Structural Estimation of a Moral Hazard Model: An Application to Business Selling

Renna Jiang*

Graduate School of Business
University of Chicago

Job Market Paper

Oct 22, 2008

Abstract

We propose and estimate a moral hazard model for contracts in business selling context, where manufacturers of industrial materials hire outside independent sales organizations as their representatives ("rep firms") to sell products to business customers. In return, rep firms receive commissions on realized sales. The question we address in this research is whether the observed commission rates are set at the optimal level, and if not, what are the economic consequences. This is different from most previous empirical contracting work which tests comparative statics predictions derived from theoretical models that impose optimality on manufacturers. A unique feature of the data is that we have a measure of salespeople's effort obtained from surveys. The effort data allow us to build a realistic model where salespeople have better information than the manufacturer about the opportunities in the field. Our empirical results show that optimal commission rates are higher than what are observed in the data, and that manufacturers could achieve greater profits by adopting the proposed rates.

Keywords: moral hazard, commission rate, business markets, sales force compensation, structural model.

JEL Classification: C13, D82, M52.

*I am grateful to my dissertation committee co-chairs Pradeep Chintagunta and Peter Rossi for their constant guidance and support. I especially thank Rob Palmatier at the University of Washington for data, without whose help this project would not have been possible. I benefited from helpful discussions with Jack Barron, Jean-Pierre Dubé, Ron Goettler, Milton Harris, Günter Hitsch, Yesim Orhun, Oleg Urminsky, and Ting Zhu. I also thank Michael Emerson from the Indian River Consulting Group for sharing with me his industry expertise. All errors are mine. Email: rjiang1@ChicagoGSB.edu.

1 Introduction

The design and implementation of an optimal contract are relevant issues in many economic situations where a "principal" hires an "agent" to undertake certain actions for the former. Moral hazard (Holmström 1979) can arise in these situations when the agent's objective is not entirely aligned with that of the principal, and the principal is unable to observe how much effort the agent puts into the job.

One marketing context in which the moral hazard issue is particularly relevant is sales force compensation. While theoretical marketing researchers have mostly focused on the design of an optimal contract (cf., Basu, Lal, Srinivasan, and Staelin 1985, Lal and Srinivasan 1993, Misra, Coughlan, and Narasimhan 2005), much of the empirical work has been conducted to test various comparative statics predictions obtained from these theoretical models (cf., Coughlan and Narasimhan 1992, Joseph and Kalwani 1995, Misra et al. 2005). For reviews on general agency relationships, their implications, and empirical findings, see Bergen, Dutta, and Walker (1992), and Prendergast (1999), Chiappori and Salanié (2003).

There is little research that investigates the optimality of an existing contract between a firm and its sales force.[1] An impediment has been the absence of sales force effort data without which one cannot directly investigate the relationship between salespeople's input (i.e., effort) with output (i.e., sales). At the same time, getting access to such effort data would seem unlikely, because if firms observe this information then the moral hazard problem would not exist at the first place. As noted in the recent meta-analysis by Albers, Mantrala, and Sridhar (2008), "data on personal selling efforts and effects have been difficult to obtain, limiting the number of salesforce-focused, market response studies relative to the volume of work on advertising and pricing."

In this research, we use a unique data set, which contains a measure of salespeople's effort, to study one form of contractual relationship between a firm and the independent sales organizations that this firm uses for purpose of selling its products. The data are

[1] There are notable exceptions in the general empirical contracting literature. Lazear (2000) observes a natural experiment where a company that installs automobile glass improves profits by switching the compensation method for its workforce from hourly wages to piece-rate pay. In heath care insurance, Vera-Hernández (2003) conducts structural estimation of a moral hazard model and uses those parameter estimates to compute the optimal contract.

in a business market context, where manufacturers (the "principal" in our model) of industrial materials hire outside independent sales organizations (or, brokers, the "agent" in our model) as representatives (hence the term "rep firms") to sell products to business customers. In return, rep firms receive commissions on realized sales. The question we aim to address in this research is whether the commission rates are set at the optimal level, and if not, what are the economic consequences.

The motivation behind our research question is an observation that we obtained from interacting with industry experts. In particular, there seems to be a convention among manufacturers to pay their reps a 5% commission rate. This is consistent with our data where both the mean and median of commission rates are 5%. Although there exists variation in the observed commission rates, the fact that these rates seem to be anchored on a fixed number suggests that manufacturers need not be optimizing their commission rate choices.

The data are obtained from surveys conducted by an independent academic researcher who is not affiliated with either manufacturing firms or rep firms. The survey collects information from each salesperson employed by the rep firm and each customer that the salesperson sells to. The main variables in the data are sales and commission rates at the customer level, as well as effort (measured as monetary cost) exerted at each customer by each salesperson.

We formulate a structural model consisting of two main equations: (1) a production function relating effort to sales; and (2) an effort equation characterized by the selling agents' incentive compatibility conditions. The second equation is at the heart of the moral hazard problem. The manufacturer cannot directly impose a desired level of effort upon the agents because actual effort is unobservable. Agents are free to choose effort levels in their best interests. In particular, agents choose effort levels based on the productivity of their efforts and on commission rates. In order to motivate the agents to take the "correct" actions, the manufacturer needs to know how agents' effort reacts to commission rates.

The effort data allow us to build a realistic model where the agent has better information than the manufacturer about the opportunities in the field. Specifically, effort's productivity, i.e., the effectiveness of effort, is determined by two components: (1) a deterministic component that may depend on characteristics of the rep firm, the field salesperson, and the customer; (2) a random shock that is customer specific. The salesperson observes both components before making the effort decision, while the manufacturer does not observe the random component.

In addition, we extend the model to allow for unobserved (to the researcher) factors that could potentially affect both the productivity of effort and the commission rate. This can be motivated by situations in which the manufacturer has some knowledge about the effort productivity (e.g., the selling difficulty of its products), and adjusts commission rates accordingly.

The advantage of using a structural model is that we can use the parameter estimates to compute optimal commission rates, and to further quantify the economic consequences of the manufacturer's seemingly sub-optimizing behavior. Our empirical results show that the optimal commission rates are on average higher than the observed rates in the data (11% vs 5%), and that manufacturers could improve profits (after paying commissions) by about 4% if they adopt the proposed optimal rates. In addition, we find support for the notion that manufacturers are willing to pay higher commission rates for more difficult to sell products.

The rest of the paper is organized as follows. We provide relevant institutional details in Section 2, and then discuss the data in Section 3. The basic model and estimation results are in Section 4. Section 5 relaxes an assumption made in Section 4. Optimal commission rates and profit implications are obtained in Section 6. In Section 7, we provide a discussion on the importance of observing effort data from the perspective of identifying model parameters. Finally, we conclude in Section 8.

