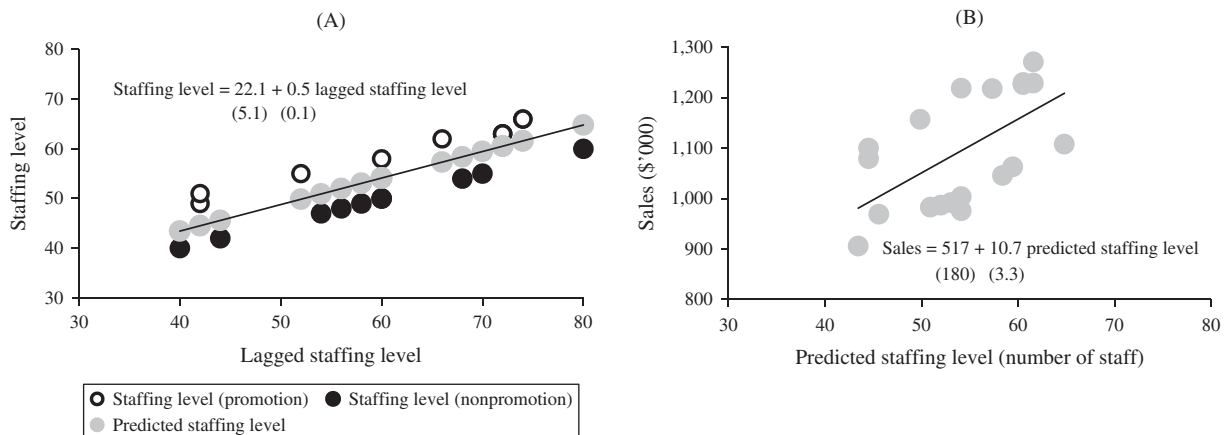
Figure 1. Illustrating the Omitted Variable Bias



Notes. Panel (A) shows the regression when the promotion variable is not available. Panel (B) shows the regression when the promotion variable is available. Standard errors are shown in parentheses.

M&SOM 2017.19:509-525.

Figure 2. IV-2SLS Estimation



Notes. Panel (A) shows Step 1 of the IV-2SLS estimation: regression of the endogenous variable on instrumental variable. Panel (B) shows Step 2 of the IV-2SLS estimation: regression of the dependent variable on predicted values generated in Step 1. Standard errors are shown in parentheses.

an estimate of X_i , \hat{X}_i , that contains the same information as in X_i but is free from the influence of Z_i .

The correction for omitted variable bias in a linear regression framework is usually conducted using a two-stage least squares (2SLS) approach.⁶ In the first step, we regress the endogenous explanatory variable X_i on the instrumental variable V_i , and other exogenous variables in the model (see Figure 2(A)). Denote the predicted values by \hat{X}_i . In the second step, we regress the outcome variable Y_i on \hat{X}_i and the other exogenous variables.

In our previous example, the staffing level is endogenous since we do not observe which store offered the promotion. Finding an appropriate set of IVs is usually a challenge. Both Perdikaki et al. (2012) and Mani et al. (2015) used the lagged staffing level as an instrument. This is a plausible IV, as stores are likely to maintain some continuity in the staffing levels over consecutive periods. In our example, we do not have other exogenous variables; therefore, it is sufficient to regress only on the lagged staffing level. Notice that the effect of the lagged staffing level on the current staffing level is significant with a coefficient estimate of 0.5 (s.e., 0.1). In Figure 2(B), we show the regression line for Step 2 of the 2SLS approach, where store sales is regressed on the predicted staffing level from Step 1. The estimated slope coefficient from Step 2 turns out to be 10.7 (s.e., 3.3), which is statistically not different from the true value of β_1 , which equals 10. In addition to correcting for the omitted variable bias, researchers may be interested in the direction of the bias. It follows from (2) that in a single equation regression model, the direction of bias will depend on the covariance between the endogenous variable and the omitted variables. If the covariance is positive, then the OLS regression will overestimate the effect, and vice versa. In our simulated example, the explanatory variable (staffing level)

is positively correlated with the omitted variable (store promotions). As a result, the OLS estimate turns out to be upward biased.

It is important to note that the IV should not be correlated with the omitted variable. While it is testable whether the IV is correlated with the endogenous explanatory variable, there are no statistical tests that can demonstrate that it is uncorrelated with the omitted variable. This is because the omitted variables are not observed. However, there are statistical tests (e.g., Anderson's Lagrange multiplier (LM) test, Cragg–Donald F -test) for checking whether the IV is sufficiently correlated with the explanatory variables. Standard software packages report these test statistics routinely with IV-regression outputs.

3.2. Simultaneity Bias

The second cause of endogeneity bias is simultaneity. This occurs when there are multiple dependent variables each influencing at least one other dependent variable. An example of the simultaneity problem in empirical OM research is found in Kesavan et al. (2010), in which the authors argue that the cost of goods sold, gross margin, and inventory levels are simultaneously determined.

To demonstrate the problems associated with simultaneity, we use an example of the causal relationship between airline ticket sales and the price of airline tickets. The price of airline tickets is an important determinant of the demand for air travel. However, airline companies adjust ticket prices based on expected demand. Therefore, airline ticket sales are a function of price from the demand perspective, but the price of airline tickets is also a function of expected airline ticket sales from the supply perspective. Suppose that we have data on the number of tickets sold for a particular route originating in a particular city for various quarters,

along with the average prices of those tickets sold. We can use the following system of equations to represent the demand and supply functions:

$$\begin{aligned} \text{Demand: } Q_t &= \beta_{10} + \beta_{11}P_t + \beta_{12}X_{1t} + \epsilon_{1t}, \\ \text{Supply: } P_t &= \beta_{20} + \beta_{21}Q_t + \beta_{22}X_{2t} + \epsilon_{2t}, \end{aligned} \quad (4)$$

where Q_t and P_t are the log-transformed values of quantity of tickets sold and average price per ticket for periods $t = 1, \dots, T$. We use X_{1t} and X_{2t} to denote factors that affect customer demand for tickets and airline companies' pricing decisions, respectively, and ϵ_{1t} and ϵ_{2t} are random shocks. Consider the demand function. Regressing Q_t on P_t and X_{1t} will generate a biased estimate of the demand elasticity β_{11} . This is because P_t is also determined by Q_t in the supply function. Consequently, we cannot claim that $\text{Cov}(P_t, \epsilon_{1t}) = 0$. In other words, P_t is endogenous in the demand equation, and the estimate of β_{11} using such a regression model would be subject to the simultaneity bias.

We illustrate this problem using simulated data in Figure 3 using $\beta_{10} = 10$, $\beta_{11} = -0.6$, $\beta_{12} = 0.2$ and $\beta_{20} = 1$, $\beta_{21} = 0.5$, $\beta_{22} = 0.3$. In equilibrium, the demand for air travel equals the supply. We generated the data using this equilibrium condition and the previously mentioned parameter values. The terms X_1 and X_2 and the errors were drawn from independent standard normal distributions. A direct regression of quantity on price yields a slope estimate of -0.1 (s.e., 0.2), which is not significant and reflects neither the elasticity of demand ($dQ_t/dP_t = -0.6$) nor the elasticity of supply ($dQ_t/dP_t = 2$).

3.2.1. Solution. To correct for the simultaneity bias, we need one IV for every endogenous variable. Since there are two endogenous variables, price and quantity, in our example, we need two IVs. Importantly, for each endogenous variable there must be at least one

IV that does not directly affect any other endogenous variable. This is known as the exclusion restriction. In our case, we assume that per-capita income affects only the quantity demanded, and fuel costs affect only the pricing strategy of airlines. Therefore, per-capita income and fuel costs satisfy the exclusion restrictions and can be used as IVs. We can use these IVs to estimate both the demand and supply function parameters, using two separate sets of regressions, via the 2SLS method. Consider estimating the demand function. In the first step, the endogenous variable (ticket prices) is regressed on all the instrumental variables (crude oil prices and per capita income) to generate the predicted ticket prices.⁷ In the second step, we regress the quantity of tickets sold on these predicted ticket prices. The estimated coefficient is -0.6 (s.e., 0.2), which is statistically no different from the true parameter value of -0.6 .

