This article was downloaded by: [160.39.34.174] On: 05 April 2025, At: 13:22 Publisher: Institute for Operations Research and the Management Sciences (INFORMS) INFORMS is located in Maryland, USA



Marketing Science

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

Automating the B2B Salesperson Pricing Decisions: A Human-Machine Hybrid Approach

Yael Karlinsky-Shichor, Oded Netzer

To cite this article:

Yael Karlinsky-Shichor, Oded Netzer (2024) Automating the B2B Salesperson Pricing Decisions: A Human-Machine Hybrid Approach. Marketing Science 43(1):138-157. <u>https://doi.org/10.1287/mksc.2023.1449</u>

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

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

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

Copyright © 2023, INFORMS

Please scroll down for article-it is on subsequent pages



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

Practice Prize Paper

Automating the B2B Salesperson Pricing Decisions: A Human-Machine Hybrid Approach

Yael Karlinsky-Shichor,^{a,*} Oded Netzer^b

^a D'Amore-McKim School of Business, Northeastern University, Boston, Massachusetts 02115; ^bColumbia Business School, New York, New York 10027

*Corresponding author

Contact: y.karlinsky@northeastern.edu, https://orcid.org/0000-0003-0098-3273 (YK-S); onetzer@gsb.columbia.edu, https://orcid.org/0000-0002-0099-8128 (ON)

Received: April 8, 2019 Revised: November 16, 2020; June 13, 2022; December 22, 2022 Accepted: February 22, 2023 Published Online in Articles in Advance: June 1, 2023 https://doi.org/10.1287/mksc.2023.1449 Copyright: © 2023 INFORMS	Abstract. We propose a human-machine hybrid approach to automating decision making in high human-interaction environments and apply it in the business-to-business (B2B) retail context. Using sales transactions data from a B2B aluminum retailer, we create an automated version of each salesperson, which learns and automatically reapplies the salesperson's pricing policy. In a field experiment with the B2B retailer, we provide salespeople with their own model's price recommendations in real time. We find that, despite the loss of private salesperson information, reducing intertemporal behavioral biases by providing the model's price to the salesperson increases profits for treated quotes by 11% relative to a control condition. Using counterfactual analyses, we show that although the model's pricing leads to higher profitability in most cases, salespeople generate higher profits when pricing out-of-the-ordinary or complex quotes. Accordingly, we propose a machine learning <i>hybrid pricing strategy</i> with two levels of automation. First, a random forest model automatically allocates quotes to either the model or the salesperson based on its prediction of whose pricing would generate higher profits. Then, if the quote is allocated to the model, the model or the salespeople.
	 History: K. Sudhir served as the senior editor for this article. This paper was selected as a finalist in the 2018 Gary L. Lilien ISMS-MSI Practice Prize Competition. Supplemental Material: The online appendix and data are available at https://doi.org/10.1287/mksc.2023.1449.

Keywords: automation • salesforce • pricing • business to business (B2B) • field experiment

1. Introduction

In the past century, automation has changed the labor market by consistently substituting for predictable and repetitive human tasks. In the early days of automation, its goal was first and foremost scalability and efficiency of well-defined tasks with clear inputs and outputs. Recent advances in computational methods and artificial intelligence (AI) allowed automation to tap into occupations that involve nonroutine aspects, such as judgment, perception and manipulation, creative intelligence and social intelligence (Brynjolfsson and McAfee 2012, Chui et al. 2016, Frey and Osborne 2017). Consequently, automation is bound to transform a significant share of soft-skills-based occupations in the near future (Nedelkoska and Quintini 2018).

Recent applications of automation and AI methods include tasks such as screening resumes (Cowgill 2018), identifying irregularities in computed tomography scans (Chilamkurthy et al. 2018), and replacing judges in deciding whether defendants will await trial at home or in jail (Kleinberg et al. 2018). Yet, although these examples require a high level of expertise (medical doctors, human resource personnel, or court judges), the task is still relatively well-defined, and subjective cues in the environment should play little role in the decision process. That is, the X-ray image or the resume file should contain all (or most) of the information needed to make the judgment.

In this research, we ask whether automation, either in the form of replacing the human agents or supporting them, could be applied to domains where soft skills and interpersonal interactions play an important role in the decision-making process and where interpretation of environmental cues may provide valuable information. Specifically, we introduce automation to one such domain with high importance to marketers: pricing decision making in business-to-business (B2B) retail. The B2B market is estimated at trillions of dollars, yet it largely lags behind the business-to-consumer (B2C) market in adopting technology and automation (Asare et al. 2016). Pricing decisions in B2B are often based on a combination of sales expertise and soft skills. On the one hand, B2B salespeople's pricing decisions are good candidates for automation because they are often repetitive and arguably predictable. On the other hand, such pricing decisions may be difficult to automate because they involve a high degree of interpersonal communication, interpretation of behavioral cues, and persuasion skills.

We collaborate with a B2B aluminum retailer, where salespeople interact with business clients on a daily basis and price incoming requests for products to maximize profitability. The company has thousands of stockkeeping units (SKUs), customizable products, and varying commodity prices, giving salespeople pricing autonomy on a quote-by-quote basis. The pricing process is relationship-based (Zhang et al. 2014), and, in determining prices, salespeople often respond to case-based information available to them. Thus, the same product may be sold to the same client at a different price over time, and in the same time period, the same product may be sold to different clients at different prices. On the one hand, this variation in pricing decisions may be justified, as during the interaction with the client, salespeople may use soft skills to adjust prices according to their assessment of the client's willingness to pay. On the other hand, such variations may be caused by a host of human behavioral decision-making intertemporal biases (e.g., Payne et al. 1993), overweighing contextual information in making pricing decisions, or incentive scheme moral hazard effects. Examples of such biases reported in the context of pricing decisions include, for example, higher loss aversion in the afternoon to recover from morning losses (Coval and Shumway 2005), intertemporal incentive scheme misalignment (Misra and Nair 2011, Larkin and Leider 2012), and adverse selection of easy-to-acquire, but less profitable, customers (Kim et al. 2019). Thus, automation of the salesperson pricing decisions poses a performance trade-off. On the one hand, it can eliminate behavioral intertemporal biases and, consequently, improve performance, but on the other hand, it may miss justifiable response of the salesperson to relevant private information that cannot be automated, but on which salesperson can capitalize. Given this trade-off, it is unclear whether automating the B2B pricing process would be beneficial.

We propose a hybrid approach to automation, in which the salesperson and an automated pricing algorithm participate in the pricing process, utilizing the algorithm's reliability in consistently applying pricing rules and eliminating intertemporal biases and using human judgment for interpreting relevant noncodable contextual cues. In a field experiment and in counterfactual simulations, we show that combining automated and human pricing can lead to higher profits than using either approach separately.

Our automated algorithm is a model version of the B2B salesperson that mimics their past pricing behavior and applies it systematically to new pricing decisions.

We create a representation of each salesperson in the company by regressing the salesperson's past pricing decisions on different variables observed by the salesperson when making the pricing decision (e.g., cost of the material, order size, or the identity of the client). The approach that uses the decision variable (price margin), rather than the outcome (whether the client accepted the price or gross profit conditional on acceptance), is called *judgmental bootstrapping* in the judgment and decision-making literature (Dawes 1979). It allows one to easily capture and consistently apply the salesperson's expertise and pricing knowledge.

In order to test the profit performance of the bootstrap automated-pricing model relative to that of the salesperson, we worked with the B2B retailer to conduct a realtime pricing field experiment. Over the course of eight business days, involving over 2,000 price quotes and over 4,000 product requests (lines), each incoming quote was randomly assigned to either treatment (receive price recommendation based on the model) or control (do not receive price recommendation) to test the causal effect of providing salespeople with the model-based pricing. We worked with the company to integrate our pricing model for each salesperson into their customer-relationship management (CRM) system and provide price recommendations in real time for quotes assigned to the treatment condition. After receiving the price predicted by the model-of-themselves in the treatment condition, the salespeople could decide whether to accept it, adjust it, or keep their original price.

Providing salespeople with price recommendations of their own model led to substantially and statistically significantly higher profits than not providing such recommendations. Specifically, treatment increased profitability by 11%, totaling in added profits to the company of over \$26,000 during the eight days of the experiment, or over \$1.4 million when extrapolated yearly. An instrumental variable (IV) analysis with treatment as the exogenous instrumental variable demonstrates that treatment (price recommendations) affected client acceptance by anchoring salespeople to offer prices that are closer to the recommended model prices.

Looking at salespeople's compliance with the model's suggested prices, we find that salespeople are more likely to comply with the model following a treatment quote that was converted, having observed the usefulness of the model. Similarly, we find that success in converting a control quote leads to lower compliance rates. That is, the salesperson did not use the model and succeeded, and, therefore, is not inclined to use it for the current quote.

To support the findings of the field experiment, we perform a counterfactual analysis that allows us to examine the potential benefits of hybrid automation. We find that although pure automation performs better than the salespeople in terms of profitability, in some cases, using human skills will lead to higher profits than using a model. Consequently, we propose a twolevel human-machine hybrid that combines automation and human decision making to increase profitability: First, we train a machine learning (ML) Random Forest (RF) model that predicts whether human or model pricing would lead to higher expected profitability based on the quote's and client's characteristics (e.g., quote weight or client purchase frequency) and allocates each quote accordingly. Next, if the quote is assigned to the model, the model prices the quote. This hybrid model generates expected profits that are 7.8% higher than those of the salespeople. Aligned with our theoretical predictions that salespeople" shine" when human judgment is required, we find that the hybrid model allocates to the salesperson out-of-the-ordinary or complex quotes (e.g., infrequent clients, quotes that include multiple items or require processing). The proposed hybrid model pushes a step forward on the human-machine continuum in automating not only the pricing decision itself, but also the decision of who should price the quote: the salesperson or the model.

Thus, in this work, we demonstrate that a humanmachine hybrid approach to automation could transform B2B sales. Through a field experiment and supporting counterfactual analyses, we show that using both model pricing (for routine cases) and human judgment (for special cases, possibly with private information involved) generates higher profits to the company than pure human pricing. The company with which we collaborated is implementing our model permanently into its CRM system.

The remainder of the paper is organized as follows: Section 2 discusses our contribution to the work on B2B pricing and automation. Section 3 lays out the specification of the bootstrap model of the salesperson and the empirical context for evaluating it. Section 4 describes the field experiment conducted with the company and explains the mechanism underlying the improved performance due to automation. Section 5 describes the counterfactual analysis used to create the human-machine hybrid, and Section 6 concludes by discussing implications of our findings to salesforce automation.

2. B2B Pricing and Automation 2.1. B2B Marketing

Our work builds on and contributes to several streams of literature. We add to the relatively limited literature on B2B marketing (Grewal et al. 2015, Lilien 2016), and specifically on B2B pricing. The B2B market was estimated at nearly \$9 trillion in transactions in 2018. Nevertheless, B2B pricing decisions remain a relatively understudied topic in the literature. Increasingly, sellers face business clients that prefer to interact and place orders via e-commerce (Forrester 2018). It is therefore of great interest to examine the possibility of automating pricing decisions in a B2B context.

Buyer-seller interactions in B2B are typically longterm and relationship-based (Morgan and Hunt 1994, Lam et al. 2004). Variation of prices across clients and across purchases is common in B2B (Zhang et al. 2014). Consequently, responding to clients' needs and understanding their state of mind are essential to the B2B salesperson's job when it comes to making pricing decisions. Although automation has gone a long way with respect to emulating human behavior, "the real-time recognition of natural human emotion remains a challenging problem, and the ability to respond intelligently to such inputs is even more difficult" (Frey and Osborne 2017, p. 262). Therefore, the potential benefit from automating B2B pricing decisions is unclear.