2 Institutional Background

In this section, we first provide some background information on business markets, then discuss two important features that are relevant for the current paper. Namely, (1) salespeople's effort plays an important role in influencing business customers' purchase decisions;

(2) manufacturers pay their rep firms on a commission basis, and there are typically no fixed payments.

2.1 General Background Information

Anderson and Narus (1999) define business markets as "firms, institutions, or governments that acquire goods and services either for their own use, to incorporate into the products or services that they produce, or for resale along with other products and services to other firms, institutions, or governments." The importance of business marketing, also known as industrial marketing or business-to-business (B2B) marketing, is underscored by Hutt and Speh (2001) who note that "business marketers serve the largest market of all; the dollar volume of transactions in the industrial or business market significantly exceeds that of the ultimate consumer market." These authors mention that companies such as GE, DuPont, and IBM spend more than $60 million a day on purchases to support their operations. Consistently, according to Dwyer and Tanner (2009), purchases made by companies, government agencies and institutions "account for more than half of the economic activity in industrialized countries such as the United States, Canada and France."

This research focuses on the part of the business markets that are served by professional selling agencies. In our sample (refer to Figure 1), manufacturers (or, suppliers) of industrial materials hire outside independent selling organizations as representatives (hence the term reps) to sell products to business customers. According to the United States Census Bureau (2007), reps account for 10.5% of U.S. wholesale sales volume in 2002; the rest of sales are carried out by wholesaler-distributors (both independent and in-house).

Figure 1: Rep firms as intermediaries in business market

Rep firms are small business enterprises that specialize in selling.[2] Unlike wholesalers or

[2]Salespeople working within the rep firm are generally referred to as "manufacturers' representatives,"

distributors, rep firms do not take title to the merchandise they sell, nor do they handle the merchandise. Rep firms work mainly in a given geography (e.g., a state), and typically sell hundreds of products representing ten or more manufacturers. Unlike most distributors, a rep firm cannot carry products from two directly competing manufacturers. For each manufacturer, the rep firm carries a catalogue of products. Customers on average buy a wide range of products from three to five different manufacturers that are represented by a rep firm.

2.2 The Role of Salespeople's Effort

One unique feature of the business markets, compared to consumer markets, is the emphasis on personal selling. Different from individual consumers, the industrial buyers in general consider salespeople, as opposed to advertising, to be the most important promotional element when making a purchase decision (Jackson, Keith, and Burdick 1987). According to Dwyer and Tanner (2009), the greater emphasis on personal selling in business marketing is the result of stronger relationships between the buyer and seller, as well as shorter distribution channels. They also note that "salespeople are the members of the organization responsible for coordinating their company's efforts at satisfying their customers." In what follows, we discuss how salespeople play their role in facilitating sales.

Manufacturers have a list of recommended prices for products. These recommended prices also reflect volume discounts. Thus, the salesperson has little room to cut price. In addition, rep firms have no control over manufacturing costs or product features. In other words, salespeople try to sell products by actively interacting with customers, providing special information and/or reports on the products, building personalized relationships by providing meals, entertainment or gifts, as well as offering value-added benefits like customized order-processing. Such efforts have been found to affect a customer's value to the rep firm by increasing the length, breadth, and depth of the buying relationship (Bolton, Lemon, and Verhoef 2004).

Indeed, since rep firms have few tangible assets, and the manufacturer typically can ter-

"reps," or "agents." In financial services, the counterpart label is "brokers." In this paper, we sometimes use the term "reps" or "agents" to loosely refer to the rep firm and/or its salespeople unless when a clear distinction between the rep firm and its salespeople is necessary.

minate a contract simply with a 30-day notice, building quality relationships with customers is important for these rep firms in retaining a manufacturer's contract. This is directly related to another feature of the manufacturer-rep firm relationship. Namely, the rep firms absorb all selling expenses and are paid by commissions on realized sales; manufacturers typically do not make any fixed payments (i.e., salaries) to rep firms. We discuss this in the following section.

2.3 Commissions Only

The fact that rep firms are paid by commissions only stands as a marked contrast to the manufacturer's employee or direct sales force, where 90% of the compensation plans involve both salary and commission components (Dwyer and Tanner 2009). Salary is necessary when there is need to either encourage nonselling activities such as taking care of the customers, or provide security to salespeople. Neither appears essential when the sales function is outsourced to rep firms. First, rep firms are professional selling agencies whose main assets are their relationships with the customers. Thus, it is likely that they are already motivated to engage in a variety of activities to enhance their bonds with the customers. Second, compared to individual salesperson in the direct sales force, rep firms as an organization have greater ability at absorbing income shocks, hence need less insurance from fixed payments. In fact, according to Anderson and Trinkle (2005), turning fixed costs into variable costs is one of the most cited reasons why a manufacturer outsources the sales function to rep firms.[3]

Industry experts reveal that the commission rates are largely set by following an industry convention of a 5% commission rate, which is consistent with our data: both the mean and median of observed commission rates are 5%. The variation in commission rates might be explained by at least two factors. First, rep firms differ in their negotiating power and selling strength (e.g., some have better brands, more synergistic lines, better salesperson, etc.). Second, the variation can also be caused by heterogeneity in products, e.g., products' selling difficulties (Anderson and Trinkle 2005). In some cases, the manufacturer may be

[3] These authors list three other reasons as: improving efficiency through the third party's economy of scale, performing the function better with the third party's specialization, and freeing up resources which allows a firm to focus on its core competencies.

willing to pay a high commission rate to push new or hard-to-sell products, while in other cases, the commission rate can be low if the product is relatively easy to sell.

Note that when a manufacturer hires a rep firm to sell products, moral hazard issues are likely to emerge because it requires costly effort to generate sales, and the rep firm receives only a fraction of each incremental sale. Thus, it is likely that the agents exert less effort than they "should" (from the manufacturer's perspective). But the manufacturer cannot force the desired level of effort upon the agents because actual effort is unobservable. Indeed, Dwyer and Tanner (2009) quote "loss of control" as manufacturers' major concern for using reps. To maximize the net sales income (or other objectives), the manufacturer needs to design a compensation scheme that properly motivates the agents.

3 Data

3.1 Data Collection Procedure

The data used in this paper are those described in Palmatier, Scheer, and Steenkamp (2007). The data were collected during early 2003 by surveying industrial customers, salespeople, and sales managers of rep firms in the United States. With support from the Manufacturers' Representatives Educational Research Foundation (MRERF), managers from 41 rep firms serving electronic and electrical components, industrial cleaning supplies, utility subcomponents, and telecommunication industries provided contact information for their salespeople and for customers serviced by each salesperson.[4] From an initial pool of 13,850 customers, a stratified random sample (based on sales and industry) of 3,000 customers was generated. These customers received a four-wave mailing: a presurvey notification postcard, the survey one week later, a follow-up reminder postcard one week later, and a duplicate survey three weeks later. A personalized cover letter identified the rep firm and buyer's salesperson, discussed endorsements by the industry association, and offered a summary report and entry into a raffle for one of twenty $25 Amazon gift certificates. From the 2,780 delivered surveys (with 220 returned as undeliverable), 511 completed surveys were returned for an effective

[4] Managers were requested to randomly select customer firms representing a range of sizes and sales; many managers provided their full customer contact database. Managers also identified the individual buyer responsible for most of the rep firm's business with each customer firm, and these buyers later received mailed surveys. For most of our discussion, we do not distinguish between the terms "buyer" and "customer."

response rate of 18.4%.