In the discussion on omitted the variable bias correction, we mentioned how the direction of bias is dependent on the covariance between the explanatory variable and the omitted variable. However, in the presence of simultaneity, the direction of bias is not as easy to establish. In such cases, the direction of the bias will depend on both the signs and magnitudes of the actual parameters, which are unknown a priori. Thus, in our example, the direction of bias will depend on the signs and magnitudes of β_{11} and β_{21} .

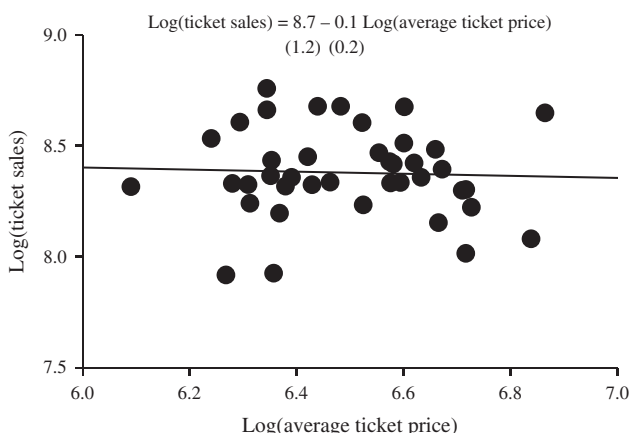
4. Selection Bias

Many OM questions involve investigating the effects of adopting a program or policy on one or more dependent variables. Researchers have investigated the effects of adopting programs such as total quality management or International Organization for Standardization (ISO) certification on productivity and other aspects of firm performance (Levine and Toffel 2010, Gray et al. 2015). When participants self-select into different programs, versus when they are assigned randomly, standard regression models do not adequately estimate the effect of the program.

Let us consider the example of studying the effects of firms obtaining ISO 9000 certification. In this case, researchers may be inclined to introduce an indicator variable W_i , which takes the value of 1 if firm i is exposed to a treatment (i.e., adopted ISO 9000 standards in this context) and 0 otherwise, in the regression model $Y_i = \beta_1 W_i + \beta_2 X_i + \epsilon_i$. However, such straightforward inclusion of the treatment indicator in the regression model will generally not lead to a consistent estimate of the actual impact of ensuring ISO 9000 standards on firm productivity, because of self-selection by firms into the programs.

Why is self-selection an issue for generating a consistent estimate of the causal effect? For any firm i , we

Figure 3. Illustration of the Simultaneity Bias



Notes. The figure is based on simulated data. Standard errors are shown in parentheses.

M&SOM 2017.19:509-525.

observe the outcome conditional on the adoption decision. Letting $Y_i(W_i)$ denote the outcome conditional on the treatment $W = \{0, 1\}$, we observe $Y_i = Y_i(1)\mathbf{1}[W_i = 1] + Y_i(0)\mathbf{1}[W_i = 0]$. In the program evaluation literature, this equation is known as the Rubin causal model. If a firm adopts the treatment, then the actual observed outcome is $Y_i(1)$, but we do not observe the potential outcome in the nonadoption condition, $Y_i(0)$. Similarly, if a firm does not adopt the treatment, then the actual observed outcome is $Y_i(0)$, but we do not observe the potential outcome in the treatment condition, $Y_i(1)$. Since $Y_i(1)$ and $Y_i(0)$ are never observed simultaneously, the causal effect of the treatment $E[Y_i(1) - Y_i(0)]$ cannot be directly estimated using the observed data. If a firm self-selects into the treatment group based on another factor that can also influence the outcome, such as the quality of its management, then we cannot be sure whether the estimated effect is due to the treatment itself or due to the firm's management quality.

We illustrate the self-selection bias using an example of the decision to obtain ISO 9000 certification. Suppose we are interested in estimating the causal effect of obtaining the certification on firm output. If it is possible to estimate the outputs of firms having obtained and not obtained the certificate, that is, if we were to observe both the actual outcomes and potential outcomes, then it is fairly straightforward to estimate the causal effect. In Table 2, we provide an example with two firms, where we observe the actual and potential outcomes for both firms. Firm B is ISO certified, and its actual output is 800, and the potential output without certification is estimated to be 700. Therefore, the treatment effect for firm B is 100. Firm A is noncertified, and its actual output is 400, and the potential output with certification is estimated to be 500. Therefore, the treatment effect for firm A is also 100. However, the potential outcomes are not observable. What we observe are only the actual outcomes, also shown in Table 2. In this example, if we assume that the two firms are similar, then we will incorrectly infer that the treatment effect

is $800 - 400 = 400$, which is the difference between the outputs of firms A and B. Without accounting for selection, a regression model will generate biased estimates.

To identify the causal impact of a treatment, the estimation depends crucially on the assumptions behind the assignment mechanism. If assignment in the treatment group can be assumed to be dependent only on observable covariates, we can use models applicable for selection on observables. If there exists some unobservable covariate that influences the outcome and treatment assignment rule, then we can use models that are able to address selection on unobservables. In the following section we discuss a few popular models for these two types of selection problems.

4.1. Selection on Observables

Consider the same question of estimating the causal effect of obtaining ISO 9000 certification on firm output. Suppose that we have data on both certified and noncertified firms. Firms that have obtained the certification are in the treatment group, and firms that have not obtained the certification are in the control group. ISO certification is granted on the basis of a firm's performance on eight criteria: (1) customer focus, (2) leadership, (3) people involvement, (4) quality management process, (5) management system, (6) continual improvement, (7) approach to decision making, and (8) supplier relationships. If we observe data on all of the eight criteria, then in the estimation model we can control for these characteristics that determine the assignment rule (to the treatment group). Once these characteristics have been controlled for, the firms' decisions to obtain ISO certification can be considered to have random differences, and in this case selection is no longer a confounding factor. We can then compare the outputs of the firms belonging to the treatment and control groups.

4.1.1. Solution: Matching Model. To estimate the treatment effect given the previous scenario, we can use a matching estimator. For every treated unit, the goal of the matching estimator is to find a comparison unit among the controls that has similar values of observable characteristics X_i . This comparison unit need not be a single unit; rather, it can be a composite (i.e., a weighted average) of several different control units that have similar values of X_i . After computing the average difference between the treated units and the control units, we can use a function of this distance as weights to construct the composite comparison units. More detailed discussion on different matching methods can be found in Todd (2010). Note, however, that when the observable characteristics X_i has many dimensions, it will be difficult to find a comparison unit that is comparable to the treated unit in every dimension, an issue known as the "curse of dimensionality."

Among the matching models, the propensity score matching model has gained significant popularity, as it

Table 2. Estimating Treatment Effect of Obtaining ISO Certification

	Output without certification $Y_i(0)$	Output with certification $Y_i(1)$	Treatment effect $E[Y_i(1) - Y_i(0)]$
<i>Both actual and potential outcomes are known</i>			
Firm A (noncertified)	400	500	100
Firm B (certified)	700	800	100
<i>Only actual outcomes are known</i>			
Firm A (noncertified)	400	—	?
Firm B (certified)	—	800	?

is intuitive and does not suffer from the curse of dimensionality. The essential idea behind the PSM model is that if one can estimate (1) the propensity (or probability) of selecting the treatment conditional on the observed covariates and (2) the expected outcomes with and without treatment for units with similar propensities, then the treatment effect can be estimated by comparing the average outcomes of the treatment group and control group firms with similar propensities. To estimate the treatment effect, we need to group the firms according to their similarity in characteristics. The primary advantage of PSM is that the propensity score is a scalar; thus, it is easy to group the firms on the basis of their propensities to participate in the treatment.

Continuing with the example of estimating the impact of obtaining ISO 9000 certification on firm output, suppose that in addition to certification status and output, we also observe data on all of the eight certification criteria. Data on these characteristics will allow us to calculate each firm’s propensity to obtain ISO 9000 certification. This can be done through a simple probit or logit model. Once these propensities are estimated, we can group firms according to their similarity in propensities. In Table 3, we present the sample data for our illustration. As shown, for each group of firms based on high or low propensities, we can find the average outputs for firms that have and have not obtained certification. Using this procedure, we can see that firms with low propensities can expect a treatment effect of an output increase of 75, while firms with high propensities can expect a treatment effect of 150. This highlights an additional advantage of the PSM method. Not only are we able to match firms using scalar propensities, which can be derived from multiple characteristics, but we are also able to see whether the treatment effects differ across propensities.