2.2. Automation and Judgmental Bootstrapping

We add to the literature on automation by providing an empirical test for automating the B2B salesperson's job. Although automation has come a long way in substituting for human tasks, automation of soft skills is still sparse (Deming 2017). Research in labor economics shows that automation can substitute for workers in performing tasks that follow explicit rules, whereas it complements them in performing nonroutine problem solving and communication-based tasks (Autor et al. 2003). The salesperson's job is a combination of repetitive, technical calculation of prices based on quote characteristics and delicate use of soft skills and communication to understand the client's state of mind and maximize profits.

The roots of our approach to automation lie in the behavioral judgment as well as the decision models literature. The former stressed the idea that models of experts trumpet experts in judgments and decision making (Meehl 1954, Dawes 1979). In a judgmental bootstrapping (JB) model, the judgment (e.g., price), rather than the outcome (e.g., profit), is used as the dependent variable in the model of the expert. Consequently, model coefficients reflect the weight that the expert puts on each variable in making the judgment, creating a paramorphic representation of the expert's decision policy (Hoffman 1960). Applications of JB include predicting students' performance (Wiggins and Kolen 1971) and bootstrapping psychiatric doctors (Goldberg 1970) and financial analysts (Ebert and Kruse 1978, Batchelor and Kwan 2007), as well as some limited applications to managerial tasks (Bowman 1963, Kunreuther 1969, Ashton et al. 1994).

Why should automation of the salesperson through JB perform better than the expert? Ultimately, JB uses less information (only codable information) and may repeat inefficiencies in the experts' past decisions. Empirical demonstrations for the superior performance of JB over experts proposed in the decision-making

literature point out to its ability to eliminate *intertem*poral biases. People are inconsistent decision makers, who overweigh noisy, but salient, inputs where deviation is not needed. Although human judgment of contextual cues can be useful in some cases, it hurts reliability in most cases (Dawes et al. 1989, Kahneman 2011). In the context of salesforce behavior, such biases may include overcorrecting for losses earlier in the day (Coval and Shumway 2005), intertemporal incentive scheme misalignment (Misra and Nair 2011), or adverse selection of clients in light of private information (Kim et al. 2019). The JB model may perform better in such cases by appropriately and consistently weighing the information according to rules extracted from the human decision policy and limiting the effect of casebased biases and nonrelevant contextual information on the human decision maker's judgment (Meehl 1954, Armstrong 2001).

We formulate the JB model and the trade-off between case-based inconsistencies and private information as follows. Let p_t be the salesperson's price for quote request:

$$p_t = g(X_t) + f(z_t) + \epsilon_t, \tag{1}$$

where X_t is a vector of observed transaction characteristics; z_t is a mean centered private signal observed by the salesperson, but not by the researcher; and ϵ_t is an unobserved random error with mean zero and standard deviation σ_{ϵ} that accounts for intertemporal inconsistencies in pricing. Our bootstrap model is the $E[p_t | X_t] = g(X_t)$. This model integrates over both ϵ_t and $f(z_t)$.

 ϵ_t is a random noise that is almost always damaging to the salesperson's pricing behavior, and eliminating it using the JB model is likely to improve decision making. The effect of $f(z_t)$ is more nuanced. As mentioned earlier, salespeople may use private information in their decision making. Such information is captured by $f(z_t)$. $f(z_t)$ may be a useful response to information like increasing the price for an urgent order or a response to, for example, the misalignment of the incentive system structure. $f(z_t)$ may also be harmful if it is overweighted relative to $g(X_t)$, which is known in the JB literature as overweighing case-based information—for example, overreacting to the client's interpersonal interactions.

A condition to the superiority of the JB model over the expert is that the benefits from eliminating the noise induced by σ_{ϵ} and possible negative effects of the response to private information in $f(z_t)$ overcome the loss induced by possible positive effects of $f(z_t)$. This condition is guaranteed if most of the information used in the decision-making process is available and codable, as has been the case in much of the JB literature. In the B2B context, it is believed that salespeople, to a large extent, use private signals in their pricing. The empirical question of whether automating the B2B salesperson is beneficial is at the heart of this paper. Our experiment and the supporting counterfactual analysis explore hybrid approaches that balance the benefits of private signals, salesperson expertise, and the consistency offered by automation.

It should be noted that because our model mimics the salesperson behavior, rather than an "optimal" pricing model, any time-invariant biases exhibited by the salesperson (e.g., a salesperson who consistently overprices/underprices) will continue to exist in the model. Another question is why automate via JB of the salesperson and not use optimal prices based on an estimated model of demand. There are several reasons for this choice. First, it allows us to directly test the tradeoff between σ_{ϵ} and $f(z_t)$. Although this automation approach does not eliminate systematic biases, it captures the decision maker's expertise and applies it consistently, preserving human knowledge. Second, deciding on optimal prices requires strong assumptions about the nature of demand, such as the degree to which consumers are forward looking or the nature of competitive responses. Our approach to automation does not require making such assumptions, as it only builds on salespeople's historical decisions. Third, and related to the second point, our goal is to implement the pricing decisions in real time, which may not be feasible with optimal pricing calculation. Finally, given the sensitivity of replacing salespeople with an automated solution, we opted for a model that relies on the salesperson's expertise, rather than a black-box model, which is likely to face more resistance from salespeople.

3. The Model of the Salesperson

Our approach to automation is to create a model of each salesperson that will learn their pricing policy based on their pricing history and apply it to new incoming quotes. For every salesperson separately, we estimate a model of previous pricing decisions as a function of a set of variables available to the salesperson at the time of decision. Although we observe the outcome of the offered price quote—that is, whether the client accepted it or not—it is not included in the model because the goal is to create a JB model that mimics the salesperson's pricing behavior. Then, the model can be used to replace every salesperson with a consistent and automated version of themselves to price a new set of quotes.

3.1. Data

The empirical context and data we use to calibrate the model of the salesperson come from a U.S.-based metals retailer that supplies to local industrial clients. The company has sales teams in three locations in Pennsylvania, New York, and California. In each of these locations, there is a team of salespeople servicing mostly, but not exclusively, clients from the area. The retailer buys raw aluminum and steel directly from the mills, cuts them according to the specification provided by the client, and ships the products to the client. Clients may be small to medium-sized industrial firms (e.g., machine shops, fabricators, or small manufacturers). The company sells thousands of SKUs in nine product categories, seven of which are subcategories of aluminum (the other two, stainless steel and other metals, represent less than 2% of the lines in our data; see Table A1 in Online Appendix 1). Aluminum categories vary in terms of the shape of the metal, their thickness, and their designation (e.g., aerospace versus commercial). Because of the large number of SKUs, the dynamic nature of this industry in terms of varying commodity prices, and the high customization of products, there is no price catalog available. Thus, despite the products themselves being commodities, the service that the retailer provides makes these transactions more customized and unique. The salesperson has a high degree of autonomy in pricing products on a quote-by-quote basis, providing different prices to different clients and even different prices to the same client over time.

A client may request a price quote via email, fax, or by calling the supplier. Although the workflow in the firm allows any available sales agent to pick up the call and provide a price quote, most clients interact with the same salesperson on most purchase occasions. When requesting a price quote, the client specifies the requested metal, size of the piece, whether cutting is required, and the quantity. A quote from a client may include only one SKU or multiple SKUs, which we define as lines. After receiving the order's specifications, the salesperson provides a price quote.¹ Salespeople are guided and incentivized to maintain high price margins. After receiving the price quote, the client decides whether to accept or reject the quote, given the price in the quote. In this industry, price negotiations

 Table 1. Descriptive Statistics of Quotes and Orders per Line

beyond the first-level negotiation of price quote and acceptance are rare. We verify this empirically by comparing the initial price from the quote to the final invoice price and find the prices to be identical in over 99% of the cases.

The data include transaction-level information of price quotes spanning 16 months from January 2016 to April 2017. The sample includes 3,863 clients, with an average of 36 product requests per client.² Each of the 17 salespeople in the sample made, on average, over 8,000 pricing decisions. A sales order may include one or more products (lines); each line is priced separately. The sample includes 67,851 price quotes, with an average of about two lines per quote, totaling 139,869 pricing decisions (every line is a "pricing decision"); 56.9% of the quotes were accepted by the clients (i.e., converted into sales orders). See Table 1 for line-level summary statistics of the data.

3.2. Model Specification

To standardize across products and order sizes, the firm uses price margins, as opposed to price or price per pound to evaluate its pricing strategy. Therefore, we use price margin in building the automated pricing model. The margin provided by salesperson s for line l in quote q for client i is defined as:

$$m_{lqis} = \frac{p_{lqis} - c_{lq}}{p_{lqis}},\tag{2}$$

where c_{lq} is the cost per pound (known to the salesperson) that the company paid to buy the material and p_{lqis} is the price per pound provided by the salesperson for the specific price quote q.³ Because the firm always prices above cost, price margins could range from zero to one and are somewhat skewed to the left. The average line price margin in the data is 41%, and the median is 36%. Consequently, we use the logarithmic transformation of price margin as the dependent variable in the

Variable	Mean	Std. dev.	Lower 10%	Median	Upper 90%
Line margin	0.41	0.20	0.20	0.36	0.72
Price per lb.	4.78	25.06	1.67	2.60	7.19
Cost per lb.	1.98	10.64	1.18	1.40	2.74
LME ^a price per lb.	0.76	0.07	0.68	0.75	0.86
LME price volatility	0.01	0.00	0.00	0.01	0.01
Weight (in lbs.)	352.30	683.54	16.09	117.00	892.77
Client recency (in days) ^b	61.86	207.92	1.00	13.00	120.00
Client frequency (per week) ^b	0.62	0.68	0.08	0.41	1.39
Client previous order \$ amount (log) ^b	6.52	1.39	4.88	6.39	8.37
% of quotes priced by same salesperson	0.78	0.31	0.14	0.93	1.00
Cut required	0.32	0.46	0.00	0.00	1.00
Feet base	0.03	0.18	0.00	0.00	0.00
Sale (quote converted)	0.57	0.50	0.00	1.00	1.00
Total = 139,869					

^aLondon Metal Exchange.

^bCalculated at the product category level.

pricing model. In building the model, we attempt to include all the information available to the salesperson at the time of the pricing decision. We conducted several interviews with senior management and salespeople in the firm to get an idea of the information flow along the pricing process. Additionally, we capture all of the information recorded on the firm's CRM software that salespeople use when determining prices (see a screenshot of the CRM system in Figure 1). The model includes the following variables:

a. *Product category*. Dummy variables for eight out of nine product categories the retailer sells (Baseline category is Aluminum–Cold Finish).

b. Weight. Log of total line weight in pounds.

c. *Relative weight*. Because pricing may be affected not only by the weight of each line, but also by the weight of the entire order, we include the relative weight of the line in the overall order to capture possible quantity discount at the quote level.

d. *Cut*. Made-to-order piece often require processing. We include *cut* in the margin equation as an interaction between the cut dummy variable and 1/weight.

e. *Cost*. The cost per pound for the requested part number, as displayed to the salesperson in the CRM system, which reflects the price the company paid for the material.

f. *Commodity market prices*. Salespeople have access to market prices published by the London Metal Exchange

Figure 1. (Color online) Screenshot of the CRM System

(LME). We include the daily LME price per pound, as well as the volatility of LME prices in the week prior to the date of the quote (measured by the LME standard deviation during the past five business days).

g. *Foot-base products*. A dummy variable for whether the product is priced per feet rather than per pound (3.5% of the items).

h. Client characteristics

a. *Priority*. The firm prioritizes clients based on orders volume in the preceding 12 months.⁴ We include priority in our model using a set of dummy variables. A client's priority may change over the data window because it is updated by the firm every six months (baseline priority is Priority A).

b. *Recency, frequency,* and *monetary—RFM. Recency* is defined as days since the client's last quote request from the same product category; *frequency* is defined as the client's running average of requests from the product category per week since their first quote request; and *monetary* is defined as the log of the total dollar amount of the client's last order in the product category.⁵

c. *Client random effect*. One of the most prominent characteristics of B2B pricing is that prices can vary across clients (Khan et al. 2009). To account for client-specific pricing based on the client's identity, we include *client random effect* in the model.