A presurvey postcard was then mailed, followed by a survey one week later to the 195 salespeople who handled these 511 customers. In the survey, the salesperson provided two crucial measures for the current study (i.e., the monetary cost of exerting effort, and average commission rate at the customer), as well as items addressing the salesperson's relationship with the counterpart customer. For cases in which the same salesperson covered multiple customers, the salesperson completed all measures for each customer. After a second mailing and follow-up, 165 salesperson surveys were received for a 84.6% response rate.

Concurrently, a customized survey was sent via registered mail to the sales manager at each rep firm, listing each customer and salesperson by name, and requesting annual sales revenue data for all 511 responding customers and other data regarding the rep firm. The cover letter stressed that each buyer had already completed a survey. After follow-up contacts using phone and email, 34 of 41 sales managers provided the requested data, a 82.9% response rate.

After cases with missing data were removed, the final data set included 295 triads from three sources (295 customers covered by 141 salespeople from 33 rep firms), for an effective response rate of 10.6%. A comparison of early (first 25%) versus late (last 25%) responding buyers and salespeople (Armstrong and Overton 1977), as well as buyers included in the final data set versus other responding buyers not included (because of the lack of corresponding salesperson or sales manager data) revealed no significant mean differences ($p > 0.05$) for the demographic or study variables. Therefore, response bias does not appear to be a major concern in this sample.

3.2 Description of Variables

Two crucial variables for this research were obtained from each salesperson: (1) effort, i.e., the monetary cost to take various actions at a customer in order to bring in sales, and (2) the average commission rate that his or her rep firm gets from each customer.

Effort This unique aspect of the data was obtained by asking each salesperson to report the monetary cost to him and his rep firm to exert effort at each customer in order to facilitate sales. Before giving the monetary cost number, in order to facilitate reporting, the salesperson was first reminded of different dimensions of effort by providing qualitative answers (on 1 to 7 scales) to items addressing those dimensions. These items include concrete ones as well as generic ones. The concrete items include: "This customer often receives special reports and/or information." "This customer is often provided meals, entertainment, or gifts." "Special structural changes (e.g., customized order processing) have been instituted for this customer." Other items are more generic and meant to reflect idiosyncratic aspects of each salespeople's effort. Such items include: "This customer often receives special treatment or status." "This customer often receives special financial benefits and incentives." "Dedicated personnel are assigned to this customer beyond what is typical for our rep firm." After rating the above qualitative items, the salespeople then reported the average monthly cost (in dollars) for him and his rep firm to exert these efforts. We then put the responses on an annual basis.

Commission rate The manufacturer sets a commission rate for each product that the rep firm sells for it. Commission rates are set on the basis of the industry norm of 5%, but could vary for reasons that we discussed in the institutional details. As a customer often buys a wide range of products produced by several manufacturers, it is difficult for the salesperson to report dollar sales and commission rate for each individual product. Instead, the salesperson was asked to report a weighted (by sales) average commission rate at each customer calculated as follows. Suppose S is the total sales revenue at a customer, and S_i is the sales of the i^{th} product that this customer purchased from the rep firm; then the sales weight for the i^{th} product is $w_i = S_i/S$. If the commission rate for the i^{th} product is d_i, then the reported average commission rate at a customer is $d = \sum w_i d_i$. This is relatively easy for salespeople to report because they usually have this number in record.

Managers of each rep firm provided information on the average tenure (years) of salespeople in the firm, annual advertising expenditure (including tradeshows and brochures, etc.), the industry that it serves, as well as annual sales revenue at each customer. Thus, the rep firm's revenue from each customer can be obtained by multiplying sales with the average commission rate.[5] Salespeople also reported their total selling experience (years).

[5] Revenue from the customer is $\sum d_i S_i = \sum d_i (w_i S) = \sum (d_i w_i) S = dS$.

Descriptive statistics are presented in Table 1. The customer firm's size (annual sales) was on average \$323 million, and it purchased \$322.6K worth of products from a rep firm a year. The average commission rate that a rep firm got was 5%, and the value of effort put into a customer was \$1.85K a year. The average tenure of salespeople in a rep firm was just under 9 years, and an average salesperson had been in the selling business for 12 years. In addition, 55% of customers are in electrical industry, 35% in electronics, and the rest 10% are in other industries.

It is worth noting that 19 out of 295 customers (6.44%) received zero total effort. Since effort directly drives sales, we examine these 19 observations to see if they are associated with behavior that is inconsistent with our expectations. It turns out that the sales at these customers are well below the population average (\$41.1K vs \$322.6K), which is what one would expect to observe. In addition, the rep firms that are spending zero effort at these customers are spending non-zero effort elsewhere. Therefore it is not the case that some rep firms systematically exert zero effort. Further, the salespeople handling these customers do not always exert zero effort either. Based on these checks, we do not believe that the zero effort observations are outliers or results of mis-reports, and therefore we need a model that can incorporate zero effort in our analysis.

On Observing the Effort Data As we pointed out, the moral hazard problem emerges because the manufacturer cannot directly observe whether the selling agents are shirking from their duties. If effort is observable, manufacturers would have forced their agents to work in the manufacturers' best interests. One might then wonder how can we as researchers have been able to obtain effort data that the manufacturer firms cannot. The answer is that salespeople in each rep firm are willing to reveal their effort information only to individuals they believe have no vested interests in obtaining that information. This appears to be true for outside academic researchers, but not for the manufacturers that the salespeople work for. In fact, the data do not record the identities of manufacturers, which eliminates salespeople or rep firms' concern that effort information may be leaked to their manufacturers.

3.3 Preliminary Evidence of Moral Hazard in Data

When a manufacturer hires a rep firm to sell products, moral hazard issues are likely to emerge because the manufacturer usually does not observe the rep firm's effort. Intuitively, since it requires costly effort to generate sales and the selling agents receive only a fraction of every incremental sale, it is likely that the salespeople "optimally" (from salespeople's perspective) choose to exert less effort than they "should" (from the manufacturer's perspective).

In our context, if the agents optimally choose effort levels based on commission rate, we would expect to observe positive correlation between the commission rate and effort. That is, the higher the commission is, the more aligned the incentives are, and therefore the more effort is exerted. We run a Tobit model to test this. Note that the Tobit model is more appropriate than OLS because some effort observations are zero.