Denote the propensity of firm i self-selecting into a treatment by $P_i(\mathbf{X}_i)$, where \mathbf{X}_i is the vector of characteristics that determine assignment to the treatment group. Assume that there are firms with similar values of \mathbf{X} in both the treatment group (ISO certified) and the control group (noncertified). Thus, we should be able to estimate $E[Y_i(1) | P_i(\mathbf{X}_i) = p]$ and $E[Y_i(0) | P_i(\mathbf{X}_i) = p]$, where p is the value of propensity score. One can then proceed to estimate the treatment effect, which is given by $\beta_{PSM} = E[Y_i(1) | P_i(\mathbf{X}_i) = p] - E[Y_i(0) | P_i(\mathbf{X}_i) = p]$.

In the first step of the PSM method, researchers need to estimate the propensities. Typically, a probit or a logit model is used to generate these estimates. In the next step, once the propensities are estimated, we need to estimate the average outcomes of the treatment and control group firms around the various propensity measures. For this purpose, we need to find groups of firms matched according to their propensities. We can use a specific matching algorithm, such as nearest-neighbor matching, caliper and radius matching, classification and interval matching, a kernel weighting function, or regression weighting. To estimate the treatment effect, one can then compare the average difference between the treated units and the composite comparison units that are matched by the chosen algorithm. Alternatively, one can use the classification approach suggested by Rosenbaum and Rubin (1983). This approach first divides the estimated propensities into J groups, and then it estimates the average outcomes for treated and untreated firms within each group. Following the estimation of the treatment effect for each group, we can then find the average treatment effect in the sample.

4.2. Selection on Unobservables

If the assignment to a treatment condition depends on factors that are not observed and if such factors also affect the outcome, then there is selection on

Table 3. Illustration of Propensity Score Matching Model for Estimation of the Treatment Effect of Obtaining ISO Certification

	Obtained certification	Output without certification	Output with certification	Propensity score (%)	Group	Subgroup	Average output within subgroup	PSM estimate of treatment effect by subgroup
Firm 1	Yes	—	800	85	High propensity (propensity score > 50)	Certified	700	150
Firm 2	Yes	—	700	75				
Firm 3	Yes	—	600	70				
Firm 4	Yes	—	700	85		Not certified	550	
Firm 5	No	600	—	80				
Firm 6	No	500	—	70				
Firm 7	Yes	—	500	40	Low propensity (propensity score ≤ 50)	Certified	450	75
Firm 8	Yes	—	400	30				
Firm 9	No	450	—	50				
Firm 10	No	400	—	40		Not certified	375	
Firm 11	No	350	—	35				
Firm 12	No	300	—	30				

Note. The previous values are based on simulated data for illustration purposes.

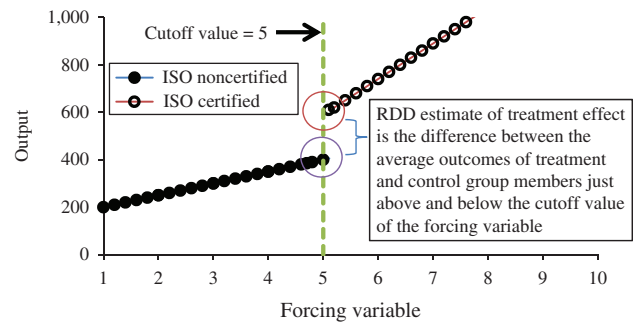
M&SOM 2017.19:509-525.

unobservables. In such situations, a matching estimator such as PSM is not applicable, since assignment to the treatment group is not solely determined by observable factors. There is no general approach for causal inference or estimation of the treatment effect in the presence of selection on unobservables (Imbens and Wooldridge 2009). Researchers have to rely on quasi-experimental designs to identify the causal effect of treatment. Two popular methods for estimating treatment effects are the regression discontinuity design and the difference-in-differences method. For RDD, the researcher observes a threshold or cutoff value of a continuous variable that separates the treatment and control group units. The units around this cutoff could be considered to be similar. Comparison of the outcomes of these units allows estimation of the treatment effect. To apply the DID method, the researcher observes data for at least two periods, one before exposure to the treatment and one after exposure to the treatment, for both the treatment and control group units. Under the assumption of a common time trend, comparison of pre- and posttreatment outcomes for the two groups yields the treatment effect. We discuss the two methods in the following section.

4.2.1. Solution 1: Regression Discontinuity Design. RDD allows for an estimation of the causal effect of a treatment, when the treatment is assigned only above (or below) some cutoff or threshold of an observable variable X_f , known as the forcing variable. We estimate the treatment effect by comparing observations lying closely on either side of the threshold. Such forcing variables exist in many contexts, often based on the application of some administrative criterion.

In the previous example of examining the effect of obtaining ISO 9000 certification on firm output, suppose now that we do not observe all of the eight characteristics. In this case, PSM may not be a useful approach since the unobserved variables may also have an impact on firm output. However, suppose that we are able to obtain data on another variable that is continuous in nature, such that above a certain value of this variable, all firms are in the treatment group. For example, if the certification granting agency generates a score for each firm, on the basis of which the certification will be awarded, then surely there will be a cutoff value above which firms will be in the treatment group. If the researcher has access to these scores, then the score can be used as a forcing variable. Suppose that this score is generated on a scale of 1–10, and that there is a cutoff value of 5, above which the firms are able to receive certification. It is plausible to assume that firms scoring just above 5 and just below 5 are similar in many ways. Therefore, we can compare the average output of firms with scores just above 5 (certified) and those just below 5 (noncertified) to estimate the effect of ISO certification on output. We illustrate this in Figure 4.⁸

Figure 4. (Color online) Regression Discontinuity Design



It is recommended that one start with a visual presentation of the data to evaluate the possibility of implementing RDD. If the plot of the being-treated ratio against any continuous-valued covariate suggests that there is a clear threshold at a particular value of the covariate, then one can consider that covariate to be the forcing variable. It is then advisable to plot the forcing variable and other observable covariates. If there is any indication of discontinuity around the threshold for any other covariate, then the treatment effect is considered to be not identified. One routine procedure for RDD is to plot every covariate against the forcing variable and test the existence of discontinuity.

The advantage of RDD is that it allows for selection on unobservables. Firms may self-select to be in a treatment regime if the realized value of a forcing variable is above a threshold c , i.e., the selection equation is given by, $W_i = 1[X_{fi} \geq c]$, where X_{fi} is the forcing or assignment variable. While various unobservable factors may determine the value X_{fi} , we can assume that around $X_f = c$, firms are similar. Firms with realizations of the forcing variable just above c are in the treatment group, and firms with realizations just below c are in the control group. Since we assume that firms belonging to these two groups are similar, we can estimate the potential outcome in the nontreatment regime using the average outcomes of the firms just below c . This is the essential idea behind RDD. Firms just below the cutoff are considered to be the quasi-control group. The treatment effect is given by $\beta_{RDD} = \lim_{\epsilon \downarrow 0} E[Y_i | X_{fi} = c + \epsilon] - \lim_{\epsilon \uparrow 0} E[Y_i | X_{fi} = c + \epsilon]$.

Once the sample near the cutoff is selected, one can then use the following regression to estimate the treatment effect that is given by β in the model $Y_i = \alpha + \beta W_i + \gamma_1(X_{fi} - c) + \gamma_2 W_i \cdot (X_{fi} - c) + \epsilon_i$. It is not necessary to include other covariates in the regression, even if those covariates are important in the selection criterion. However, including available covariates can help reduce any small-sample bias (see Imbens and Lemieux 2008). Finally, one would also need to consider how to determine the optimal intervals around the cut-off point. We refer readers to Imbens and Kalyanaraman (2012) for more detailed discussion.

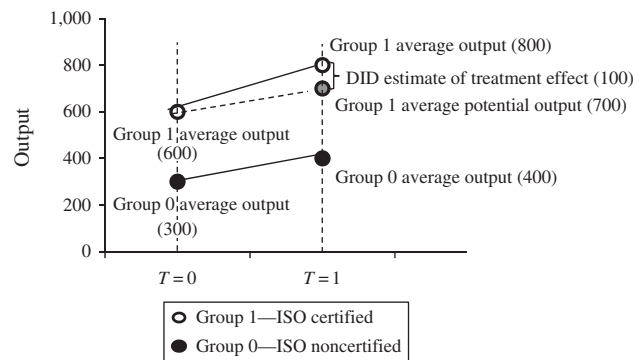
4.2.2. Solution 2: Difference-in-Differences Method.