File Sales	Purchasing Inventor	y Financials Operation	s My Pentago	n CRM	Administratio	on Internet	Window	Help			
Customers List	t (E) Beceiving	Auto	matic Invoicing	Custom	ver Special Conti	acts 🗿 Advice	of Delivery	Sales Forecast	Archive History	Inter Company	
1 Quotes	👧 Pick & Ship 👘 Invoid	ce 🗇 📶 Shipi	per Consolidator	Contrac	t Invoicgs	Advice	of Shipment	Phone Order	Import SPEC2000		
Sales Orders	B.M.A.	Print Documente + 10 0. Re	ceiver / WO by PN	Contrac	t Management	P- Custon	ner Survey	2. e-Comm Pend	ing Export SPEC2000	DLA Solicitation	
Sal	les Documents	occuments .									
	0 52 20	0/ S/ O		6							
1 W					You You	have 4982 Unsav	ed Queries!				
New Add	Qui <u>c</u> k <u>D</u> ei Ob <u>I</u> r	n Quote S.Q. Srch	29X6 Peeb	Settings	Close	(7) DI A 7	Lille.		hice		
Acct: 317814	1	Contact :	2.1	Het	NO: 324	Ibc Phonty: E	E	Curren	cy: USD		The second
D Line Type	P/N	Description	on	1 7075 7054		Spec			MFG Q. Req	UM Unit Price	UM Q
•] 01	H7531651ND-144	3 DIA X	144" Aluminum Ho	and 7075 1651		AMS	QQ-A-225/9, A	STM 8211	1.1	00 18 0.0	1000 LB
4	C D (10.1	11 (P		0010	0		0.4			
lotal Lines 10	Cust P/N	Category ALU/U	11 01	7N Type (Hi	UBAH	Part Set	1,253 BFFJJ	Lube ev	ery i		
History Qust Info	P/N Memo Ket Elec	ctronic Catalog (y) <u>R</u> epair Rej	ections Capability	BOM Ex	ternal Source						
	Quote History				Stock Info				R.F.Q Hist	ory	
2 2 4 0	11 🐌 🔊 🖭 📆 10	÷ # 🖬	Qty Cost S	Exch 0	utbound Inbour	nd Stock UM		V 12 Q2	TT 🐚 🗐 🖭 📆 10 🛟	12 C	
Customer	Qty. Bid UM Unit Prid	ce (Base) Priced 🔺	In Church a	0.000	P/N Screen	Officia		Vendor	Brk Qty. Bid	UM Cond	Unit Price
	2.000 EA	31.6500 2/26/2015 4:05	In Stock :	0.000	Images 1.4	Elect Catalog					
	1.000 EA	3.6000 1/14/2015 2:29	Internal Use:	0.000	Deserved by	Elect Calaby					
A CONTRACT OF CONTRACT	1.000 EA	3.5000 11/7/2014 9:58	Consigned :	0.000	Documents	111					
	11.000 EA	3.3000 11/4/2014 3.42	Other Stock :	0.000	Uncomm	utted P.U.		10	«No data to di	plano	
and the second	1 000 EA	3 4500 9/30/2014 1.10	QA/Inspc :	0.000	Consumpt	ion Analysis					
	5.000 EA	3 6100 8/29/2014 1:04	Transport :	0.000	Calculate 0	Consumption					
	1.000 EA	395.0000 7/28/2014 2:16	Quarantine :	0.000	Consumption	n					
TANGOGAG META	101 0000 54	0.0000 0.000 000 0.000	Available Stock :	0.000	This Year(2015)	517.000		4			
Tupe : 01 - Inventor	ru ltem		Available ToSell	2,793.392	Last Year(2014)	: 1,653.00C		Tune : 01 - Inve	entory litera		
1990.01.101.000	dy insure		Private Toren .		2 Tears Ago(20	13; 7,540.000		1300.01 . 01	arrively risks		
	Sales History (Select For	(Details)			W/H location				Purchasing History (Sel	ect For Details 1	
1 6 0 4	10 11 1 2 3	Beenflag	网周期山	· · ·	9			000	10 1 1 1 1 2 2		
Customer	Oty Order UM	Unit Price (Base) Ship On	All Emoty C	and W/H	Mfa Outside F	lepairs		Vendor	Qty Order + UM	Unit Price (B) State	as Ship On
	2.000 EA	31.6500 3/4/2015	W/H Loc M	1	FACON	(Avail	Cost	-	1,131.000 LB	2.6500 CL	6/13/2014
	3.000 EA	368.5000 1/15/201			less (all	Pretrain.	- www	-	1,135.000 LB	2.5700 CL	5/13/2013
	1.000 EA	398.7500 12/30/20							1,135.000 LB	2.5700 CL	5/7/2013
	1.000 EA	3.4500 10/6/201						-	2,000.000 LB	2.9300 CL	8/15/2013
	3.000 EA	360.5000 6/23/20					Concernence of		2,261.905 LB	2.3900 CL	5/28/2014
	2.000 EA	406.0000 6/9/2014					Show all wa	rehouse locations wh	ere quantity>0 3,000.000 LB	2.4400 0	5/20/2015
	5.000 EA	372.0000 5/24/20			No data to displa	62		-	29,152.000 LB	2.5700 CL	9/27/2013
	4.000 EA	3.3000 3/10/201									
	4.000 EA	300.0000 10/29/20									
	JULUOU EA	345.0000 10/14/20									
								1			
		•						4 11			

i. *Client-salesperson history*. Relationship with the client could affect the salesperson's pricing behavior. On the one hand, long-term familiarity with the client can increase the salesperson's persuasion power. On the other hand, it may bias their pricing decisions (e.g., pricing may become too lenient). As a measure of the salesperson-client relationship, we calculate the proportion of quotes up to date that the salesperson priced with the focal client out of the total number of client quotes. (i.e., we measure to what extent this is the client's regular salesperson). On average, the same salesperson handles the client nearly 80% of the time.

j. *Time dummies*. To control for any time trends, we include quarter dummies (baseline quarter Q1 of 2016).

3.3. Model Estimation and Results

We estimate a linear regression separately for each salesperson to extract the weight each salesperson puts on each variable in setting the price margins for the requested quote. For each line *l* of each quote *q* priced by salesperson *s* for client *i* in the sample, we regress the logistic transformation of the price margin m_{lqis} (as defined in Equation (2)), on the set of line characteristics and time-varying client characteristics, x_{lqir} , as well as salesperson-client random effect, α_{is} :

$$\log\left(\frac{m_{lqis}}{1-m_{lqis}}\right) \sim \alpha_{is} + \boldsymbol{\rho}_{s} \boldsymbol{x}_{lqi} + \zeta_{lqis}, \tag{3}$$

where ζ_{lqis} is a normally distributed random shock. Note that the subscript *s* in Equation (3) means that we estimate Equation (3) for each salesperson *s* separately. However, to get a sense for the effect each variable has on the log price margins, Table 2 presents the results of

Table 2. Bootstrap Pricing Model with Salesperson Fixed

 Effects

Variable	Coefficient	Std. err.
Cost per lb.	-0.003***	(0.000)
LME per lb.	0.860***	(0.076)
LME volatility	-1.454^{**}	(0.462)
Weight (log)	-0.469^{***}	(0.001)
Relative Weight	0.270***	(0.005)
Cut/weight	0.303***	(0.007)
Foot base	-0.232***	(0.009)
Recency	0.00001	(0.000)
Frequency	-0.077***	(0.004)
Monetary (log)	0.003*	(0.001)
Regular salesperson	-0.018*	(0.008)
Intercept	0.646***	(0.068)
Observations	139,869	
R^2	67.1%	

Notes. DV is Logit-transformed price margins. Regression includes client random-effect, salesperson fixed effect, client priority dummies, category dummies, and quarter dummies.

p < 0.05; p < 0.01; p < 0.01

a mixed model with client random effect and salesperson fixed effect estimated on the whole sample.⁶

The automated version of the salesperson captures salespeople's pricing policy well—the regression model explains nearly 70% of the variation in the pricing policy. Thus, $g(X_t)$ in Equation (1) captures 70% of the variance salesperson's pricing decisions. Furthermore, when converting log price margins back to price margins, the average predicted line price margin of 41.96% is very similar to the average observed line price margin of 41.14%.

We find that when cost increases, salespeople decrease price margins. However, when the daily metal price increases, salespeople pass through some of the increase to the consumers (controlling for the cost of the material to the company). High variability in market prices leads to lower price margins. Salespeople employ quantity discount in pricing, such that larger orders have lower price margins. As expected, processing (cut) increases price margins. With respect to client behavior, the company provides lower price margins to customers who buy frequently, but salespeople charge higher price margins for clients whose previous order was large. We find that clients receive lower price margins from their regular salesperson, suggesting that relationship building may lead to pricing "softening."

4. Randomized Field Experiment

To assess the value of automating the salesperson pricing decisions through the individual pricing models, we collaborated with the company to conduct a largescale field experiment. Ideally, the automated prices would replace salespeople's prices altogether. However, due to the immediate impact that such a pricing experiment can have on the company's profits, we were only able to provide the model's prices as (real-time) recommendations and allow salespeople to adjust their original prices accordingly.

4.1. Experimental Design

In collaboration with the B2B retailer's information technology team, we created a "price calculator," which, upon receiving a new quote, calculates the model's predicted price margins per Equation (3) based on the quote, client, and salesperson characteristics. The calculated price per pound⁷ is then displayed in real time as a recommendation to the salesperson. The experimental design randomly allocates incoming quotes into treatment (60% of the quotes) and control (40% of the quotes).⁸ The regular pricing workflow is as follows: When a client puts in a new quote request, the salesperson enters the new quote information (client ID, SKUs requested, etc.) into the CRM system. The salesperson then provides a price quote, saves it to the system, and is able to edit prices as needed. When done editing, the Figure 2. (Color online) Emails Sent to Salespeople as Part of the Field Experiment



Accept suggested prices Accept original prices Edit quote prices

Notes. (a) Treatment email format. (b) Control email format.

salesperson generates a price-quote document and sends it to the client via email.

In the experimental intervention for quotes in the treatment condition, after the salesperson entered their quoted pricing information, the following message was emailed to the salesperson: "Based on your previous pricing decisions, the prices recommended for this quote are:", and below was a table displaying the product information for every line of the quote, the price that the salesperson had just entered into the system, per pound and per unit, and total per line, as well as the model's price per pound and per unit and total per line (see Figure 2(a) for a screenshot of the email). The salesperson could then either click Accept suggested prices to update the sales system to reflect the model's prices, Accept original prices to keep her original prices, or Edit to edit any price manually (see Figure A2 in Online Appendix 2 for the edit form). Prices were automatically updated in the sales system and sent to the client, therefore not requiring an extra step on behalf of the salesperson. The flow of the experiment is depicted in Figure 3.