Results are presented in Table 2. The coefficient on the commission rate is indeed positive. Thus, higher commission rate tends to be accompanied by higher effort. We consider this as being consistent with the notion that salespeople choose effort levels in consideration of commission rates, which reflects the concern of moral hazard. This finding conforms to Prendergast (1999) who concludes after reviewing recent empirical studies across a wide range of contexts that, "there are strong responses of (agents') output to the use of pay-for-performance contracts."

4 Basic Model

We introduce our model in this section. Four key assumptions that we make are as follows.

(i) Each salesperson is a "perfect" agent for the rep firm that he or she belongs to. That is, we assume that the incentives of the rep firm and of its salespeople are entirely aligned, and there is no agency issue between these two parties. Many previous researchers have studied moral hazard problems within an organization. We do not model this internal relationship for three reasons. First, the goal of this paper is to study the upper level relationship between manufacturers and their rep firms (as reflected in a specific contractual form—commission), i.e., a organization-to-organization relationship, rather than the lower level internal relationship within rep firms. Second, we do not have data on such internal transfers. Third, rep firms are relatively small enterprises. A rep firm typically has fewer than fifteen field salespeople who work in a local territory and meet with each other and the sales manager about once a week. Therefore, it is easier for rep firm's managers to monitor their salespeople's action/effort, compared with a manufacturer monitoring the rep firm's action/effort from long distance.

(ii) Both manufacturers and their rep firms are risk neutral.[6] A typical moral hazard model (e.g., Holmström 1979) assumes that the principal is risk neutral, but the agent is risk averse. The argument here is that the principal is often a firm/organization that seeks to hire managers or workers as agents, and since organizations are much better at absorbing income shocks than individuals, it is deemed reasonable to consider an agency model with a risk neutral principal and a risk averse agent. In our context, however, the agent itself is an organization. As we explained in the institutional details in the previous section, rep firms do not even receive any fixed payments from the manufacturer, suggesting that risk aversion is not likely to be an important characteristic of rep firms.

(iii) Salespeople's effort is endogenously determined by their utility maximizing behavior. Compared with manufacturers who base their commission rate decisions on the industry convention of 5%, salespeople have far fewer social norms, if any, to follow when making their effort decisions, as salespeople mostly work alone to serve their customers. Also, salespeople have more autonomy and face less bureaucracy when choosing their effort levels. In addition, salespeople can easily learn from frequent interactions with the customers about the effect of their effort on generating sales. The supporting evidence on this point comes from the vast empirical literature that documents the agents' active responses to their incentive plans.[7]

(iv) In the absence of time-series data, we assume that salespeople are not forward-looking. This assumption is appropriate because the linear compensation structure in our data does not encourage dynamic considerations into agents' actions. In a recent paper by Misra and Nair (2008), the authors investigate the effect of quota plans on salespeople's

[6] In our setting, we do not allow negative transfers. That is, the agent does not have money to buy manufacturer's firm. This is a realistic assumption.

[7] For instance, Prendergast (1999) concludes after reviewing recent empirical studies across a wide range of contexts that, "there are strong responses of output to the use of pay-for-performance contracts."

intertemporal allocation decisions.

In what follows, we first specify the production function, and then outline the agent's decision problem.

4.1 Sales Response Function

Modeling the sales process is much like modeling the production process. The objective is to link input (effort) to output (sales). The literature has mainly considered three forms of sales response function: linear additive (e.g., Basu et al. 1985, Lal and Srinivasan 1993), multiplicative (e.g., Rao 1990), and one with a saturation level of sales (e.g., Mantrala et al. 1994).[8] Linear additive specification is often attractive to theoretical researchers for model tractability considerations. We choose to adopt the multiplicative formulation because it ensures the positiveness of sales and does not require specifying a maximum sales level.

Assume that the sales at customer j follow a multiplicative formulation as follows,

$$S_j = \widetilde{\theta}_j E_j^{\alpha} \varepsilon_j, \tag{1}$$

where

$\widetilde{\theta}_j$ = effort's productivity (i.e., effectiveness of effort) on customer j, $\widetilde{\theta}_j \geq 0$;

E_j = total effort that influences sales at customer j;

α = a scalar parameter, $0 < \alpha < 1$;

ε_j = a random sales shock, $\log(\varepsilon_j) \sim N\left(0, \sigma_\varepsilon^2\right)$.

To allow for non-zero sales at zero effort, we assume that E_j has two components,

$$E_j = e_j + \underline{e}(z_j),$$

where e_j is the observed effort in data; $\underline{e}(z_j)$ can be interpreted as the base-level effort, which is shifted by the advertising expenditure at the rep firm level, z_j. We parameterize $\underline{e}(z_j) = \exp(\underline{e}_0 + \underline{e}_z z_j)$, where the constant $\underline{e}_0$ captures the intrinsic demand of products.

[8] Basu et al.(1985) first investigate optimal compensation plans for cases where sales distribution is gamma or binomial, then transform expected sales to be linear in effort when discussing comparative statics, $E[S] = h + ke$. Lal and Srinivasan (1993) assume $S \sim N\left(h + ke, \sigma^2\right)$. Mantrala et al. (1994) assume $E[S] = S_{\min} + (S_{\max} - S_{\min})(1 - \exp(-ke))$ without an error term.

We further assume that the effectiveness of effort on sales can be decomposed into a deterministic part and a random part:

$$\widetilde{\theta}_j = \exp\left(\overline{\theta}' X_j + \eta_j\right), \qquad \eta_j \sim N\left(0, \sigma_\eta^2\right).$$

The deterministic part of effort's productivity is a function of various observed characteristics X_j. The random part, η_j, represents field opportunities that are observable to salespeople before they makes effort decisions. But η_j is never observable to the manufacturer or researcher.

Taken together, the sales response function can be summarized as follows,

$$\log(S_j) = \overline{\theta}' X_j + \alpha \log\left(e_j + \underline{e}(z_j)\right) + \eta_j + \log(\varepsilon_j). \tag{2}$$

The above equation has two error terms, η_j and ε_j, which represent *ex-ante* and *ex-post* randomness, respectively. Neither shock is ever observable to the researcher or manufacturer. But salespeople observe η_j and take it into account before making effort decisions.

In standard moral hazard models a la Holmström (1979), there is only one error term: the ex-post term ε_j. Observing effort data enables us to build a more realistic model that incorporates an additional source of randomness η_j, which captures the following aspect in the real world selling environment. Salespeople, who personally interact with customers, have more information than the manufacturer on field sales opportunities, and they will use that information to their advantages when making effort decisions. Examples of such field opportunities include changes in local conditions, changes in personnel of the customer firm, productivity shocks due to adding or removing of machinery in the customer firm, or other idiosyncratic shocks that only salespeople can observe through face-to-face interaction with customers. Any other randomness in sales that is not even observable to salespeople prior to their effort decisions is left in ε_j. After sales are realized, ε_j's realization becomes observable to salespeople, and remains unobservable to manufacturer or researcher.