We discussed the usefulness of the PSM model and RDD for causal inference on the treatment effect using cross-sectional data. If there are observations on pre- and posttreatment periods for both the treatment and control group firms, then the DID method is a very useful technique for estimating the causal effect of the treatment. This method allows for the calculation of the treatment effect by comparing the difference in the change in average outcomes over the two periods for the treatment and control group firms.

We continue with the same example of estimating the treatment effect of obtaining ISO 9000 certification. One needs to address the potential self-selection of high-performing firms into the treatment group (ISO certified) and low-performing firms into the control group (noncertified) group. Suppose that we observe data on these two groups of firms before and after ISO certification was obtained. Let $G_i = \{0, 1\}$ denote the group that firm i belongs to, where $G = 0$ is for the control group and $G = 1$ is for the treatment group. Let $T_i = \{0, 1\}$ denote the observation period for firm i , where $T = 0$ is the precertification period and $T = 1$ is the postcertification period. In this case, we can define $W_{iT_i} \equiv G_i \cdot T_i$, which takes a value of 1 if firm i is in the treatment group and the observation is for the posttreatment period, and 0 otherwise. Given this setting, we can obtain the DID estimate of the treatment effect on outcome Y_i through the coefficient of W_{iT_i} in the regression $Y_{iT_i} = \alpha + \beta_{DID}W_{iT_i} + \gamma G_i + \delta T_i + \epsilon_{iT_i}$. It follows from this equation that $\beta_{DID} = (E[Y | G = 1, T = 1] - E[Y | G = 1, T = 0]) - (E[Y | G = 0, T = 1] - E[Y | G = 0, T = 0])$.

Essentially, β_{DID} captures the difference between two components. The first component is the difference in population average outcomes between the pre- and posttreatment periods for the firms in the treatment group, while the second component is the difference in population average outcomes between the pre- and posttreatment periods for the firms in the control group. We illustrate this in Figure 5. Suppose the average output of firms in group $G = 0$ increased by 100, from 300 in $T = 0$ to 400 in $T = 1$. On the other hand, the average output of firms in group $G = 1$ increased by 200, from 600 in $T = 0$ to 800 in $T = 1$. Inferring that all of this output increase was due to obtaining the certification would clearly be misleading. Instead, we can assume that the output increase in group $G = 1$ would have followed the same time trend as that for firms in group $G = 0$ had the former not obtained the certification. Specifically, if firms in $G = 1$ had not obtained the certification, they would have potentially produced 700 in $T = 1$. The DID estimate of the treatment effect of certification is thus calculated as the difference between the actual and potential outcome for firms in $G = 1$, which in this case is 100. The direct DID estimate of the treatment effect uses the sample averages before

Figure 5. Difference-in-Differences Estimation



and after treatment according to the formula given previously. In practice, researchers usually adopt the regression approach using the aforementioned regression model, as it allows for the incorporation of control variables.

While the DID regression model has been widely used, it also faces two common critiques. First, researchers are often asked to justify the validity of the DID model. Specifically, the most important assumption of the DID model is that of the parallel trend between the control and treatment groups.⁹ To examine the DID model's validity, if the data span multiple periods, one can visually examine the time trends in the treatment and control groups. In addition, one can conduct a placebo test by reestimating the model with randomly chosen treated groups. A valid DID design would generate zero causal effect of any "placebo treatment." The second critique to the DID model is about the estimated standard errors of the treatment effect, and these need to be adjusted for potential serial correlation. Bertrand et al. (2004) discuss how to estimate the correct standard errors using block bootstrap or arbitrary variance-covariance matrix corrections.

5. Does Big Data Alleviate Causal Inference Concerns?

Wide-scale digitization of information is rapidly changing the data landscape. Digital data are now collected and stored at unprecedented levels of variety, velocity, and volume (3Vs), and we are now unquestionably in the "big data" era. While big data has become the buzz word in empirical research, the challenges to causal inference, such as endogeneity and selection-bias, still exist no matter how big the 3Vs of the data are. However, the richness of the data can help researchers better address these challenges, and we believe that the combination of big data and causal inference tools represents the future of empirical OM research.

To illustrate our point better, we define the 3Vs of big data in the OM context. A relevant data set for OM usually consists of various information on a set of firms

or individuals, possibly over several time periods. Let the number of firms be N , the number of time periods be T , and the set of information be M . We will characterize a data set to be big data if N, T , and M are large, which includes that (1) the dimension of information M is large for each firm or individual (variety), (2) the sampling rate $1/T$ is granular with observations available for every instance of time within the sample period (velocity), and (3) the sample size N constitutes a large proportion of the population (volume). We use an example to demonstrate how the 3Vs may help causal inference models.

5.1. Example: IV Estimation to Address the Endogeneity Bias

Our example relates to services management and investigates how a customer's choice to join an on-call taxi queue is determined by the number of empty taxis. In the taxi service context, there is a pool of taxi drivers who serve a common group of customers. The customers can choose to be in either of two queues: a physical queue and an on-call queue. In the physical queue, the waiting time is uncertain, whereas in the on-call queue, the waiting time is more certain once the customer receives a booking confirmation. Customers have to pay an additional fee to join the on-call queue, so that taxi drivers also earn more revenue per trip from an on-call booking. Our research question carries policy implications on optimal price setting and taxi allocation for the taxi operator. For instance, if the number of empty taxis is found to have a negative impact on on-call sales, then the taxi operator should keep the number of empty taxis low to boost on-call revenues.

Our data set contains electronic records of taxi usage recorded by 3,184 taxi drivers over the span of three consecutive months, in a city with 5.5 million people spread across 29 geographical districts. The 3,184 taxi drivers represent 12% of all taxi drivers in this city. The taxi usage records were collected from information stored in a console installed in each taxi. The console updates the information every 15 seconds. The recorded information includes taxi location (via GPS) and vehicle status (i.e., on break, empty, on call, or passenger on board). Our data set consists of 588,764 GPS location points and 18,747,792 observations on vehicle status. Given the nature of the data set, we consider this to be an example of big data with the 3V characteristics. For our study, we can construct the following variables for each district and each hour in a day:

- *On-call sales*. This is the dependent variable. We calculate this as the number of fulfilled requests for taxis through the on-call service, for each district and hour of the day.
- *Number of empty taxis*. This is the key explanatory variable. We calculate this in two steps. In the first step, we calculate the number of sessions a taxi recorded the status as "free" in each district for each hour of the day.

Then we compute the total number of free sessions for all the taxis in each district for each hour.

- *Average speed (kilometers/hour) of the taxis*. This is a control variable that captures congestion. For each taxi we calculate the average speed while located in a district during each hour of the day. We then compute the average speed of all taxis in each district for each hour of the day.

- *Average duration (in minutes) of on-call trips and non-on-call trips*. We use these as instrumental variables. For each trip originating from a district in a given hour, we calculate the duration of the trip. Then we compute the average duration of the on-call and non-on-call trips.

In our study we examine the effect of the number of empty taxis on on-call sales. However, our dependent variable is potentially censored. This is because we do not observe unfulfilled demand for on-call taxi services. To address this, we use data only from the off-peak hours (all hours except for 8:00 A.M.–10:00 A.M. and 5:00 P.M.–7:00 P.M. on weekdays), because the demand for on-call service during these hours is almost always met (i.e., the potential censoring problem is more likely to occur during the peak hours).

We assume that customers who wish to take a taxi can correctly estimate the number of empty taxis in their district during a given hour. We wish to test whether this estimate has any influence on their decision to join the on-call queue. Other observable factors that possibly influence their decision include their location, the time of day, the day of the week, and the level of congestion. Let Y_{ldt} , X_{ldt} , and Z_{ldt} denote the on-call sales, the number of empty taxis, and the average speed of taxis (proxy for congestion), respectively, at location l , day of the week d , and hour t . Let \mathbf{L} be the vector of dummy variables for the districts, let \mathbf{D} be the vector of dummy variables for the days in the week, and let \mathbf{H} be the vector of dummy variables for the hours in a day. Finally, let $\epsilon_{Y,ldt}$ denote the random error term. A linear regression model for causal inference would be as follows:

$$Y_{ldt} = \beta_0 + \beta_1 X_{ldt} + \beta_2 Z_{ldt} + \delta'_1 \mathbf{L} + \delta'_2 \mathbf{D} + \delta'_3 \mathbf{H} + \epsilon_{Y,ldt}. \quad (5)$$

In the previous model, the parameter β_1 captures the causal effect of the number of empty taxis on on-call sales. This is our parameter of interest.