Because treatment involved an extra step of evaluating the original prices, which may, in and of itself, generate higher attention of the salespeople to their pricing decisions, an email was also sent to quotes in the control group. The control email was similar to that of the treatment, except that it did not include the columns displaying the model's recommended price (see Figure 3(b) for a screenshot of the control group email). Similar to the treatment condition email, the control condition

Figure 3. (Color online) Flow of Field Experiment



u bject ello C uote	: Pricing Calculator: Quote #737659 athleen, No: 737659	D)		
uston ased Line	ner: Description on your input, the prices recommended P/N & Description	for this qu Qty Bid	ote are: Your Price	Your Tota
1	P52.25H32-96-48 .250 X 48 X 96 Aluminum Plate 5052 H32	2.000 EA	\$201.00/EA (\$1.80/LB)	\$402.00
2	S52.19H32-96-48 .190 X 48 X 96 Aluminum Sheet 5052 H32	1.000 EA	\$149.00/EA (\$1.75/LB)	\$149.00



S

Н

C

C

В

email allowed the salesperson to either Accept or Edit their original prices. If edited, prices were updated directly in the system. The salesperson's next step in both control and treatment flows was to go back to the system, generate the price-quote document, and send it to the client, as they would have done without the experiment.

Note that when entering their original price quotes, salespeople did not know whether each quote belonged to treatment or control (i.e., whether they will receive a price recommendation or not); hence, the original price quotes are independent of the experimental manipulation. This unique design gives us knowledge of three data points for each quote (control and treatment): the original price set by the salesperson, the model's recommended price (calculated in both control and treatment, but made available to the salesperson only in the latter), and the final price that the salesperson provided to the client. We use this information in subsequent analyses to shed light on salespeople behavior in the experiment.

Prior to the commencement of the experiment, we let the salespeople experience the tool for four business days, during which we adjusted the tool to fit best into their workflow and corrected any technical issues that arose. During those pretest days, we visited two out of the three company locations (New York and Pennsylvania) and conducted several phone conversations with the salespeople in the third location (California) to make sure salespeople were comfortable using the tool and understood its flow.

We ran the experiment for eight consecutive business days, from June 2 to June 13, 2017. Our data include 2,075 quotes by 1,045 clients, with a total of 4,142 pricing decisions (each line is a pricing decision).⁵ The average compliance level with the tool (i.e., quotes for which salespeople either fully accepted the recommended prices or edited prices in the direction of the recommendation) was 19.48%. We note that in our analyses, we use intention to treat (price recommendation), as opposed to compliance (whether the salesperson adopted our price recommendation), because

Total

compliance is endogenous. Hence, considering the compliance levels, our results may underestimate the true effect of automation. We further discuss salespeople's compliance behavior in Section 4.3.2.

4.1.1. Randomization. Every incoming quote was assigned to the treatment group with probability 0.6 or to the control group with probability 0.4. Indeed, 58.3% of incoming quotes were assigned to the treatment condition. As with any experimental design, the first order of business is to examine that the randomization was performed correctly. We performed a randomization check for different quote variables, such as average cost, total weight, number of lines requiring cut, and number of lines per quote, as well as the original price set by the salesperson, the model's price, and the difference between them. We find no statistically significant differences between the treatment and control conditions on these variables (all p > 0.23; see Online Appendix 2.2).

4.1.2. Stable Unit Treatment Value Assumption. The relatively small number of salespeople in the company was the key reason to randomizing at the quote level, rather than at the salesperson level. When choosing a within-subject design, where some of the salesperson's quotes are treated, whereas others are not, there is a risk of violating the stable unit treatment value assumption (SUTVA; Rubin 1980). That is, that treatment of quotes in the treatment group "contaminates" the quotes in the control group because the same salesperson prices both the treatment and the control quotes. One possible mechanism through which such contamination may occur is learning. If, for example, the salesperson receives a few consecutive treatment emails recommending higher prices than their original prices, they may adjust their pricing upward in the following treatment and control quotes.

To evaluate the extent to which learning is affecting pricing, we compare the difference between the model's price per pound and the salesperson's original price per pound over time, for control and treatment quotes. Although we expect that the model maintains the same pricing rule, if the salesperson learns over the course of the experiment to price more systematically and more similarly to the model, the difference between the salesperson's original prices and the model's prices will decrease over time. Figure 4 shows the average absolute difference between the model's price and the original salesperson's price over the eight days of the experiment. We see no apparent pattern in the difference between the model and the salesperson pricing in either of the experimental conditions over the course of experiment, suggesting that violations of SUTVA due to learning are likely to be minimal

To statistically test possible violations of SUTVA via the effect of one quote on a subsequent quote, we tested **Figure 4.** Average Absolute Difference between Model Price per Pound and Original Price per Pound Over the Eight Days of the Experiment: Treatment vs. Control



whether the treatment given to a quote affects the pricing by the same salesperson in the following quote. For each line in a quote, we regress the absolute difference between the model's price per pound and the salesperson's original price per pound on a dummy variable indicating whether the previous quote priced by the salesperson was treated, controlling for the set of line characteristics, time-varying client characteristics, salesperson fixed effect, and salesperson-client random effect. If SUTVA violations exist, the salesperson will price more similarly to the model following a treatment quote, as they can learn from the treatment quote pricing. However, we do not find a statistically significant relationship between whether the previous quote belongs to the treatment condition and the difference between the salesperson's and the model's prices in the current quote ($\beta_{previous_quote_treated} = 0.0019, p = 0.959$). See Online Appendix 2.3 for full details.

Although the results of the SUTVA analysis are encouraging for the purpose of causally estimating the treatment effect, they may be disappointing from a training point of view, in the sense that, at least within the eight days of the experiment, observing the model's prices was not sufficient to change salespeople's inherent pricing decisions.

4.2. Field Experiment Results

4.2.1. Nonparametric Test. To test the effectiveness of the treatment (providing price recommendation), we compare the gross profit (GP) between treatment and control quotes. GP can go from zero to a large number. Because quotes that were not converted to sales (i.e., the client declined the offered price) have zero GP, the distribution of GP has a mass at zero. Accordingly, we use a nonparametric test to compare the GPs between the treatment and control conditions. In addition, although randomization was done at the quote level, pricing was done separately, but not independently, for each line within the quote. To account for such interdependence,

we cluster the standard errors across lines of the same quote. Specifically, we use a nonparametric Wilcoxon rank-sum test with clustered standard errors for lines within a quote (Datta and Satten 2005, Jiang et al. 2017) to compare mean line gross profits between treatment and control conditions. We find that quotes in the treatment group have statistically significantly higher gross profits per line relative to quotes in the control group (Diff = \$10.95, $GP_{control} = \$94.16$, $GP_{treatment} = \$105.11$, Z = -2.132, p = 0.033). Overall, the increase in profits corresponds to over \$26,000 for the treated quotes during the eight days of the experiment and over \$1.4 million when extrapolated to all quotes handled by the company in a year. Thus, automation in the form of recommending salespeople their own model's prices can result in significant and substantial increase in the company's profits.

4.2.2. Cragg Hurdle Regression Analysis. The positive effect of treatment on profits and margins could come from increasing the number of quotes that were accepted and/or from higher price margins of accepted quotes. In order to further understand the mechanism underlying the positive effect of providing price recommendations to quotes in real time, we estimated a Cragg hurdle regression (Cragg 1971). The Cragg hurdle model enables the estimation of the treatment effect separately on the two observed processes: acceptance

Table 3. Cragg Hurdle Regression Analys	is
---	----

of the suggested price by the client and GP level conditional on acceptance of the price (GP is zero if the client rejects the price offer).¹⁰ Specifically, we use a normalized log(1 + GP) as dependent variable (DV) and define its distribution using the following model:

$$f(\log(1 + GP) | \mathbf{x}_{lq}^{1}) = \begin{cases} \Phi(\mathbf{x}_{lq}^{1} \delta^{1}) [\Phi(\mathbf{x}_{lq}^{1} \delta^{2})/\sigma]^{-1} \\ \phi[\log(1 + GP) - \mathbf{x}_{lq}^{1} \delta^{2})/\sigma]/\sigma, & \text{if } GP > 0, \\ 1 - \Phi(\mathbf{x}_{lq}^{1} \delta^{1}), & \text{if } GP = 0, \end{cases}$$
(4)

where the top part of the equation reflects the cases in which the client accepted the quote—and, hence, the *GP* is positive—and the bottom part the client's quote-acceptance process. x_{1q}^{1} includes a dummy for whether the quote was treated or not, a set of dummy variables to control for day of the experiment fixed effect, line weight, cost per pound, and whether the quote required a cut (divided by the weight).

The results of the Cragg hurdle model analysis are shown in the left column of Table 3. Controlling for line characteristics and for day fixed effect, the effect of the treatment (i.e., providing the model's price recommendation) on the probability that the client will accept the quote is positive and statistically significant. The effect of the treatment on gross profit for the lines that were converted is not statistically significant.¹¹ Thus, the

Variable	Base me	odel	Interaction with <i>PI</i> and <i>N</i> train	
Client acceptance of price				
Treatment	0.154*	(0.076)	0.135	(0.074)
Line weight (log)	-0.0886^{***}	(0.022)	-0.0844^{***}	(0.021)
Cost per lbs.	-0.0510	(0.040)	-0.0145	(0.040)
Cut/weight	-3.338*	(1.558)	-2.788	(1.571)
Prediction interval		. ,	-1.566^{***}	(0.389)
$Treatment \times PI$			0.820	(0.485)
N train (in thousands)			0.0475*	(0.019)
Treatment × N train			-0.0218	(0.024)
Constant	0.473*	(0.191)	0.437*	(0.187)
Log line gross profit		. ,		
Treatment	0.00448	(0.009)	0.00117	(0.009)
Line weight (log)	0.115***	(0.003)	0.115***	(0.003)
Cost per lbs.	0.0386***	(0.006)	0.0375***	(0.006)
Cut/weight	1.541***	(0.222)	1.483***	(0.219)
Prediction interval		. ,	0.125***	(0.035)
$Treatment \times PI$			-0.173^{***}	(0.050)
N train (in thousands)			-0.00096	(0.002)
$Treatment \times N$ train			-0.00136	(0.0026)
Constant	0.973***	(0.022)	0.975***	(0.021)
ln <i>sigma</i>		. ,		. ,
Constant	-2.189***	(0.039)	-2.195***	(0.039)
Observations	4,142		4,142	. ,
R^2	27.99%		30.10%	

Notes. Standard errors are in parentheses. Day fixed effects are included.

p < 0.05; p < 0.001.

Variable	Base m	odel	Interaction with model price low	
$\Delta Price$	-0.938***	(0.048)	-1.575***	(0.102)
Line weight (log)	-0.284^{***}	(0.027)	-0.211^{***}	(0.026)
Cost per lb.	0.174***	(0.039)	-0.00381	(0.029)
Cut/weight	13.86***	(2.275)	24.41***	(3.414)
Model price lower			0.883***	(0.054)
Model lower $\times \Delta Price$			-1.429***	(0.109)
Constant	1.655***	(0.205)	1.926***	(0.179)
Observations	4,142		4,142	

 Table 4. Instrumental Variable Analysis for Line Conversion

Note. Standard errors are in parentheses.