In what follows, we specify salespeople's utility function and model their effort decisions.

4.2 Agents' Utility Function and Optimal Effort

Since at least as early as Marschak and Andrews (1944), applied researchers have worried about the potential correlation between inputs and some unobserved (to researchers) productivity shocks. The worries arise from the fact that decision makers usually have more information than researchers. For example, the productivity shocks may be actually observable to the decision maker, who then adjusts the input level accordingly.

In our context, the decision maker is the selling agent, who interacts with the customer and thus knows more about field opportunities than the manufacturer and researcher. Therefore, the salesperson could use that information (i.e., the effectiveness of the effort in terms of generating sales at a customer) when deciding on the level of effort to exert on that customer. In what follows, we specify agents' utility function and model their decision problem.

Assume that an agent's profit from customer j is as follows,[9]

$$\begin{aligned} U_j &= d_j S_j - e_j \\ &= d_j \widetilde{\theta}_j \left(e_j + \underline{e}\left(z_j\right)\right)^{\alpha} \varepsilon_j - e_j, \end{aligned}$$

where d_j is the commission rate at customer j, and S_j represents sales. The first part of the profit function is revenue: the agent receives a fraction of d_j from sales realized in the current period, S_j. Before deciding on how much effort to put into a customer, the agent observes d_j and the realization of $\widetilde{\theta}_j$, and knows that his effort can influence the deterministic part (from his point of view) of sales in the following fashion, $\widetilde{\theta}_j \left(e_j + \underline{e}\left(z_j\right)\right)^{\alpha}$. The second part in the profit function is cost of effort. Since effort is observed in monetary values in data, we let it enter linearly into the agent's profit function.

Assume that the agent is risk neutral, and that the utility is equal to profit, U_j. Then the expected utility is

$$\begin{aligned} EU_j &\equiv E\left[U_j\right] \\ &= d_j E\left[S_j\right] - e_j \\ &= d_j \exp\left(\log\left(\widetilde{\theta}_j \left(e_j + \underline{e}\left(z_j\right)\right)^{\alpha}\right) + \frac{1}{2}\sigma_{\varepsilon}^2\right) - e_j \\ &= \overline{d_j}\widetilde{\theta}_j \left(e_j + \underline{e}\left(z_j\right)\right)^{\alpha} - e_j, \end{aligned}$$

where $\overline{d_j} \equiv d_j \exp\left(\frac{1}{2}\sigma_{\varepsilon}^2\right)$. Then, an agent facing customer j tries to maximize the expected

[9] A salesperson typically handles several customers. We assume that the effort exerted on one customer does not affect sales at another customer.

utility by solving the following problem:

$$\underset{e_j}{Max}\ EU_j\left(e_j\left|\widetilde{\theta}_j\right.\right).$$

As discussed earlier, the agent observes the realization of the shock in $\widetilde{\theta}_j$, η_j, but not of ε_j, before choosing effort. Take first order derivative of the above expected utility with respect to effort, we get

$$g_j \equiv \frac{\partial EU_j}{\partial e_j} = \alpha\overline{d_j}\widetilde{\theta}_j\left(e_j + \underline{e}\left(z_j\right)\right)^{\alpha-1} - 1.$$

Since the agent's objective function is globally concave in effort, the optimal effort can be characterized by Kuhn-Tucker conditions

$$\begin{cases} g_j\left(\widetilde{\theta}_j, e_j^*\right) = 0 & \text{if } e_j^* > 0 \\ g_j\left(\widetilde{\theta}_j, 0\right) < 0 & \text{if } e_j^* = 0 \end{cases}$$

or equivalently,

$$\alpha\overline{d_j}\widetilde{\theta}_j\left(e_j^* + \underline{e}\left(z_j\right)\right)^{\alpha-1} - 1 = 0 \quad \text{if } MB_0 > 1 \qquad \text{(IC1)}$$

$$e_j^* = 0 \quad \text{if } MB_0 \leq 1 \qquad \text{(IC2)}$$

where $MB_0 = \alpha\overline{d_j}\widetilde{\theta}_j\left(\underline{e}\left(z_j\right)\right)^{\alpha-1}$. A graphical illustration of the incentive compatibility conditions is:

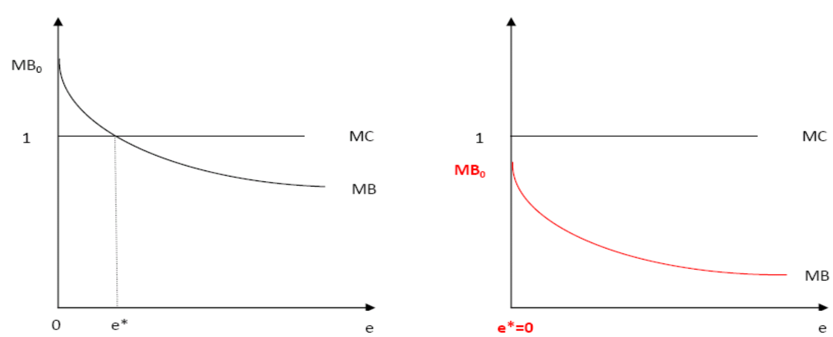

Figure 2: Left: IC1. Right: IC2

First, notice that the marginal cost of exerting effort is a constant, 1, because the effort is itself measured in dollar. The marginal benefit of effort decreases with effort because $0 < \alpha < 1$. Efforts are by nature non-negative. When the marginal benefit at zero effort (denoted as MB_0) is smaller than the marginal cost, one cannot further increase the marginal benefit by cutting effort—the optimal effort is achieved at the corner, 0, which is what Figure 2(right) depicts. When MB_0 is greater than the marginal cost, the principle of $MC = MB$ pushes optimal effort above zero, and we get an interior solution in Figure 2(left) Thus, the model allows for the possibility that an agent exerts zero effort. This is important because 19 out of 295 (6.44%) total observed effort is zero.

The two IC conditions imply several properties of effort. First, it weakly increases in commission rate. The greater share of sales that salespeople get (versus manufacturers get), the higher effort they will exert. This is consistent with what we presented as preliminary evidence for moral hazard in the previous section. Second, it weakly increases in sales' responsiveness to effort, i.e., effort's productivity, $\widetilde{\theta}_j$. The easier it is to sell, the higher the effort. In other words, difficult selling tasks discourage salespeople's effort. Third, it weakly decreases with the base level effort $\underline{e}(z_j)$, because it is the "total" effort that matters. All these properties make intuitive sense.

Next, we use the IC conditions to derive the density of effort in the estimation step. They also serve as constraints later in searching for optimal commission rate.