However, an important question to ask is whether the key explanatory variable (X_{ldt}) is affected by the dependent variable (Y_{ldt}). In our case, it is possible that the number of empty taxis is itself determined by on-call sales. In general, the taxi drivers would incorporate their knowledge of the demand side for their supply choices. In other words, we cannot rule out simultaneity in the aforementioned regression model. If there is simultaneity between these two variables, a direct estimation of this model using OLS regression will not yield a consistent estimate of β_1 .

To address the simultaneity bias problem in the estimation of β_1 , we need at least one IV that is correlated with X_{ldt} but not with $\epsilon_{Y,ldt}$. Therefore, the IV should not affect on-call demand, but should affect taxi availability in the vicinity of a waiting customer. Once a suitable IV is found, we can then proceed to estimate the regression model (5) using the 2SLS approach. In the following section we demonstrate how the three attributes of big data helped us in implementing this procedure.

5.1.1. Variety. The most crucial component in addressing the issue of potential endogeneity is the availability of IVs. A greater variety of data allows us to find suitable IVs. In our context, we need IVs that affect the decisions of taxi drivers, but not those of the customers. Specifically, a taxi driver's decision can be seen as having to choose among three mutually exclusive options: accept an on-call request, accept a passenger from the physical queue or roadside, or remain empty for the time being. We assume that the taxi drivers form an expectation of the benefits and costs for these options before making a choice. A key factor in whether to accept an on-call request is the expected revenue from the trip, which in turn depends on the expected duration of the on-call trip. Note that the destinations of most on-call requests are revealed to the taxi drivers before they decide whether to accept the request. Similarly, the expected revenue of a non-on-call pickup is dependent on the expected duration of such a trip. Therefore, we can assume that the observed number of empty taxis is related to the expected durations of on-call and non-on-call trips originating in each location, day, and hour. In fact, we hypothesize that trips with potentially longer durations are more attractive for the taxi drivers during the off-peak hours, since there are fewer customers during these hours. Consequently, the correlation between the expected durations (for both on-call and non-on-call trips) and the number of empty taxis should be negative.

To operationalize this conceptual connection between the taxi driver's decision to remain empty and the expected durations of the trips in the estimation model, we assume that the actual mean durations of the trips can be a reasonable proxy for the drivers' expectations; that is, we assume that, on average, the taxi drivers make correct estimates about the trip durations. Because of the variety in data, we are able to calculate the amount of time taken until completion for all the trips originating in any location, district, and hour. Then we calculate the average durations of on-call and non-on-call trips for each location, day, and hour combination. Importantly, these variables (i.e., the average durations) are not observable by the passengers, and thus can only affect the supply of empty taxis, and not the demand for on-call taxis. Hence, we can use these variables as IVs.

Before applying the IV-2SLS method to estimate the causal effect of the number of empty taxis on on-call sales, we estimate the model in Equation (5) using OLS. We log transform the variables Y_{ldt} , X_{ldt} , and Z_{ldt} in our estimation, so that our coefficient of interest, β_1 , captures the elasticity of on-call sales. In Table 4, column (1) displays the OLS estimates, which are potentially biased, and column (2) displays the estimates using IV-2SLS, which corrects for endogeneity bias. Our results indicate that on-call sales decrease in locations where there are more empty taxis. Specifically, the bias-corrected coefficient is -2.739 (s.e., 0.134), which suggests that during off-peak hours, if the number of empty taxis increases by 1%, then on-call sales decrease by 2.739%. In contrast, the estimated coefficient using OLS is -0.132 (s.e., 0.008), which is upward biased.

5.1.2. Velocity. High-velocity data permit the investigation of questions that are not feasible with low-velocity data. Suppose that instead of having data at the hourly level, we had access to only lower-velocity data at the daily level. In this case, we would lose important variation in both the dependent and independent variables, and would also lose the hourly dummies as control variables. In Table 4, column (3) reports the estimates generated by the low-velocity data. Note that the estimated effect of empty taxis on on-call sales is markedly different with the low-velocity data, with the sign of the coefficient reversed. A major driver of this is that an important set of control variables, the hourly dummies, is absent in the model. In the model estimated on the high-velocity data, we find that many of the hourly dummies are statistically significant. We also note that the Sargan statistic for the overidentification test indicates that the null hypothesis of the instruments being jointly valid is rejected. In other words, the IVs are not admissible using the low-velocity data.

5.1.3. Volume. The primary advantage of having a high-volume data set is that it helps in generating more precise estimates. To see this, in column (4) of Table 4 we report the model estimates obtained by using a randomly generated subsample of our data, set at 10% of the size of the original data. Although the estimated effect of empty taxis on on-call sales is similar to that for the full sample, the standard errors with the smaller sample are more than twice as large. In some cases, the larger standard errors due to lower-volume data may result in low power for testing the effect of an explanatory variable on the dependent variable.

6. Research Implications for OM and Conclusion

An important question is which areas within OM will benefit from more causal empirical research. To examine this, we review the empirical literature published in

Table 4. Estimates of the Various Specifications in the Example in Section 5

DV: log(<i>on-call sales</i>)	Variety:	Low	High	High	High
	Velocity:	High	High	Low	High
	Volume:	High	High	High	Low
		(1) OLS	(2) IV 2SLS	(3) IV 2SLS	(4) IV 2SLS
Log(<i>number of empty taxis</i>)		−0.132*** (0.008)	−2.739*** (0.134)	1.002*** (0.161)	−2.073*** (0.271)
Log(<i>average speed</i>)		−0.685*** (0.022)	−0.845*** (0.042)	−1.049*** (0.082)	0.760*** (0.108)
District fixed effect (28 district dummy variables)		Yes	Yes	Yes	Yes
Day of week fixed effect (6 day dummy variables)		Yes	Yes	Yes	Yes
Hour of the day fixed effect (23 hour dummy variables)		Yes	Yes	Not feasible	Yes
Observations		42,116	42,116	2,592	4,211
Anderson's LM test		NA	523	103	87.410
P-value of Anderson's LM test		NA	<0.001	<0.001	<0.001
Cragg–Donald weak instrument test statistic		NA	264.4	52.9	43.98
Sargan overidentification test statistic		NA	0.016	22.42	0.088
P-value of Sargan statistic		NA	0.901	<0.001	0.767

Notes. Standard errors are in parentheses. Log(*number of empty taxis*) is instrumented by log(*duration of on-call trips*) and log(*duration of non-on-call trips*), where *duration* is the average duration of trips of on-call sales and non-on-call sales generated at each location and hour combination. Anderson's LM test is the underidentification test.

* $p < 0.1$, ** $p < 0.05$, *** $p < 0.01$.

three top OM journals, *Management Science*, *Manufacturing & Service Operations Management*, and *Production and Operations Management*, from 2000 to 2015.¹⁰ We include only the papers that focus on establishing causal relationships among variables of interest using observational data; thus, papers that are experiment based, simulation based, or forecasting oriented and those employing structural equation modeling are excluded. For papers published in 2000–2009, we include only those with Google Scholar citation counts of 50 or more (as of May 2017). For articles published during 2010–2015, we include only those with 30 or more Google Scholar citations. We categorize these papers into the following broad topics: supply chain management, quality management, services management and retailing, pricing and revenue management, workforce management, and miscellaneous topics. We list these papers in Tables A1 to A6 in the online appendix.

6.1. Supply Chain Management (Online Appendix Table A1)

The major research questions in supply chain management have centered on (a) which factors are the critical determinants of inventory and (b) how various aspects of inventory management impact firm performance.¹¹ Gaur et al. (2005) show that inventory is simultaneously determined with other firm operational outcomes, which suggests that simultaneous equations models would be appropriate empirical tools. Kesavan et al. (2010) apply simultaneity bias correction in their estimation model to study the relationship among cost of goods sold, gross margin, and inventory levels.