***p < 0.001.

treatment worked through setting prices that increase the likelihood of the client accepting the quote, but not through setting prices that lead to higher profits given quote acceptance.¹²

To investigate the mechanism by which treatment led to increase in quote acceptance by the client, we run an instrumental variable Probit regression (Amemiya 1978) of quote acceptance on the absolute value of the difference between the model's price per pound and the final price per pound quoted to the client, with treatment as an exogenous IV for the endogenous price difference. We clustered standard errors for lines within a quote:

$$P(sale_{l} = 1) = Probit(\Delta Price_{l}\beta_{1} + x_{l}^{2}\beta_{2}), \qquad (5(a))$$

$$\Delta Price_l = I_T \pi_1 + x_l^2 \pi_2 + v_l, \qquad (5(b))$$

where *sale*_{*l*} is the client's decision to accept line *l* (in quote *q*), $\Delta Price_l$ is the absolute value of the difference between the model and the final price per pound,¹³ and x_l^2 is the same set of controls used in Equation (4). The Gaussian function for $\Delta Price_l$ includes the same set of controls x_l^2 , a treatment dummy I_T , and a random shock normally distributed, v_l .

The results of the IV analysis are shown in the left column of Table 4. As expected, the term that captures the difference between the model and final price has a negative coefficient in the quote acceptance, suggesting that when the salesperson prices closer to the model (final price is closer to model's price), client acceptance increases, and confirming that the treatment works through making the salespeople's pricing closer to their model.

In addition, we run an IV regression, in which we include an interaction term between $\Delta Price_l$ and a dummy variable for whether the model recommended a lower price than the salesperson (38.2% of cases). We find that the instrumented price difference variable is more strongly related to quote acceptance when the model recommends lower prices than the salesperson relative to when it recommends higher prices ($\beta_{\Delta Price_l \times I_{mdd_light}} = -1.43, p < 0.001$; right column of Table 4), suggesting that the model affects quote acceptance by recommending lower prices to the salespeople.

4.3. Explaining the Automation Mechanism

Although we have now established that following the model's recommendation leads to higher profitability (through increased acceptance), a question may arise: How does a model that merely mimics the salesperson's pricing policy, rather than provides optimal prices, lead to better outcomes? As mentioned earlier, the performance of the automation approach relies on a trade-off between improved performance due to increased consistency (Meehl 1954) and possible deteriorated performance due to missing relevant noncodable cues (see also Equation (1)).

A plethora of previous work (e.g., Dawes et al. 1989) suggests that systematically applying the expert's decision policy will lead to better predictions by mere consistency of reapplication of the expert's judgment. Consistent with this account, we find that the coefficient of variation of the model's predicted price margins (0.372) is significantly smaller than that of the salespeople's price margins (0.432; p < 0.001). That is, the model leads to lower variance in the pricing decisions by eliminating ϵ_t , the random error component in the price, and possibly some casebase information in $f(z_t)$ in Equation (1). Taken together with the increased profitability in the treatment condition and the IV analysis that shows that the treatment effect works through salespeople pricing closer to their model, we conclude that the model's benefit in reducing the noise component of the pricing decision outweighs the model's possible loss due to not accounting for private information. In this subsection, we demonstrate how the consistency of the model may have helped salespeople's decision making in the experiment. We start by showing how the model's recommendations make salespeople more consistent and how the model helps correct intertemporal biases. We then provide a compliance analysis that, although not causal, sheds light on the trade-off between consistency and missed private information and how the treatment effect is materialized.

4.3.1. Heterogeneity in Treatment Effect

4.3.1.1. Prediction Intervals. We would expect the model to perform better and help salespeople in

situations where the model makes more accurate predictions. For example, when orders are complex or odd, the model has less relevant data to calibrate on, and, hence, predictions are likely to be less accurate and less helpful. To investigate this conjecture, we calculate prediction intervals (PIs) for each of the model's pricemargin recommendations, based on each salesperson's own data. We then include these mean-centered PIs as heterogeneity in treatment effect in x_{lq}^1 in Equation (4) of the Cragg analysis. Prediction intervals, by definition, are larger when model covariates are extreme, and, thus, the model's prediction is less certain, leading to larger intervals and weaker treatment effects. To control for the fact that PIs may be negatively correlated with the amount of data available, we include in the regression the mean-centered number of pricing decision (lines) for the salesperson during the calibration period.

Indeed, we find that when PIs are large, and controlling for the number of pricing decisions made by the salesperson in the calibration period, the treatment is weaker, and gross profit is lower while there is no effect for conversion (see right column in Table 3). The joint effect of the PIs' width and treatment is possibly driven by compliance, as salespeople were less likely to comply with the model when the model was unsure about the pricing (large PIs). We find a weak negative relationship between the size of the PIs and compliance with the model's recommendation (*Pearson'sr* = -0.05, *p* = 0.011), meaning that salespeople do have an insight in terms of when to comply with the model.

4.3.1.2. Salesperson Characteristics. Theoretically, it would be informative to investigate heterogeneity in treatment effect by salesperson characteristics, such as consistency of past pricing decisions, expertise, or tenure with the company. However, with only 17 salespeople in the experiment, such analyses can be suggestive at best. Directionally, we find that salespeople for whom our model of the salesperson had a higher coefficient of determination, R^2 (i.e., salespeople with an average lower ϵ_t and $f(z_t)$ in Equation (1)), had lower treatment effect (lower increase in quote acceptance and GPs due to treatment). That is, the model contributes less to salespeople who are consistent and use primarily observables to make their pricing decisions. In addition, we verify that the number of observations in the salesperson's training set does not drive the relationship between R^2 and the treatment effect. See Online Appendix 2.4 for full details.

4.3.2. Compliance Analysis. One of the largest risks when conducting a field experiment that requires cooperation of participants is lack of compliance. Specifically, when offered to rely on algorithmic decision aids, people may demonstrate *algorithm aversion* and limit

their use of the aid tool. Among the reasons for this aversion are the belief that humans can reach nearperfection in decision making (Einhorn 1986) and that human predictions improve through experience (Highhouse 2008). The latter is especially important when it comes to experts' decision making. Experts tend to overweigh their experience and expertise, which often leads to poor predictability (Arkes et al. 1986, Camerer and Johnson 1991). Moreover, when facing (inevitable) algorithmic errors, people are less likely to trust and use the algorithm (Dietvorst et al. 2015).

The experimental design, in which salespeople received the model's prices as recommendations and could use it at their own discretion, posed a risk of low compliance to our experiment. During the experiment, salespeople expressed great confidence in their own judgment. For example, one salesperson said, "I am not likely to follow the recommended price because I had already put a lot of thought into pricing the quote and considered everything there is to consider." Moreover, many salespeople said that although the tool may be useful for other salespeople, "their clients" (or the quotes they typically price) are "different." Overall, salespeople's reluctance to accept the model's price could make it harder to identify the true effect of the treatment.

Our experimental design allows us to assess compliance because we have information about the salesperson's original pricing decision made prior to exposure to treatment (only after the salesperson inputs into the system a price for the new quote, the quote is randomly assigned to treatment or control, and the model's price is displayed for treated quotes). In what follows, we analyze compliance patterns to shed more light on the observed treatment effects. However, because compliance is endogenous to the decision maker and to the quote and client characteristics, the analysis in this section should be taken as descriptive, rather than causal.

Table 5 depicts the compliance patterns based on whether salespeople changed their price, relative to their original price, in a direction that is consistent or inconsistent with the model's recommendation.¹⁴ First, looking at the control condition, we find that salespeople have an insight into adjusting their price in the right direction. In the control condition, salespeople did not see the model's recommendation, yet, upon further thinking about their price after receiving the control email prompt, they inherently adjusted prices in higher rates in the direction of the model than in the opposite direction (9.52% price decreases versus 6.25% price increases when the model recommended a lower price than the salesperson price; 16.93% price increases versus 3.78% price decreases when the model recommended a higher price). Overall, in the control condition, salespeople adjusted their original price in the direction of the model in 14.05% of the cases (64 price decreases and 179 price increases). The magnitude of change in price (see median change in price in Table 5) in the treatment condition is also higher when salespeople complied with the model versus when they went against the model. This result highlights that when reconsidering their prices, even without seeing the model price, salespeople had the intuition to revise their pricing decisions according to the model. It also highlights that the model predictions are in line with the salespeople's decision-making process. Turning now to the treatment condition, we see an even higher rate of "compliance" with the model's recommendation. In 19.48% of the cases (133 price decreases and 337 price increases; see boldface numbers), salespeople changed their price in the direction of the model, a lift of 37.7% in compliance with the model relative to the control condition (5.4 percentage points).¹⁵

Moving to the relationship between compliance patterns and demand, Table 6 shows quote conversion rates by model recommendation and salesperson behavior. Cases in which the salesperson changed the price in a direction congruent with the model's recommendation are in boldface. As expected, and in line with the results of the Cragg analysis, the largest increase in conversion, from 25% in control to 50.38% in treatment, comes from following the model in decreasing the price (p < 0.001). When increasing the price following the model's recommendation, we do not expect an increase in conversion because the price was increased (39.17% in treatment versus 37.99% in control, *p* = 0.79).

The off-diagonal, in which salespeople went against the recommendation of the model, also reveals an interesting pattern. In these cases, we find significantly higher conversion rates in treatment relative to control. Because of selfselection, we can only speculate that these are cases in which salespeople had relevant private information about the client's likelihood to accept the quote and responded accordingly $(f(z_t)$ in Equation (1)). When reconsidering their pricing decisions following the email prompt, the salesperson may have re-evaluated the information and,

in some cases, may have discussed the case with a manager, possibly making the salesperson weigh the information differently and consequently deviate from the model even more. Of course, had this information z_t been codable, it could have been incorporated in the model to improve predictions. We also investigated how gross profits vary by compliance. Consistent with the insignificant effect of treatment on gross profits, given quote conversion in the Cragg model, we do not find significant differences in gross profits by compliance (see Online Appendix 2.4 for details).

One of the reasons suggested by the JB literature for why a model of the expert improves the expert's decision making is that it helps the expert avoid intertemporal biases due to, for example, reacting to previous successes in independent decisions (Coval and Shumway 2005). In the context of compliance, salespeople may be more likely to comply with the model if they recently observed the model's success. To investigate such patterns, we estimate a mixed logit model regressing salesperson compliance per the definition of compliance in Table 5 on a dummy indicating whether the previous quote by the salesperson was converted, a dummy indicating whether the previous quote by the salesperson was a treatment quote, and the interaction between the treatment and conversion dummies.

We find that success in converting a control quote leads to lower compliance rates in the following treatment quote ($\beta_{prev_conversion} = -0.97, p = 0.001$; see Table 7): The salesperson did not use the model and succeeded, and, therefore, is not inclined to use it for the current quote. However, following a treatment quote that was converted ($\beta_{conversion \times treatment} = 0.793, p = 0.026$), salespeople are more likely to follow the model, having observed the usefulness of the model. Indeed, overconfidence was found to be prevalent among salespeople (Bonney et al. 2016), but automation in the form of AI recommendations can mitigate this bias as salespeople realize its benefits.