4.3 Likelihood

The likelihood can be obtained by multiplying the joint density of sales and effort across observations:

$$\begin{aligned} L\left(\alpha, \overline{\theta}, \underline{e}, \sigma_\eta^2, \sigma_\varepsilon^2\right) &= \prod_{j=1}^{295} f\left(S_j, e_j\right) \\ &= \prod_{j=1}^{295} f\left(S_j \left| e_j \right.\right) f\left(e_j\right). \end{aligned}$$

Next, we derive $f(e_j)$ from the IC conditions, and $f\left(S_j \left| e_j \right.\right)$ based on the sales response function and IC conditions. Some complication arises from the fact that optimal effort can be achieved at the corner solution of zero.

4.3.1 Marginal density of effort

The model predicted effort is determined by the two IC conditions, one for non-zero effort, the other for zero effort. Effort inherits its randomness from the productivity shock η_j. For $e_j > 0$, calculating its density is straightforward. The condition (IC1) implies that

$$\begin{aligned}\log\left(e_j + \underline{e}\left(z_j\right)\right) &= \frac{1}{1-\alpha}\log\left(\alpha\overline{d_j}\widetilde{\theta}_j\right) \\ &= \frac{1}{1-\alpha}\log\left(\alpha\overline{d_j}\right) + \frac{1}{1-\alpha}\overline{\theta}' X_j + \frac{1}{1-\alpha}\eta_j \\ &= \mu_j + \frac{1}{1-\alpha}\eta_j \\ &\sim N\left(\mu_j, \frac{\sigma_\eta^2}{\left(1-\alpha\right)^2}\right), \end{aligned} \tag{3}$$

where we define $\mu_j \equiv \frac{1}{1-\alpha}\log\left(\alpha\overline{d_j}\right) + \frac{1}{1-\alpha}\overline{\theta}' X_j$. Introduce a variable $y_j \equiv \log\left(e_j + \underline{e}\left(z_j\right)\right)$, and use the change-of-variable technique:

$$\pi\left(e_j\right) = \phi\left(y_j \left| \mu_j, \frac{\sigma_\eta^2}{\left(1-\alpha\right)^2}\right.\right)\frac{\partial y_j}{\partial e_j} = \frac{1}{e_j + \underline{e}\left(z_j\right)}\phi\left(\log\left(e_j + \underline{e}\left(z_j\right)\right) \left| \mu_j, \frac{\sigma_\eta^2}{\left(1-\alpha\right)^2}\right.\right),$$

where $\phi\left(\cdot\left|\cdot,\cdot\right.\right)$ denotes the Normal probability density function (pdf) with the specified mean and variance.

For $e_j = 0$, the condition (IC2) implies that the density calculation involves an integral:

$$\begin{aligned}\Pr\left(e_j = 0\right) &= \Pr\left(MB_0 \leq 1\right) \\ &= \Pr\left(\alpha\overline{d_j}\widetilde{\theta}_j\left(\underline{e}\left(z_j\right)\right)^{\alpha-1} \leq 1\right) \\ &= \Pr\left(\exp\left(\eta_{jk}\right) \leq \frac{\left(\underline{e}\left(z_j\right)\right)^{1-\alpha}}{\alpha\overline{d_j}\exp\left(\overline{\theta}' X_j\right)} \equiv M_j\right) \\ &= \Phi\left(\frac{\log\left(M_j\right)}{\sigma_\eta}\right), \end{aligned} \tag{4}$$

where $\Phi\left(\cdot\right)$ denotes the standard Normal cumulative distribution function. After some algebra we see that $\log\left(M_j\right) = \left(1-\alpha\right)\left(\log\left(\underline{e}\left(z_j\right)\right) - \mu_j\right)$.

Thus, the marginal density of effort e_j can be obtained as

$$f\left(e_j\right) = \left[\pi\left(e_j\right)\right]^{\mathbf{1}\left(e_j>0\right)} \times \left[\Pr\left(e_j = 0\right)\right]^{\mathbf{1}\left(e_j=0\right)}, \tag{5}$$

where $\mathbf{1}\left(\cdot\right)$ is an indicator function.

4.3.2 Density of sales conditional on effort

The derivation of $f(S_j | e_j)$ also differs across zero and non-zero e_j. When $e_j > 0$, it is straightforward to calculate the sales shock ε_j:

$$\varepsilon_j = \frac{S_j}{\widetilde{\theta}_j \left(e_j + \underline{e}(z_j)\right)^\alpha} = \frac{\alpha \overline{d_j} S_j}{e_j + \underline{e}(z_j)},$$

where the first line follows from the sales response function, the second line from (IC1) condition. We then derive the density of sales from the density of ε,

$$f(S_j | e_j > 0) = f^\varepsilon(\varepsilon_j) \frac{\partial \varepsilon_j}{\partial S_j} = f^\varepsilon(\varepsilon_j) \frac{\varepsilon_j}{S_j}, \tag{6}$$

where $f^\varepsilon(\cdot)$ is the pdf of lognormal$\left(0, \sigma_\varepsilon^2\right)$.

When $e_j = 0$, the conditional density of sales involves an integral:

$$\begin{aligned} f(S_j | e_j = 0) &= \Pr\left(\widetilde{\theta}_j \left(\underline{e}(z_j)\right)^\alpha \varepsilon_j = S_j\right) \\ &= \Pr\left(\exp\left(\overline{\theta}' X_j\right) \exp\left(\eta_j\right) \varepsilon_j = S_j / \left(\underline{e}(z_j)\right)^\alpha\right) \\ &= \Pr\left(v_j \varepsilon_j = \frac{S_j}{\exp\left(\overline{\theta}' X_j\right) \left(\underline{e}(z_j)\right)^\alpha}\right) \\ &= \int_0^{M_j} f^v(v_j) f^\varepsilon\left(\frac{S_j}{v_j \exp\left(\overline{\theta}' X_j\right) \left(\underline{e}(z_j)\right)^\alpha}\right) dv, \end{aligned} \tag{7}$$

where $v_j \equiv \exp\left(\eta_j\right)$, $f^v(\cdot)$ is the pdf of lognormal$\left(0, \sigma_\eta^2\right)$, and $f^\varepsilon(\cdot)$ is the pdf of lognormal$\left(0, \sigma_\varepsilon^2\right)$. The limits on the last integral are based on (IC2). This integral is approximated using the recursive adaptive Lobatto quadrature (Gander and Gautschi, 2000).

4.4 Model Estimation

The model is estimated using maximum likelihood procedure. Results are presented in the first column in Table 3A, where there are only two observable heterogeneity variables in X: two industry dummy variables (later in Table 3B we will include more variables for robustness check). The parameter α in the sales response function is estimated to be 0.075, suggesting that the sales function is relatively concave, or, in other words, the return of effort diminishes quickly. σ_ε is significant, which confirms effort's unobservability and reinforces that moral hazard is a relevant concern in this context. A significant estimate of σ_η is consistent with our preliminary Tobit analysis: observed variables in the data do not explain all the variation in effort, and agents have their private information.