Other papers on inventory management employ single equation models, including those by Ton and Raman (2010), Randall et al. (2006), Rumyantsev and Netessine (2007), and Jain et al. (2013). Industry-specific papers include those by Olivares and Cachon (2009) and Cachon and Olivares (2010), which identify the key drivers that explain the variation in the finished goods inventory within the automotive distribution system. There is also work that investigates the factors that affect inventory record accuracy (DeHoratius and Raman 2008) and when managers decide to deviate from the inventory recommendations of an automated ordering system (van Donselaar et al. 2010).

Since simultaneous equations models can be more useful in modeling inventory decisions, it is important to use appropriate IVs and exclusion restrictions for the identification and estimation of the parameters associated with the endogenous variables. Consequently, access to a greater variety of data will be particularly useful. In addition, access to a greater velocity of data will allow for the construction of more precise variables. For instance, van Donselaar et al. (2010) explain how access to more accurate inventory data would have improved the construction of their key dependent variable.

Other well-cited empirical works in the supply chain management area include topics on purchase operations and e-procurement (Boyer and Olson 2002, Mithas and Jones 2007), the bullwhip effect (Cachon et al. 2007, Bray and Mendelson 2012), the effect of supply disruptions on stock market and firm performance

(Hendricks and Singhal 2005a, b), buyer–supplier relationships and communications (Terwiesch et al. 2005, Jira and Toffel 2013), and the effect of adopting a buy-online, pickup-in-store distribution strategy (Gallino and Moreno 2014). The questions that study both the buyers' and the suppliers' decisions are potentially subject to endogeneity issues, as one side's decisions can affect those of the other side. Moreover, research that measures the impact of adopting particular procurement or distribution strategies must account for potential selection bias (Gallino and Moreno 2014). Finally, we note that while there are many analytical models on inventory management, there are only a handful of empirical studies that directly test the predictions of these models (e.g., Olivares et al. 2008).

6.2. Quality Management (Online Appendix Table A2)

The empirical papers published in the quality management area can be divided into two groups. The first group focuses on examining the impact of quality standards on firm outcomes. Corbett et al. (2005) use the event study method to investigate the impact of ISO 9000 certification on financial performance. Levine and Toffel (2010) use a matching model to examine how the adoption of ISO 9001 quality management standards affects employee outcomes such as employee earnings, turnover, and safety. Thirumalai and Sinha (2011) examine what happens to firms when quality drops and applies it to the context of product recalls of medical devices. They apply an event study methodology and find that the product recalls do not have a significant impact on aggregate stock returns. Guajardo et al. (2015) examine the moderating effect of quality on the effects of service attributes on demand in the auto sector. They use the random coefficient model following Berry et al. (1995), which has been widely applied in economics and marketing for demand analysis. An advantage of the approach of Berry et al. (1995) is that it suggests various instruments that can be generated from the data.

The second group of papers examine quality outcomes in the healthcare context. The papers that have attempted to address the omitted variables problem include those by KC and Terwiesch (2011, 2012), and Kim et al. (2015). They all use IV estimation to investigate the causal effects of hospital focus, discharge strategies, and admission policies. Other studies include those by Theokary and Ren (2011), who investigate whether hospital volume and teaching status affect service quality, and Chandrasekaran et al. (2012), who use a random-effects regression to examine how process management affects the quality of hospital stays.

6.3. Services Management and Retailing (Online Appendix Table A3)

In the area of services management, the main topics that have been empirically examined are (a) the causal relationship between customer satisfaction and various operational metrics and (b) the effects of introducing self-service technology and e-services. Two notable empirical studies on the first topic are by Lapré (2011) and Gu and Ye (2014). The notable papers on the second topic include those by Tsiriktsis et al. (2004), Xue et al. (2007), Buell et al. (2010), and Campbell and Frei (2010). In particular, Campbell and Frei (2010) examine how self-service (specifically, online banking) alters services consumption, cost to serve, and customer profitability. Because customers self-select whether to adopt online banking, the authors use the propensity score matching method to control for selection.

In the area of retailing, Heim and Sinha (2001) use data from 52 electronic food retailers to explore how website navigation, timeliness of delivery, and ease of return affect customer loyalty. Chong et al. (2001) use shopping trips and purchase records to estimate a category assortment model that can be used by managers to assess the revenue implications of alternative category assortments. Perdikaki et al. (2012) use a dynamic panel data model to study the relationship between store traffic, labor, and sales performance. They find that while the effect of store traffic on sales exhibits diminishing returns, having more in-store labor can alleviate this effect.

6.4. Pricing and Revenue Management (Online Appendix Table A4)

Notable works include those by Anderson and Xie (2012), Subramanian and Subramanyam (2012), Li et al. (2014), and Phillips et al. (2015). In general, the problems examined in this literature relate to the factors that affect pricing decisions and how different pricing schemes affect revenues. For example, Phillips et al. (2015) investigate the effect of centralized versus decentralized pricing strategies on firm performance. They note the potential endogeneity of price and consumer response, and address this using a control function approach in their model. Notice that pricing mechanisms are typically self-selected by firms. Therefore, the tools discussed in Section 4—propensity score matching, difference in differences, and regression discontinuity design—would be particularly germane.

6.5. Workforce Management (Online Appendix Table A5)

There has also been recent empirical work examining how firms can better manage their workforce to achieve better outcomes (see Siemsen et al. 2009, Bendoly 2014, Narayanan et al. 2011, Staats 2012). KC and Terwiesch (2009), Powell et al. (2012), and KC (2014) examine the

effect of workload on various operational outcomes in the healthcare sector. KC and Staats (2012) and Staats and Gino (2012) examine the impact of specialization on performance. Another line of research examines how the structure of teams, specifically the team's diversity, affects firm performance. Huckman and Staats (2011) answer this question in the context of fluid teams (i.e., teams that are formed and disassembled quickly) in the software services industry. Green et al. (2013) examine how the workload of the team affects the participation of team members. Although the endogeneity issue is likely to be absent given the specific contexts of their studies, future research should account for the fact that workers self-select into teams and that there are potentially unobserved factors that guide how managers construct teams.

6.6. Miscellaneous Topics (Online Appendix Table A6)

Other OM topics that have been empirically investigated, but with a relatively fewer number of papers, include innovation management, information technology (IT) management, queue management, and operations strategy. In innovation management, two notable studies are those by those by Bajaj et al. (2004) and Boudreau et al. (2011). Bajaj et al. (2004) assess the impact of management levers such as oversight, design specialization, and customer interaction on cost savings and scheduling during the two phases of new product development—the design phase and the manufacturing phase. Since the design phase outcomes are explanatory variables in the manufacturing phase regressions, the authors employ a two-stage regression approach. Boudreau et al. (2011) use data from software contests to examine the optimal number of competitors in innovation contests. They rely on their quasi-experimental setting to directly estimate the effect of the number of competitors on innovation outcomes.

The well-cited empirical papers in IT management include those by Ahmad and Schroeder (2001) and McAfee (2002). McAfee (2002) exploits a natural experiment conducted at a U.S. high-tech manufacturer and documents the longitudinal effect of IT adoption on operational performance. The event-study approach used in the paper follows a DID design, where the causal effect of IT adoption is captured by the difference between the control and treatment groups. The causal link between queuing time and firm revenues is an important topic in queue management research. Recent well-cited studies on this topic include those by Allon et al. (2011), Lu et al. (2013), Akşin et al. (2013), Batt and Terwiesch (2015), and Song et al. (2015). Song et al. (2015) use a DID model to show that a dedicated queuing system reduces waiting time relative to a pooled queuing system. On issues related to operations strategy, empirical researchers have examined how various

strategies affect operational performance (see Stratman 2007, Lapré and Scudder 2004, Tsikriktsis 2007, Rawley and Simcoe 2010, Kroes et al. 2012). In particular, Rawley and Simcoe (2010) investigate whether firm diversification leads to increased outsourcing. They employ both the PSM model and IV regression to account for the fact that firms self-select into diversifying their operations.