Salesperson's behavior

Increased price

44 (\$0.09)

4.84%

337 (\$0.4)

22.41%

381

15.79%

42 (\$0.39)

6.25%

179 (\$0.44)

16.93%

Total

909

100%

1,504

100%

2,413

100%

672

100%

1.057

100%

No change

732 (-)

80.53%

1,110 (-)

73.80%

1,842

76.34%

566 (-)

84.23%

838 (-)

79.28%

able 5.	Compliance	Patterns l	bv N	/lodel	Recomme	ndation
	compnance	1 autorito i	U Y 1 V	iouci	neconnic.	ilaalion

Model's recommendation

Decrease price

Increase price

Total

Decrease price

Increase price

T (1

Experimental condition

Treatment

Control

Total	104	1,404	221	1,729
	6.02%	81.20%	12.78%	100%
<i>Notes.</i> Median changes in price per condition are in parenthes are in boldface.	es. Cases in which the sa	alesperson complied w	vith the model's recon	nmendation

Decreased price

133 (-\$0.2)

14.63%

57 (-\$0.13)

3.79%

190

7.78%

64 (-\$0.24)

9.52%

40 (-\$0.99)

3.78%

		Salesperson's behavior				
Experimental condition	Model's recommendation	Decreased price	No change	Increased price	Total	
Treatment	Decrease price	50.38 %	48.36%	52.27%	48.84%	
	Increase price	59.65%	54.23%	39.17 %	51.06%	
	Total	53.16%	51.90%	40.68%	50.23%	
Control	Decrease price	25.00%	42.40%	26.19%	39.73%	
	Increase price	42.50%	47.37%	37.99 %	45.60%	
	Total	31.73%	45.37%	35.75%	43.32%	

Table 6. Conversion Rates by Compliance with Model Recommendation

Note. Cases in which the salesperson complied with the model's recommendation are in boldface.

Overall, our analyses suggest a moderate level of compliance, which leads to higher quote conversion, particularly when the model recommended to decrease the price. When salespeople decide to go against the model's price, it is often when the quote has a higher chance of conversion, hinting toward the existence of noncodable private information. Finally, we find that observing the model's success in converting quotes can mitigate salespeople's overconfidence with their own behavior and establish confidence with the model's advice.

5. Counterfactual Analyses

The experiment allowed us to directly investigate the causal effect of automation on profitability. However, as with any field experiment, there were some limitations to it, including not being able to fully replace salespeople's pricing due to the high stakes for the company, being dependent on salespeople's endogenous compliance, and not being able to test different specifications of the pricing automation model. To overcome these issues, we build a demand model that mimics the client's decision to accept or reject the quote, given the quoted price. We then run a counterfactual analysis comparing profitability under different pricing schemes based on versions of full and hybrid automation.

Using observed historical client decisions to accept or reject quoted prices, we estimate a demand model.¹⁶ To create the counterfactual, for each quote *q* requested by client *i*, based on observed prices p_{qi} and predicted prices \hat{p}_{qi} (calculated based on the model's predicted margins), we calculate predicted acceptance probabilities, based on the actual price, $Pr(p_{qi})$, and the model's price, $Pr(\hat{p}_{qi})$. We then calculate for quote *q* requested by client *i*:

$$\Pi_{qi} = (p_{qi} - c_q) \times Pr_{qi}(p_{qi}), \tag{6}$$

$$\hat{\Pi}_{qi} = (\hat{p}_{ai} - c_q) \times Pr_{qi}(\hat{p}_{ai}), \tag{7}$$

and compare expected profits based on the difference between Π_{qi} and $\hat{\Pi}_{qi}$.

5.1. Data for Counterfactuals

To allow for both calibration and validation data for the counterfactual, we use a longer data period that spans two years of transactions between 2015 and 2016, using

the first 18 months for calibration and the last 6 months for validation (prediction). Overall, the calibration data include 21 salespeople making 104,336 pricing decisions for 3,787 clients over the course of 18 months. See Online Appendix 3.1 for summary statistics of the counterfactual analyses data. Similar to Equation (3), we estimate a price-margins model using counterfactual calibration data. The estimated price-margins model is very similar to the pricing model in Table 2. See Online Appendix 3.1 for details.

5.2. The Demand Model

To calculate expected profits, we need to estimate the probability of quote acceptance, given price (the last term in Equations (6) and (7)). Given the salesperson price quote, the client decides whether to accept or reject the offer. For each client, we observe multiple quote requests and the corresponding accept or reject decisions. We specify the utility for client *i* from accepting quote *q* as:

$$u_{qi} = \beta_{1i} + \beta_{2i} \ p_{qi} + \boldsymbol{\beta}_z \ \boldsymbol{z}_{qi} + \gamma \ \Delta P_{qi} + \sigma \ \eta_{qi} + \xi_{2qi},$$
(8)

where β_{1i} is a random intercept for client *i*, and p_{qi} is the price per pound offered for quote *q* made by client *i*.¹⁷ z_{qi} is a vector of covariates that includes recency (days since the last quote request by client *i*), regular salesperson (the ratio of quotes priced by the salesperson out of the total number of quotes by this client up to the date of the current quote), log weight of quote *j*, LME price on the day of quote *j*, LME volatility on the week prior

 Table 7. Compliance by Conversion and Treatment

Variable	Coefficient	Std. err.
Previous quote conversion	-0.970***	(0.292)
Previous quote treatment	0.0713	(0.222)
<i>Prev. quote: conversion</i> \times <i>treatment</i>	0.793*	(0.356)
Constant	-2.569***	(0.334)
$ln(\sigma)$	1.442***	(0.181)
Observations	2,406 ^a	· · · ·

Note. Day fixed effect and client random effect are included. ^aLines in the treatment condition.

p < 0.05; p < 0.001.

to quote *j*, and a set of dummies, one for each product category included in the quote.

To control for possible endogeneity of the price due to either targeted pricing for specific clients or unobserved random shocks that may affect both pricing and demand, we use a control function approach (Petrin and Train 2010) with wholesale cost, cut, and quarter fixed effect as instrumental variables, as well as client random effects to control for potential endogenous effect in targeting prices to clients based on their estimated likelihood to accept. The cost that the company paid for the product is a good exclusion instrumental variable for price, as its effect on clients' demand should primarily go through the price of the product. Given that the cost is determined based on the price the company paid when buying the product, and the fact that products tend to stay in the company's warehouse for as long as six months, correlation between wholesale price and competitive prices is likely to be relatively low. To further test the validity of this instrument, we run the Hausman specification test adapted for control function estimation (Hausman and McFadden 1984) for our main IV, cost. The test suggests validity of instrumental variable approach (Chi-Squared = 18.26, p < 0.001).

The Gaussian control function price equation for client *i* and quote *q* is:

$$p_{qi} = \lambda_i + \lambda_{cost} \cos t_q + \lambda_{cut} \cot t_q + \lambda_{quarter} \ quarter_q + \xi_{1qi},$$
(9)

where p_{qi} is the actual price quoted to client *i* for quote q, λ_i is a client *i* random-effect intercept, $cost_q$ is the cost of the material for quote q, cut_q is the ratio of lines in the quote that require special processing, and $quarter_q$ is a set of dummy variables for five out of the six quarters in the calibration data. ξ_{1qi} is a random shock normally distributed with a zero mean and a variance σ_{1q} . The last two terms prior to the random shock ξ_{2qi} in Equation (8) reflect the specification of the control function approach. $\Delta P_{qi} = p_{qi} - \tilde{p_{qi}}$ is the residual of the control function price equation, where $\tilde{p_{qi}}$ is the fitted value of Equation (9) for the specific values of quote *j*, and η_{qi} is independent and identically distributed standard normal.

Finally, assuming that ξ_{2qi} is extreme value distributed, the probability that client *i* will accept quote *q*, Pr_{qi} , follows the binary logit specification.

We estimate the demand model in two stages. First, we estimate a control function random-effects model to estimate $\Delta P_{qi} = p_{qi} - \tilde{p_{qi}}$; then, we use Hamiltonian Monte Carlo (HMC) with No U-turn sampler (NUTS) to estimate the demand model. Online Appendix 3.2.1 includes the full details of the demand model estimation and results. In what follows, we use results from the demand model estimation to calculate the profit counterfactuals.

5.3. Profits of Model Pricing versus Profits of Salesperson Pricing

Using the price-margins model (Equation (3)) together with the demand model that predicts the client's acceptance behavior, we can compare expected profits based on the model-of-the-salesperson predicted prices and based on salespeople's prices (Equations (6) and (7)). We use the holdout sample of the last six months of the data, which were not used in estimating the demand or the pricing models, with a total of 11,621 quotes to assess the performance of the demand model. The expected acceptance probability based on the original pricing scheme, 61.1%, is comparable to the actual observed acceptance probability, 59.3%. The demand model's hit rate for out-of-sample quotes is 69.9% overall, pointing to a fairly good aggregate demand model accuracy.

Using Equations (6) and (7) and aggregating across quotes, we find that the model's prices generate expected profits that are 4.9% higher than those of the salespeople's prices ($\Pi[\hat{p}] = \$2,536,058$ compared with $\Pi[p] = \$2,417,149$). This difference is statistically significant, based on the 95% posterior confidence intervals (PCIs) across a sample of 100 draws from the output of the HMC sampler. The actual profits for the same set of quotes are \$2,345,479.¹⁸

Thus, consistent with the experimental results, but now fully replacing the salespeople with their bootstrap model, the counterfactual analysis demonstrates that the model of the salesperson generates higher profits than the salesperson. This should not be taken for granted because when fully automating the salesperson, we completely eliminate the human response to private information $f(z_t)$, yet the model's consistency achieved by eliminating ϵ_t overcomes this possible loss. A possible concern with the counterfactual analysis is that our control function accounts for aggregate correlated random shocks, like targeting specific clients, but possibly not for quote-specific random shocks that may affect both the pricing decisions and the clients' demand. To examine the risk of such random shock interference, we compare the predicted profits, based on the demand model, for the actual prices with the actual profits. If a meaningful positive correlated random shock exists, the demand model at real prices should show higher profitability than the observed profits, as salespeople would price higher when they expect a conversion. We find a very small, but negative, difference between the predicted profitability at actual prices $(\Pi[p] = \$109.12 \text{ per line})$ and the actual profitability ($\Pi =$ \$103.99 per line). Thus, to the extent that a correlated random shock exists, it is likely to be small and negative. We conduct additional simulation analyses to assess the magnitude of the random shock in our data and find that to the extent that a correlated random shock exists, it is likely to be small (see Online Appendix

are statistically significant based on the 95% PCIs. Thus, we find that the human-machine hybrid scheme, in which the majority of the quotes are priced by the model, leads to higher profits than the two extreme cases (full automation or no automation). This raises the question of which quotes should be allocated to the model and which to the salesperson, which we address next.

5.4.1. Understanding the Machine Learning Hybrid. The RF algorithm is a "black box" nonlinear prediction tool. We analyze the feature importance of the RF algorithm to shed light on which quote and client characteristics determine the allocation of quotes to the salesperson or the model. Quote characteristics, such as weight, cost, cut, and number of lines per quote, as well as client characteristics, such as RFM and whether the salesperson is the regular salesperson, all affect the quote allocation decision (see Online Appendix 4.2 for a full list of feature importance).