In addition, the base-level effort coefficient associated with advertising is estimated to be 0.094. Its sign is positive as expected. The estimated $\underline{e}(z_j)$ is about a third of the magnitude of the average observed effort. The sales elasticity with respect to effort can be derived from the sales response function as

$$\frac{\partial S_j / S_j}{\partial e_j / e_j} = \alpha \frac{e_j}{e_j + \underline{e}(z_j)}.$$

At our parameter estimates, the average elasticity across all the observations is 0.045.

5 Relaxing the Assumption that d is Exogenous

In the previous section, we introduced a system of two equations: the sales equation $S(\cdot)$, and the effort equation $e(\cdot)$ that is characterized by agent's incentive compatibility conditions (IC1) and (IC2):

$$\begin{cases} S = S\left(\widetilde{\theta}, e, \varepsilon\right) \\ e = e\left(\widetilde{\theta}(X, \eta), d\right) \end{cases}$$

So far, we have taken the commission rate d as an exogenous variable and computed the density of sales and effort conditional on d. In the effort equation, effort inherits its randomness from the productivity shock η. In this section, we consider the possibility that d and η are dependent.

The dependence between productivity shock and commission rate can be motivated by an omitted variable explanation. As we described in the previous section on institutional details, the commission rate of a product is determined by at least its selling difficulty and some factors that are specific to the rep firm. To the extent that the manufacturer pay a higher commission rate for the more difficult to sell product, we should expect to observe positive correlation between product selling difficulty and its commission rate. Meanwhile, the selling difficulty by definition negatively impacts the productivity of the agent's effort, $\widetilde{\theta}$. Anything that is not included in X but affects product selling difficulty is left in η. Thus, to the extent that the correlation between d and η is driven by product selling difficulty, that correlation should be negative.

If such a negative correlation between d and η indeed exists, then only part of the variation in d helps us to identify the influence of d on agent's effort, i.e., $\frac{\partial e}{\partial d}$. The part of the variation in d that is negatively correlated with η should not be used to infer about $\frac{\partial e}{\partial d}$. Otherwise, $\frac{\partial e}{\partial d}$ will be under-estimated, which means α will be under-estimated based on (3). This will in turn cause under-prediction of the optimal commission rate, because it under-estimates how effective the commission rate is in motivating the agent. To address this issue, we consider two scenarios.

In the first scenario (A), we try to find a proxy for the theoretical construct—selling difficulty, and *assume* that the proxy is all that is common between η and d. Under this assumption, by including the proxy into the productivity term $\widetilde{\theta}$, the residual productivity shock will become totally independent of d. The problem is solved.

In the second scenario (B), however, we cast doubt on the above assumption for two reasons. (1) The proxy can only approximate the theoretical construct. There might be other dimensions of selling difficulty that is not captured by the proxy. (2) There could be factors other than selling difficulty that create dependence between d and η. For these reasons, we add a third equation—the commission rate equation—into the original two-equation-system, and allow for correlation between the commission rate's error term and residual productivity shock.

5.1 Adding A Proxy for Selling Difficulty

To the extent that less commoditized products are more difficult to sell, we should expect to see that d, the weighted (across products) average of commission rate at a customer, is *negatively* correlated with the percentage of sales resulting from commodities (denoted as m) at this customer. So, m inversely proxies for the average product selling difficulty at a customer. Under the assumption in scenario (A), it suffices to include m into $\widetilde{\theta}$ and re-estimate the two-equation-system, $f(S, e\,|d)$.

The average m in data is 69.0% (median=75%, std.=32.6%, min=0, max=1), meaning that the average buyer had other competitive sources for 69% of their purchases from the rep firm.

Estimation results are presented in the second column of Table 3A. We can see that simply adding the proxy—percentage of commodity—into the model does not make any significant changes. This is because the proxy's coefficient, β^{η}, is not statistically significant.

5.2 Adding the Proxy and Correlation

Under scenario (B), we need to specify an equation for d and model $f(S, e, d)$.

$$d_j = \beta^{rep\_j} + \beta^{d} m_j + \xi_j, \tag{8}$$

where β^{rep_j} is the fixed effect for the rep firm that handled customer j; β^{d} is the coefficient on m_j; ξ_j is an error term. Rep firm dummy variables are included to control for the rep-specific factors (e.g., negotiation) that influence commission rate. The OLS regression of (8) produces an R^2 of 0.29, and indeed β^{d} is estimated to be negative (-0.0058, $p = 0.0885$).

It is worth pointing out that equation (8) is not necessarily an *optimal* decision rule for the manufacturer. It serves merely as an approximation to the manufacturer's *current* decision rule.

Next, we include the proxy, m, into the original productivity shock of the two-equation-system:[10]

$$\eta_j = \beta^{\eta} m_j + \upsilon_j, \tag{9}$$

where β^{η} is the coefficient on m_j, and υ_j denotes the new residual in the effort productivity. Taken together, effort's productivity can be written as

$$\widetilde{\theta}_j = \exp\left(\overline{\theta}' X_j + \beta^{\eta} m_j + \upsilon_j\right). \tag{10}$$

In scenario (B), the belief is that there are common unobserved variables in ξ_j and υ_j. Assume that they are jointly distributed as follows (Villas-Boas and Winer 1999):

$$\begin{pmatrix} \upsilon_j \\ \xi_j \end{pmatrix} \sim N\left(\begin{pmatrix} 0 \\ 0 \end{pmatrix}, \begin{pmatrix} \sigma_{\upsilon}^2 & \sigma_{\upsilon\xi} \\ \sigma_{\upsilon\xi} & \sigma_{\xi}^2 \end{pmatrix}\right)$$

[10] The equation (9) is just for illustration purpose, because η_j is a shock and therefore is itself zero meaned. Of course estimate of the constant in X_j in (10) will change after including m_j into $\widetilde{\theta}_j$.

As we explained, if the omitted variable is products' selling difficulty which increases commission rate but reduces effort's effectiveness, $\sigma_{\upsilon\xi}$ should be negative. The conditional density of υ_j given d_j depends on d_j:

$$\upsilon_j \left| d_j \right. \sim N\left(\frac{\sigma_{\upsilon\xi}}{\sigma_\xi^2}\left(d_j - \beta^{rep(j)} - \beta^d m_j\right), \sigma_\upsilon^2 - \frac{\sigma_{\upsilon\xi}^2}{\sigma_\xi^2}\right).$$

Now, it is clear to see that scenario (A) essentially restricts $\sigma_{\upsilon\xi} = 0$. Under such restrictive assumption, equation (8) does not feed back to the two-equation-system (sales and effort) because the conditional distribution of $\upsilon_j \left| d_j \right.$ is the same as the unconditional distribution of υ_j.

However, when $\sigma_{\upsilon\xi} \neq 0$ as in scenario (B), we need to include the commission equation (8), and form the joint density of the three-equation-system as

$$f\left(S_j, e_j, d_j\right) = f\left(S_j, e_j \left| d_j \right.\right) f\left(d_j\right),$$

where $f(d_j)$ is a simple Normal density based on (8); and $f\left(S_j, e_j \left| d_j \right.\right)$ is the same as what we had before with the two-equation-system, except that now we need to use the conditional density of υ_j given d_j. Finally, the likelihood is $L(\Theta) = \prod_{j=1}^{295} f\left(S_j, e_j, d_j\right)$.