Conclusion

We conclude this paper with the following remarks. First, the five econometric tools discussed in this paper are nonexhaustive. We chose to focus on these because they are quite intuitive and have been applied to many empirical contexts. For a comprehensive discussion on causal inference models, readers can consult Angrist and Pischke (2009) and Imbens and Rubin (2015). Second, when applying causal inference models to analyzing big data, there are high-dimensional econometric and machine learning techniques, such as LASSO (least absolute shrinkage and selection operator), the post-double-selection method, random forest, and bagging (bootstrap aggregating), that researchers can use to handle large data sets. Interested readers can refer to Tibshirani (1996), Belloni et al. (2013, 2014), Varian (2014), Athey and Imbens (2017), and Wager and Athey (2017) for discussions on such methods. These methods have yet to see widespread applications. However, with data sizes getting increasingly larger, we foresee greater adoption of these methods in empirical OM research over the next few years.

Acknowledgments

The authors would like to thank Ingrid Koch and Ashish Sachdeva for their constructive comments and Rehan Ali for his excellent editorial help. They also thank the editor and two anonymous referees for their helpful comments and suggestions.

Endnotes

¹ We do not discuss empirical models for predictive purposes, such as time-series models. Researchers interested in forecasting can refer to Enders (2004) for a detailed introduction to such models. We also exclude empirical models for experimental data, either from the laboratory or field. Field experiments, particularly well-designed randomized controlled trials, are ideal for causal inference. However, such experiments are usually prohibitively expensive, and statistical analyses of experimental data are relatively simpler.

² We do not discuss nonlinear models and panel data models in this review. The two empirical challenges that we highlight in this paper, endogeneity and selection biases, are also pertinent to such models. Interested readers can refer to Greene (2000) and Wooldridge (2008) for textbook-level discussions on such models.

³ A few examples of structural modeling papers in OM include those by Akşin et al. (2013), Caro et al. (2014), Golrezaei et al. (2014), Kim et al. (2014), Li et al. (2014), and Hyndman and Parmeter (2015).

⁴ A third reason for the endogeneity bias is due to measurement error in the explanatory variables. Because the solution to address this issue is similar to that for the case with omitted variables, that

is, we need to find an instrumental variable, we do not discuss the measurement error problem separately.

⁵The complete data generating process is as follows: $Y_i = 500 + 10X_i + 100Z_i + \epsilon_i$, and $X_i = 20 + 0.5V_i + 6Z_i + v_i$. We randomly assign half the observations into each of the groups $Z_i = 1$ and $Z_i = 0$. The term V_i is randomly generated from $N(50, 5)$, and ϵ_i and v_i are randomly generated from $N(0, 5)$, where N denotes the normal distribution.

⁶Alternatives to the 2SLS approach include the 3SLS method, limited information maximum likelihood estimation, the control function approach, and the generalized method of moments. Readers can refer to a graduate-level econometrics textbook such as Wooldridge (2008) for discussions of these methods.

⁷If there were any exogenous variables in any of the two equations, then those variables should be in the regression as well.

⁸Another example of having a discontinuity in the data is Bennett et al. (2013). The authors exploit changes in the testing protocols for car emissions and find a discontinuity in the passing rate based on the cars' model year.

⁹This assumption is more plausible when the observable heterogeneity among the firms is explicitly taken into account in the estimation model. Angrist and Pischke (2009) provide a discussion on the types of variables that are acceptable as control variables in a DID framework.

¹⁰For *Management Science*, we include empirical papers accepted in the OM department, based on the names of the department editors. Since the department editor information is available only from 2004, we exclude articles published during 2000–2003.

¹¹The early empirical papers were primarily concerned with examining inventory performance over time (e.g., Rajagopalan and Malhotra 2001; Chen et al. 2005, 2007). More recently, researchers have shifted toward establishing causal relationships between inventory and other operational variables.

References

Ahmad S, Schroeder RG (2001) The impact of electronic data interchange on delivery performance. *Production Oper. Management* 10(1):16–30.

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

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.

Anderson CK, Xie X (2012) A choice-based dynamic programming approach for setting opaque prices. *Production Oper. Management* 21(3):590–605.

Angrist JD, Pischke J (2009) *Mostly Harmless Econometrics: An Empiricist's Companion* (Princeton University Press, Princeton, NJ).

Athey S, Imbens GW (2017) The state of applied econometrics: Causality and policy evaluation. *J. Econom. Perspectives* 31(2):3–32.

Bajaj A, Kekre S, Srinivasan K (2004) Managing NPD: Cost and schedule performance in design an manufacturing. *Management Sci.* 50(4):527–536.

Batt RJ, Terwiesch C (2015) Waiting patiently: An empirical study of queue abandonment in an emergency department. *Management Sci.* 61(1):39–59.

Belloni A, Chernozhukov V, Hansen C (2013) Inference on treatment effects after selection among high-dimensional controls. *Rev. Econom. Stud.* 81(2):608–650.

Belloni A, Chernozhukov V, Hansen C (2014) High-dimensional methods and inference on structural and treatment effects. *J. Econom. Perspect.* 28(2):29–50.

Bendoly E (2014) System dynamics understanding in projects: Information sharing, psychological safety, and performance effects. *Production Oper. Management* 23(8):1352–1369.

Bennett VM, Pierce L, Snyder JA, Toffel MW (2013) Customer-driven misconduct: How competition corrupts business practices. *Management Sci.* 59(8):1725–1742.

Berry S, Levinsohn J, Pakes A (1995) Automobile prices in market equilibrium. *Econometrica* 63(4):841–890.

Bertrand M, Duflo E, Mullainathan S (2004) How much should we trust differences-in-differences estimates? *Quart. J. Econom.* 119(1):249–275.

Boudreau KJ, Lacetera N, Lakhani KR (2011) Incentives and problem uncertainty in innovation contests: An empirical analysis. *Management Sci.* 57(5):843–863.

Boyer KK, Olson JR (2002) Drivers of internet purchasing success. *Production Oper. Management* 11(4):480–498.

Bray RL, Mendelson H (2012) Information transmission and the bullwhip effect: An empirical investigation. *Management Sci.* 58(5):860–875.

Buell RW, Campbell D, Frei FX (2010) Are self service customers satisfied or stuck? *Production Oper. Management* 19(6):679–697.

Cachon GP, Olivares M (2010) Drivers of finished-goods inventory in the U.S. automobile industry. *Management Sci.* 56(1):202–216.

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

Campbell D, Frei F (2010) Cost structure, customer profitability, and retention implications of self-service distribution channels: Evidence from customer behavior in an online banking channel. *Management Sci.* 56(1):4–24.

Caro F, Martínez-de AV, Rusmevichientong P (2014) The assortment packing problem: Multiperiod assortment planning for short-lived products. *Management Sci.* 60(11):2701–2721.

Chandrasekaran A, Senot C, Boyer KK (2012) Process management impact on clinical and experiential quality: Managing tensions between safe and patient-centered healthcare. *Manufacturing Service Oper. Management* 14(4):548–566.

Chen H, Frank MZ, Wu OQ (2005) What actually happened to the inventories of American companies between 1981 and 2000? *Management Sci.* 51(7):1015–1031.

Chen H, Frank MZ, Wu OQ (2007) U.S. retail and wholesale inventory performance from 1981 to 2004. *Manufacturing Service Oper. Management* 9(4):430–456.

Chong JK, Ho TH, Tang CS (2001) A modeling framework for category assortment planning. *Manufacturing Service Oper. Management* 3(3):191–210.

Corbett CJ, Montes-Sancho MJ, Kirsch DA (2005) The financial impact of ISO 9000 certification in the United States: An empirical analysis. *Management Sci.* 51(7):1046–1059.

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

Enders W (2004) *Applied Time Series Econometrics* (John Wiley and Sons, Hoboken, NJ).

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.

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

Golrezaei N, Nazerzadeh H, Rusmevichientong P (2014) Real-time optimization of personalized assortments. *Management Sci.* 60(6):1532–1551.

Gray JV, Anand G, Roth AV (2015) The influence of ISO 9000 certification on process compliance. *Production Oper. Management* 24(3):369–382.

Green LV, Savin S, Savva N (2013) "Nursevendor problem": Personnel staffing in the presence of endogenous absenteeism. *Management Sci.* 59(10):2237–2256.

Greene WH (2000) *Econometric Analysis* (Prentice Hall, Upper Saddle River, NJ).

Gu B, Ye Q (2014) First step in social media: Measuring the influence of online management responses on customer satisfaction. *Production Oper. Management* 23(4):570–582.