Additionally, we run a mixed linear regression for the difference in expected profits between the salesperson and the model using the same variables used in the RF model (see details of the model in Online Appendix 4.3). The results of this analysis, shown in Table 8, indicate that salespeople generate higher profits than their bootstrap model when cost per pound and quote weight have extreme values, either very small or very large. In addition, consistent with the RF feature importance, salespeople generate higher profits than the model when there are multiple lines in the quote. These results suggest that salespeople are more successful in pricing out of the ordinary quotes with special features, and the RF

Table 8. Mixed Linear Regression of Expected Profits

 Difference (Salesperson Profits Minus Model Profits)

Variable	Coefficient	Std. err.
Weighted cost per lb.	-16.46***	(4.589)
Weighted cost per lb. squared	1.164**	(0.445)
Quote weight (log)	-60.08***	(7.486)
Quote weight (log) squared	5.054***	(0.693)
LME per lb.	-164.6*	(67.908)
LME per volatility	7.077	(3.888)
Lines per quote	5.882***	(1.060)
Regular salesperson	8.573	(6.748)
Cut ratio	8.477	(4.963)
Quote recency	-0.726	(0.897)
Quote frequency	-7.166	(6.210)
Quote monetary	2.434	(1.267)
FT base ratio	-4.806	(8.783)
Constant	261.9***	(57.546)
Observations	5,829 ^a	. ,

Notes. Regression includes client priority, product category, and salesperson fixed effects. Regression includes client random effects. ^aBased on Quarter 6 that was used for the RF training.

3.3 for details). That being said, because we cannot fully rule out the possibility of a correlated random shock, we use the counterfactual analysis as complement to the field experiment, rather than as direct evidence for the performance of automation in this context.

5.4. The Human-Machine Hybrid Approach

Allowing all quotes to be priced by the salesperson (as in the current practice in the firm and the control condition in our experiment) or all quotes to be priced by the model (as we did in the in the previous section) are two extremes on the continuum of human-machine hybrid automation. The treatment condition in the experiment demonstrates a hybrid structure in which salespeople receive the model's price recommendation and decide whether to accept or reject it. However, in light of the relatively low compliance rates observed in the experiment, it is not clear whether the salesperson's judgment on when to comply with the model's pricing was optimal. On the one hand, allowing salespeople in the experiment to make the judgment of when to use the model's price might have led to low compliance rates, which possibly limited the possible treatment effect. On the other hand, salespeople may have decided to forgo the model's prices when they had valuable private information that the model was missing.

Ideally, the company would be able to identify and allocate out-of-the-ordinary quotes to human pricing as they come in, while automatically pricing plain-vanilla quotes by the model. In order to automatically allocate to either the salesperson or the model, we train a machine learning Random Forest model that predicts which of the two will generate higher expected profits based on each quote's and client's characteristics. The dependent variable for the RF model is the difference in expected profits between the salesperson and the model based on the demand model described in Section 5.2. As independent variables, we include the quote and client characteristics used in the price-margins model (see details in Online Appendix 4). We then predict the difference in expected profits between the salesperson and the model for each of 11,621 quotes in the validation period (Quarters 3 and 4 of 2016) and allocate the quote to the pricing scheme (salesperson pricing or model pricing) that is expected to generate higher profits. Overall, the RF hybrid allocates 68% of the quotes to model pricing, a significantly higher level of automation compared with the compliance levels observed in the experiment, possible evidence of algorithm aversion among salespeople in our experiment.

Based on the validation period, we find that the total expected profits of the machine learning RF hybrid are 7.8% higher than those of the salespeople ($\Pi[p_{ML_hyb}]$ = 2,606,208 versus $\Pi[p]$ = \$2,417,149) and 3.1% higher than those of the full automation counterfactual ($\Pi[\hat{p}]$ = \$2,536,058). The differences between the profits of the RF hybrid and the salesperson or the model profits

p < 0.05; p < 0.01; p < 0.01; p < 0.001.

allocation model captures these patterns. Importantly, the superiority of the hybrid model suggests that there is merit in using the salesperson's judgment, $f(z_t)$, in some cases, while eliminating noise, ϵ_t , via automation in others. It should be noted that if the private signal, z_t is not correlated with any observables, then the hybrid will not be able to identify and redirect the quote to the salesperson.

6. General Discussion

Algorithmic pricing transformed the way sellers set prices and, in some domains, mainly in business-toconsumer context, almost fully replaced human pricing. Yet, in some cases, algorithmic pricing can lead to extreme failures (e.g., when the price of a book in Amazon peaked to \$24 million¹⁹ or when Delta Airlines was accused of price gouging during Hurricane Irma²⁰).

The B2B market lags behind the B2C market in adopting automation (Asare et al. 2016). To a large extent, pricing processes in B2B still rely on human labor, and soft skills such as communication or salesmanship are believed to be essential to B2B sales. In this paper, we examine whether in high human-relationship environments, such as B2B pricing, in which salespeople provide individual price quotes to customers, models can assist, or even replace, human pricing. Using a multimethod approach that combines a field experiment, in which we embed AI-based algorithmic pricing into the CRM system of a B2B retailer, and econometric modeling for counterfactual analysis, we demonstrate that pricing decisions in B2B settings can be automated by modeling the salesperson and reapplying their pricing policy automatically to new pricing decisions. In our field experiment, providing salespeople with automated price recommendation in real time led to an 11% increase in profits to the company. Additional counterfactual analyses show that because B2B pricing involves a high degree of soft skills and salesmanship expertise, a hybrid model that prices incoming quotes most of the time, but allows the salesperson to price complex or irregular quotes, performs better than either full automated or pure human pricing. The hybrid approach uses the model's scalability and consistency for most pricing decisions and human judgment for out-of-theordinary cases that possibly involve noncodeable information. Such an approach allows one to mitigate extreme algorithmic pricing failures, as the one described above.

We propose a two-step machine learning hybrid automation approach. In the first step, the model automates the allocation of incoming quotes to the salesperson or to the model by predicting who, the salesperson or the model, will generate higher profits and allocates each quote accordingly. In the second step, the model automates the pricing decision itself if the quote was allocated to the model. The human-machine hybrid performs significantly better than pure model pricing in generating profits to the company, with an increase of over 7.8% in profits over pure human pricing. By using machine learning to automatically identify who should price the quote, we lay the grounds for a hybrid automation solution that utilizes the benefits of automation in overcoming intertemporal human biases, but preserves human expertise and experience gained by salespeople in the company over time. Our findings provide empirical evidence in the context of B2B pricing to the idea discussed in labor economics that although automation can substitute for predictable and rule-based human labor, it can only complement human labor that is largely based on social and emotional skills (Autor et al. 2003, Autor 2015).

Our research bridges between the behavioral judgment literature and marketing science literature by building a pricing judgmental bootstrapping model (Dawes 1971) and demonstrating, using both a field experiment and econometric modeling, how such a model could be applied in real-world settings to address a major business problem. The performance of judgmental bootstrapping has rarely been tested in repeated business decision making and in settings where the expert has access to richer information than the JB model, information that can arguably lead to superior decision making on the expert's end. Moreover, our research bridges theory and practice, by demonstrating via a pricing field experiment how automation can improve the profitability of a B2B retailer. Indeed, following our experiment, the B2B retailer with which we collaborated is adding our pricing model to its CRM system to provide price recommendations to salespeople for all incoming quotes. In the longer term, and based on our work, the firm is considering using our hybrid model to move to an online sales process, which automates both the prices presented to clients online and the decision of whether to present an online price or a "call an agent" message. We call for future research to further explore these two degrees of automation in both B2B and B2C contexts, such as hospitality and services.

In our empirical application, we find that using a linear judgmental bootstrapping to "teach" the model how to price works better than more advanced machine learning models of the salespeople. An advantage of the linear model is its simplicity, which is particularly important, given that the company will need to occasionally rerun the model to update model parameters. Nevertheless, we encourage future research to explore the performance of machine learning in automating human decision making in other contexts. Additionally, we encourage future research to explore automation using profit-maximizing prices, as opposed to a judgmental bootstrapping approach that mimics the expert. Such automation would need to make assumptions about demand and is likely to be more complicated for the firm to routinely estimate and optimize.

Using a hybrid automation approach that complements the salesperson with a model of themselves can have far-reaching implications for preserving organizational knowledge in a work environment characterized by high salesforce turnover rates, especially in light of the COVID-19-driven great resignation.²¹ Salespeople develop expertise and familiarity not only with the product they sell, but also with their regular clients. By learning the salesperson's pricing policy and applying it automatically, the tool serves not only as a pricing aid, but also as a knowledge-management mechanism, a means to preserve organizational knowledge and specific expertise within the organization, and to mitigate losses in case of salesforce turnover (Shi et al. 2017). Conversations with salespeople in the company echo the benefits of the approach. For example, one salesperson commented during the course of the experiment: "When I am not in the office, other salespeople can use my tool's recommendations to price my quotes. Currently they are not willing to take my quotes because it takes them too long to price them, so I am losing business when I am not here." Future research could further explore the use of automation to preserve organizational knowledge and mitigate the negative consequences of personnel turnover and absences. Such research could have significant implications for individuals who require flexible employment structures, such as parents or individuals with special needs. As the world is moving toward more flexible and employee-centric employment structures, hybrid modes of automation, such as the one we explore in this research, may become even more useful.

Our analysis explored the potential of automation in B2B salesforce pricing decisions using a field experiment and secondary data from a metal B2B retailer. Future research could explore the generalizability of these findings to other B2B retail domains and to other managerial decision making. Potential applications include other retail environments, such as building supplies (Bruno et al. 2012), or special expertise in B2B services, such as consulting, legal services, or architectural services. The degree to which the hybrid model would fit such environments and the share of transactions that should be allocated to automation would depend on how structured the transactions are and how prevalent and prominent case-specific information is for making decisions in each context. The less structured and codable the decision-making process is to begin with, the lower the expected contribution of direct automation. That being said, our hybrid automation approach can flexibly accommodate different levels of automation that are appropriate for each domain, given that it keeps the human decision maker in the process.

One limitation of our field experiment was the relatively low compliance of the salespeople with the tool, which possibly underestimates the potential effect of automation. People, and especially experts, are often averse to using algorithms to aid them in decision making (Arkes et al. 1986, Camerer and Johnson 1991). Compliance may limit the effectiveness of any tool that relies on experts' willingness to use it. Specifically, if a hybrid approach is adopted and usage is in the discretion of the expert, the approach's effectiveness will depend on compliance patterns. We postulate that a bootstrap-type model is likely to facilitate higher compliance rates relative to a normative model because it mimics the salesperson's behavior, as opposed to some optimal algorithmic behavior. Future research could further explore the role of compliance in automation in general and in hybrid automation in particular.

In summary, our research provides first empirical evidence of the potential of automating the humanintensive work of the B2B salesforce. It suggests that although the B2B salesperson is traditionally perceived as indispensable, some sales tasks could be automated. By automating parts of the pricing task, the company could not only reduce costs associated with maintaining its sales team, but also increase profitability due to better-quality pricing decisions. Moreover, we show that the decision of when to use human expert pricing to override the model could, in and of itself, be automated. We hope this research will spark further investigation of this promising direction.

Acknowledgments

All authors certify that they have no affiliations with or involvement in any organization or entity with any financial interest or nonfinancial interest in the subject matter or materials discussed in this manuscript. The authors have no funding to report.

Endnotes

¹ Shipping costs are priced separately as an additional line in the quote. We do not model those costs.

² We removed from this analysis clients that had only one quote, and, hence, do not allow estimating a reliable pricing model; clients defined by the company as either contractual or semicontractual; and rare cases of lines with missing or negative price or cost. Additionally, and following the company's recommendation, we removed orders of over 8,000 pounds or orders at the bottom 1% of orders by weight. Such orders are treated differently by the company and are often priced by a manager or follow predefined rules.

³ A small number of SKUs are not stocked and priced by weight, but by length. We later account for that in the pricing model.

⁴ Priority A is the highest for clients with order volume of at least \$100,000, and priority E is the lowest for clients with spending of less than \$5,000 in the past 12 months. Priority P is given to clients with "E" orders volume that have a potential (judged by the management) to become high-priority clients.

⁵ In the calculation of RFM measures, we include quotes that were not converted to sales, under the assumption that the client decided to purchase the product somewhere else. To initialize the recency and monetary variables, if the client purchased before January 2016, we use the last purchase prior to January 2016. If the client is a new

client, we dropped the first purchase from this analysis and used it to initialize these variables.

⁶ Table A2 in Online Appendix 1 reports average estimates across the individual-salesperson regressions.

⁷ The company has a minimum price of \$30 per line for high-priority clients and \$150 per line for low-priority clients. If the model's calculation resulted in a total price lower than that minimum, we adjusted the price per pound and the total per line to reflect the minimum price.

⁸ Because of the relatively small number of salespeople in the company (17 salespeople at the time), randomization was done at the quote level rather than at the salesperson level. We intentionally overweighted treatment over control with anticipation of low compliance rates.

⁹ We excluded from the analysis approximately 10% of the lines with cost or price per pound larger than \$16 that often relate to irregular orders, as well as lines for which the final profit margin was negative. The results reported in Section 4.2 are robust to including these data points.

¹⁰ A Tobit II analysis would not be appropriate to separate the effect of treatment on acceptance and profits because the data are not left truncated. Not observing gross profits occurs due to client rejection of the quote and not due to truncation of the firm's profits to the negative domain.

¹¹ Note that one could not interpret the parameters of the profit equation in the Cragg hurdle model as causal, as it is conditional on quote acceptance, which is endogenous.

¹² We find similar results when running the Cragg hurdle analysis on the treatment variable without the control variables in x_{la}^1 .

¹³ Running the IV analysis using, instead of absolute difference, the absolute percentage difference between the final price and the model price yields similar results. See Table A5 in the online appendix.

¹⁴ Note that this measure of compliance based on a price change is conservative because in some of the cases in which the salesperson did not change their price, the model recommended a price that is very similar to that of the salesperson.

¹⁵ Over the eight days of the experiment, we find a slight increase in compliance (r = 0.057, p = 0.0053), driven by increased compliance in the cases in which the model recommended a higher price (r = 0.069, p = 0.0007) with no change in compliance over time when the model recommends lower prices (r = -0.0064, p = 0.753).

¹⁶ Demand is estimated at the quote level, rather than the line level, because only in about 5% of the quotes, clients partially accepted the quote (i.e., the client accepted the price only for some of the lines in the quote). In the analysis, we handle these rare quotes as two separate quotes: one accepted and one rejected.

¹⁷ Estimating the demand model with reference prices instead of price yields similar results.

¹⁸ In addition to the linear JB pricing model, we test the robustness of our results to alternative machine-learning-based pricing models and find similar results. The alternative analyses are available in Online Appendix 3.4.

¹⁹ See https://www.wired.com/2011/04/amazon-flies-24-million/.

²⁰ See https://www.nytimes.com/2017/09/17/travel/price-goug ing-hurricane-irma-airlines.html.

²¹ See https://www.xactlycorp.com/blog/great-resignations-impactsales.

References

Amemiya T (1978) The estimation of a simultaneous equation generalized probit model. *Econometrica* 46(5):1193–1205.

- Arkes HR, Dawes RM, Christensen C (1986) Factors influencing the use of a decision rule in a probabilistic task. Organ. Behav. Human Decision Processes 37(1):93–110.
- Armstrong JS (2001) Judgmental bootstrapping: Inferring experts' rules for forecasting. Armstrong, JS, ed. Principles of Forecasting, International Series in Operations Research & Management Science, volume 30 (Springer, Boston), 171–192.
- Asare AK, Brashear-Alejandro TG, Kang J (2016) B2B technology adoption in customer driven supply chains. J. Bus. Indust. Marketing 31(1):1–12.
- Ashton AH, Ashton RH, Davis MN (1994) White-collar robotics: Levering managerial decision making. *Calif. Management Rev.* 37(1):83–109.
- Autor DH (2015) Why are there still so many jobs? The history and future of workplace automation. J. Econom. Perspect. 29(3):3–30.
- Autor DH, Levy F, Murnane RJ (2003) The skill content of recent technological change: An empirical exploration. *Quart. J. Econom.* 118(4): 1279–1333.
- Batchelor R, Kwan TY (2007) Judgemental bootstrapping of technical traders in the bond market. Internat. J. Forecasting 23(3):427–445.
- Bonney L, Plouffe CR, Brady M (2016) Investigations of sales representatives' valuation of options. J. Acad. Marketing Sci. 44(2):135–150.
- Bowman EH (1963) Consistency and optimality in managerial decision making. *Management Sci.* 9(2):310–321.
- Bruno HA, Che H, Dutta S (2012) Role of reference price on price and quantity: Insights from business-to-business markets. J. Marketing Res. 49(5):640–654.
- Brynjolfsson E, McAfee A (2012) Race Against the Machine: How the Digital Revolution Is Accelerating Innovation, Driving Productivity, and Irreversibly Transforming Employment and the Economy (Digital Frontier Press, Lexington, MA).
- Camerer CF, Johnson EJ (1991) The process-performance paradox in expert judgment: How can experts know so much and predict so badly? Ericsson KA, Smith J, eds. *Toward a General Theory of Expertise: Prospects and Limits* (Cambridge University Press, Cambridge, UK), 195–217.
- Chilamkurthy S, Ghosh R, Tanamala S, Biviji M, Campeau NG, Venugopal VK, Mahajan V, Rao P, Warier P (2018) Deep learning algorithms for detection of critical findings in head CT scans: A retrospective study. *Lancet* 392(10162):2388–2396.
- Chui M, Manyika J, Miremadi M (2016) Where machines could replace humans—And where they can't (yet). *McKinsey Quart.* (July 8), https://www.mckinsey.com/capabilities/mckinsey-digital/ourinsights/where-machines-could-replace-humans-and-where-theycant-yet.
- Coval JD, Shumway T (2005) Do behavioral biases affect prices? J. Finance 60(1):1–34.
- Cowgill B (2018) Bias and productivity in humans and algorithms: Theory and evidence from resume screening. Working paper, Columbia Business School, New York.
- Cragg JG (1971) Some statistical models for limited dependent variables with application to the demand for durable goods. *Econometrica* 39(5):829–844.
- Datta S, Satten GA (2005) Rank-sum tests for clustered data. J. Amer. Statist. Assoc. 100(471):908–915.
- Dawes RM (1971) A case study of graduate admissions: Application of three principles of human decision making. *Amer. Psych.* 26(2):180–188.
- Dawes RM (1979) The robust beauty of improper linear models in decision making. Amer. Psych. 34(7):571–582.
- Dawes RM, Faust D, Meehl PE (1989) Clinical vs. actuarial judgment. Science 243(4899):1668–1674.
- Deming DJ (2017) The growing importance of social skills in the labor market. Quart. J. Econom. 132(4):1593–1640.
- Dietvorst BJ, Simmons JP, Massey C (2015) Algorithm aversion: People erroneously avoid algorithms after seeing them err. J. Experiment. Psych. Gen. 144(1):114–126.

156

- Ebert RJ, Kruse TE (1978) Bootstrapping the security analyst. J. Appl. Psych. 63(1):110.
- Einhorn HJ (1986) Accepting error to make less error. J. Personality Assessment 50(3):387–395.
- Forrester (2018) Mapping the \$9 trillion US B2B online commerce market. Technical report, Forrester, Cambridge, MA.
- Frey CB, Osborne MA (2017) The future of employment: How susceptible are jobs to computerisation? *Tech. Forecasting Soc. Change* 114:254–280.
- Goldberg LR (1970) Man vs. model of man: A rationale, plus some evidence, for a method of improving on clinical inferences. *Psych. Bull.* 73(6):422.
- Grewal R, Lilien GL, Bharadwaj S, Jindal P, Kayande U, Lusch RF, Mantrala M, et al. (2015) Business-to-business buying: Challenges and opportunities. *Customer Needs Solutions* 2(3):193–208.
- Hausman J, McFadden D (1984) Specification tests for the multinomial logit model. *Econometrica* 52(5):1219–1240.
- Highhouse S (2008) Stubborn reliance on intuition and subjectivity in employee selection. *Indust. Organ. Psych.* 1(3):333–342.
- Hoffman PJ (1960) The paramorphic representation of clinical judgment. Psych. Bull. 57(2):116.
- Jiang Y, He X, Lee M-LT, Rosner B, Yan J (2017) Wilcoxon rankbased tests for clustered data with R package clusrank. Preprint, submitted June 11, https://arxiv.org/abs/1706.03409.
- Kahneman D (2011) Thinking, Fast and Slow (Macmillan, New York).
- Khan R, Lewis M, Singh V (2009) Dynamic customer management and the value of one-to-one marketing. *Marketing Sci.* 28(6):1063–1079.
- Kim M, Sudhir K, Uetake K, Canales R (2019) When salespeople manage customer relationships: Multidimensional incentives and private information. J. Marketing Res. 56(5):749–766.
- Kleinberg J, Lakkaraju H, Leskovec J, Ludwig J, Mullainathan S (2018) Human decisions and machine predictions. *Quart. J. Econom.* 133(1):237–293.
- Kunreuther H (1969) Extensions of Bowman's theory on managerial decision-making. *Management Sci.* 15(8):B415–B439.

- Lam SY, Shankar V, Erramilli MK, Murthy B (2004) Customer value, satisfaction, loyalty, and switching costs: An illustration from a business-to-business service context. J. Acad. Marketing Sci. 32(3):293–311.
- Larkin I, Leider S (2012) Incentive schemes, sorting, and behavioral biases of employees: Experimental evidence. Amer. Econom. J. Microeconom. 4(2):184–214.
- Lilien GL (2016) The B2B knowledge gap. Internat. J. Res. Marketing 33(3):543–556.
- Meehl PE (1954) Clinical vs. Statistical Prediction: A Theoretical Analysis and a Review of the Evidence (University of Minnesota Press, Minneapolis).
- Misra S, Nair HS (2011) A structural model of sales-force compensation dynamics: Estimation and field implementation. *Quant. Marketing Econom.* 9(3):211–257.
- Morgan RM, Hunt SD (1994) The commitment-trust theory of relationship marketing. J. Marketing 58(3):20–38.
- Nedelkoska L, Quintini G (2018) Automation, skills use and training. OECD Social, Employment and Migration Working Papers 202, OECD Publishing, Paris.
- Payne JW, Payne JW, Bettman JR, Johnson EJ (1993) The Adaptive Decision Maker (Cambridge University Press, Cambridge, UK).
- Petrin A, Train K (2010) A control function approach to endogeneity in consumer choice models. J. Marketing Res. 47(1):3–13.
- Rubin DB (1980) Randomization analysis of experimental data: The Fisher randomization test comment. J. Amer. Statist. Assoc. 75(371):591–593.
- Shi H, Sridhar S, Grewal R, Lilien G (2017) Sales representative departures and customer reassignment strategies in businessto-business markets. J. Marketing 81(2):25–44.
- Wiggins N, Kolen ES (1971) Man vs. model of man revisited: The forecasting of graduate school success. J. Personality Soc. Psych. 19(1):100–106.
- Zhang JZ, Netzer O, Ansari A (2014) Dynamic targeted pricing in B2B relationships. *Marketing Sci.* 33(3):317–337.