- Guajardo JA, Cohen MA, Netessine S (2015) Service competition and product quality in the U.S. automobile industry. *Management Sci.* 62(7):1860–1877.
- Heim GR, Sinha KK (2001) Operational drivers of customer loyalty in electronic retailing: An empirical analysis of electronic food retailers. *Manufacturing Service Oper. Management* 3(3):264–271.
- Hendricks KB, Singhal VR (2005a) Association between supply chain glitches and operating performance. *Management Sci.* 51(5):695–711.
- Hendricks KB, Singhal VR (2005b) An empirical analysis of the effect of supply chain disruptions on long-run stock price performance and equity risk of the firm. *Production Oper. Management* 14(1):35–52.
- 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.
- Hyndman K, Parmeter CF (2015) Efficiency or competition? A structural econometric analysis of Canada's AWS auction and the set-aside provision. *Production Oper. Management* 24(5):821–839.
- Imbens G, Kalyanaraman K (2012) Optimal bandwidth choice for the regression discontinuity estimator. *Rev. Econom. Stud.* 79(3):933–959.
- Imbens G, Lemieux T (2008) Regression discontinuity designs: A guide to practice. *J. Econometrics* 142(2):615–635.
- Imbens G, Rubin DB (2015) *Causal Inference in Statistics, Social, and Biomedical Sciences* (Cambridge University Press, Cambridge, UK).
- Imbens G, Wooldridge J (2009) Recent developments in the econometrics of program evaluation. *J. Econom. Literature* 47(1):5–86.
- Jain N, Girotra K, Netessine S (2013) Managing global sourcing: Inventory performance. *Management Sci.* 60(5):1202–1222.
- Jira C, Toffel MW (2013) Engaging supply chains in climate change. *Manufacturing Service Oper. Management* 15(4):559–577.
- KC DS (2014) Does multitasking improve performance? Evidence from the emergency department. *Manufacturing Service Oper. Management* 16(2):168–183.
- 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.
- KC DS, Terwiesch C (2011) The effects of focus on performance: Evidence from California hospitals. *Management Sci.* 57(11):1897–1912.
- KC DS, Terwiesch C (2012) An econometric analysis of patient flows in the cardiac intensive care unit. *Manufacturing Service Oper. Management* 14(1):50–65.
- 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.
- Kim SH, Chan CW, Olivares M, Escobar G (2015) ICU admission control: An empirical study of capacity allocation and its implication for patient outcomes. *Management Sci.* 61(1):19–38.
- Kim SW, Olivares M, Weintraub GY (2014) Measuring the performance of large-scale combinatorial auctions: A structural estimation approach. *Management Sci.* 60(5):1180–1201.
- Kroes J, Subramanian R, Subramanyam R (2012) Operational compliance levers, environmental performance, and firm performance under cap and trade regulation. *Manufacturing Service Oper. Management* 14(2):186–201.
- Lapr e MA (2011) Reducing customer dissatisfaction: How important is learning to reduce service failure? *Production Oper. Management* 20(4):491–507.
- Lapr e MA, Scudder GD (2004) Performance improvement paths in the U.S. airline industry: Linking trade-offs to asset frontiers. *Production Oper. Management* 13(2):123–134.
- Levine DI, Toffel MW (2010) Quality management and job quality: How the ISO 9001 standard for quality management systems affects employees and employers. *Management Sci.* 56(6):978–996.
- Li J, Granados N, Netessine S (2014) Are consumers strategic? Structural estimation from the air-travel industry. *Management Sci.* 60(9):2114–2137.
- Lu Y, Musalem A, Olivares M, Schilkrot A (2013) Measuring the effect of queues on customer purchases. *Management Sci.* 59(8):1743–1763.
- Mani V, Kesavan S, Swaminathan JM (2015) Estimating the impact of understaffing on sales and profitability in retail stores. *Production Oper. Management* 24(2):201–218.
- McAfee A (2002) The impact of enterprise information technology adoption on operational performance: An empirical investigation. *Production Oper. Management* 11(1):33–53.
- Mithas S, Jones JL (2007) Do auction parameters affect buyer surplus in e-auctions for procurement? *Production Oper. Management* 16(4):455–470.
- 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 GP (2009) Competing retailers and inventory: An empirical investigation of general motors' dealerships in isolated U.S. 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.
- Perdikaki O, Kesavan S, Swaminathan JM (2012) Effect of traffic on sales and conversion rates of retail stores. *Manufacturing Service Oper. Management* 14(1):145–162.
- Phillips R, Şimşek AS, Van Ryzin G (2015) The effectiveness of field price discretization: Empirical evidence from auto lending. *Management Sci.* 61(8):1741–1759.
- Powell A, Savin S, Savva N (2012) Physician workload and hospital reimbursement: Overworked physicians generate less revenue per patient. *Manufacturing Service Oper. Management* 14(4):512–528.
- Rajagopalan S, Malhotra A (2001) Have US manufacturing inventories really decreased? An empirical study. *Manufacturing Service Oper. Management* 3(1):14–24.
- Randall T, Netessine S, Rudi N (2006) An empirical examination of the decision to invest in fulfillment capabilities: A study of internet retailers. *Management Sci.* 52(4):567–580.
- Rawley E, Simcoe TS (2010) Diversification, diseconomies of scope, and vertical contracting: Evidence from the taxicab industry. *Management Sci.* 56(9):1534–1550.
- Rosenbaum P, Rubin D (1983) Assessing sensitivity to an unobserved binary covariate in an observational study with binary outcome. *J. Roy. Statist. Soc. Ser. B (Methodological)* 45(2):212–218.
- Rumyantsev S, Netessine S (2007) What can be learned from classical inventory models? A cross-industry exploratory investigation. *Manufacturing Service Oper. Management* 9(4):409–429.
- Siemens 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.
- 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 (2012) Unpacking team familiarity: The effects of geographic location and hierarchical role. *Production Oper. Management* 21(3):619–635.
- Staats BR, Gino F (2012) Specialization and variety in repetitive tasks: Evidence from a Japanese bank. *Management Sci.* 58(6):1141–1159.
- Stratman JK (2007) Realizing benefits from enterprise resource planning: Does strategic focus matter? *Production Oper. Management* 16(2):203–216.
- Subramanian R, Subramanyam R (2012) Key factors in the market for remanufactured products. *Manufacturing Service Oper. Management* 14(2):315–326.

- Terwiesch C, Ren ZJ, Ho TH, Cohen MA (2005) An empirical analysis of forecast sharing in the semiconductor equipment supply chain. *Management Sci.* 51(2):208–220.
- Theokary C, Ren Z (2011) An empirical study of the relations between hospital volume, teaching status, and service quality. *Production Oper. Management* 20(3):303–318.
- Thirumalai S, Sinha KK (2011) Product recalls in the medical device industry: An empirical exploration of the sources and financial consequences. *Management Sci.* 57(2):376–392.
- Tibshirani R (1996) Regression shrinkage and selection via the LASSO. *J. Roy. Statist. Soc. Ser. B (Methodological)* 58(1):267–288.
- Todd PE (2010) Matching estimators. Durlauf SN, Blume LE, eds. *Microeconometrics*. The New Palgrave Economics Collection (Palgrave Macmillan, London), 108–121.
- Ton Z, Raman A (2010) The effect of product variety and inventory levels on retail store sales: A longitudinal study. *Production Oper. Management* 19(5):546–560.
- Tsikriktsis N (2007) The effect of operational performance and focus on profitability: A longitudinal study of the U.S. airline industry. *Manufacturing Service Oper. Management* 9(4):506–517.
- Tsikriktsis N, Lanzolla G, Frohlich M (2004) Adoption of e-processes by service firms: An empirical study of antecedents. *Production Oper. Management* 13(3):216–229.
- van Donselaar KH, Gaur V, van Woensel T, Broekmeulen RA, Fransoo JC (2010) Ordering behavior in retail stores and implications for automated replenishment. *Management Sci.* 56(5):766–784.
- Varian H (2014) Big data: New tricks for econometrics. *J. Econom. Perspect.* 28(2):3–28.
- Wager S, Athey S (2017) Estimation and inference of heterogeneous treatment effects using random forests. *J. Amer. Statist. Assoc.* Forthcoming.
- Wooldridge J (2008) *Econometric Analysis of Cross Section and Panel Data* (MIT Press, Cambridge, MA).
- Xue M, Hitt LM, Harker PT (2007) Customer efficiency, channel usage, and firm performance in retail banking. *Manufacturing Service Oper. Management* 9(4):535–558.