Estimation results Parameter estimates are presented in the last column in Table 3A. First, note that the covariance $\sigma_{\upsilon\xi}$ is estimated to be significantly negative; the implied correlation is -0.59. The negative sign indicates that, if not accounting for $\sigma_{\upsilon\xi}$, one would not fully acknowledge the role of commission rate in motivating the agent's effort. A negative $\sigma_{\upsilon\xi}$ suggests that, whatever is left out in the effort's productivity shock is negatively correlated with the commission rate. One example is what we proposed—the products' selling difficulty. In addition, α increases from 0.075 to 0.115, as expected. The average elasticity of effort across all observations increases to 0.053. Standard deviations associated with the general sales shock and agents' private information are still both significant, suggesting that moral hazard is an important issue, and that agents indeed have private information when they make effort decisions.

In sum, allowing for the endogeneity of commission rate results in a significant change in the parameter estimates. This is important because changes in parameter estimates may then affect our predicted optimal commission rate, as we will show in the next section.

Including more observed heterogeneity variables We started with a relatively parsimonious model specification in Table 3A. As a robustness check, we next include more variables into X and re-estimate the three models: basic model, basic model with proxy, and basic model with proxy and correlation. Results are presented in Table 3B. There are no substantial differences from Table 3A. The covariance $\sigma_{\upsilon\xi}$ is still significantly negative, and the implied correlation is -0.58. Accounting for such correlation increases the estimate of α from 0.074 to 0.110. Standard deviations associated with the general sales shock and agents' private information are still both significant. In addition, more experienced salespeople and rep firms that are able to retain their sales force for longer tenure appear to be better at selling. The size of the customer firm also has a positive impact.

Identification of $\sigma_{\upsilon\xi}$ In an analogue of the traditional two-equation system (effort and commission), $\sigma_{\upsilon\xi}$ is not identified without excluded "instruments" because the system has 3 moments (two variances and a covariance) and 4 parameters (see Chapter 7 in Rossi, Allenby, and McCulloch 2005). But here we can identify $\sigma_{\upsilon\xi}$ even without any excluded "instruments" because of the additional moments added by the third equation—the sales equation. See the Appendix for more details.

6 Optimal Commission Rates

After obtaining the structural parameter estimates, we proceed to search for optimal (in a "second-best" sense) commission rates.[11] When evaluating alternative rates, note that manufacturers can only specify contracts that are contingent on the output that they can observe—sales—not on effort. Therefore, the manufacturer needs to take into account that agents will choose effort according to their optimal decision rule: (IC1) and (IC2). The

[11]We restrict our attention to the set of straight commissions only, rather than more general contract form, because straight commission is the only contract form observed in data. Refer to previous discussion on institutional details in section 2 for why straight commission is a reasonable "equilibrium" in the manufacturer-rep relationship. For the role of fixed payment in a more general contract form, see discussion in the last section about risk aversion.

manufacturer's decision problem is

$$\begin{aligned} V_j^* \quad \equiv \quad & \underset{0 \leq d_j \leq 1}{Max} \; (1-d_j) \cdot E_{\widetilde{\theta}} \left[E_\varepsilon \left[S_j \left(\widetilde{\theta}_j, e_j^* \left(d_j, \widetilde{\theta}_j \right), \varepsilon_j \right) \right] \right] \\ & s.t. \; e_j^* \in \arg\max \; d_j \cdot E_\varepsilon \left[S_j \left(\widetilde{\theta}_j, e_j, \varepsilon_j \right) \middle| \widetilde{\theta}_j \right] - e_j \end{aligned}$$

Note that there are two expectations, E_ε and $E_{\widetilde{\theta}}$, in the objective function, while only one expectation, E_ε, in the constraint. This reflects the notion that neither shock in the sales response function (2) is observable to the manufacturer, but agents get to observe $\widetilde{\theta}_j$ through field interaction with customers. The constraint can be replaced by agents' (IC1) and (IC2).[12] E_ε has a closed-form expression $E_\varepsilon \left[S_j \middle| \widetilde{\theta}_j, e_j \right] = \widetilde{\theta}_j \left(e_j + \underline{e}\left(z_j \right) \right)^\alpha \exp\left(\frac{1}{2} \sigma_\varepsilon^2 \right)$, while $E_{\widetilde{\theta}}$ needs to be numerically evaluated (using its conditional distribution in the relaxed model with correlation).

Using parameter estimates in Table 3A (3B), the predicted optimal commission rates are presented in Table 4A (4B). Note that predictions in Table 4A and 4B are very similar. This is not surprising given that there are no substantial differences between the parameter estimates in Table 3A and 3B. Across Table 4A and 4B, the average optimal commission rate increases from about 7.4% to about 11% after allowing for correlation between observed commission rates and effort's productivity shock. The increase in predicted optimal rate is intuitive because $\sigma_{\upsilon\xi}$ is estimated to be significantly negative, which indicates that, by not accounting for $\sigma_{\upsilon\xi}$, one would not fully acknowledge the role of commission rate in motivating the agent's effort. By adopting the proposed optimal commission rates, the manufacturer on average could improve profit (after paying commissions) by about 4%.

7 Discussion: How Does Observing Effort Data Help?

As noted, a unique feature of the data is that there is a measure of effort. Exactly how does observing effort data help? We devote this section to a discussion on the role of effort data. Specifically, we ask the following question. Under the data generating process defined by the production function (2), (IC1) and (IC2), which parameter(s) among $\left\{ \alpha, \overline{\theta}, \underline{e}, \sigma_\eta^2, \sigma_\varepsilon^2 \right\}$ can be identified without observing effort?

[12] Armstrong, Larcker, and Su (2007) deal with situations where the agent's IC condition cannot be replaced with first-order conditions.

First, it is straightforward to see that without observing effort data e_j, the $\underline{e}$ parameters associated with base-level effort cannot be identified. $\underline{e}$ is important because its magnitude is considerable and it affects sales elasticity.

Next, we discuss identification issues for the rest of the parameters. For illustration purpose, only consider the positive effort scenario. Substitute the term $\log\left(e_j + \underline{e}\left(z_j\right)\right)$ in (2) with (IC1) and obtain the following reduced form equation:

$$\log\left(S_j\right) = \frac{\alpha}{1-\alpha}\left(\log\left(\alpha\right) + \frac{1}{2}\sigma_\varepsilon^2\right) + \frac{1}{1-\alpha}\overline{\theta}' X_j + \frac{\alpha}{1-\alpha}\log\left(d_j\right) + \frac{1}{1-\alpha}\eta_j + \log\left(\varepsilon_j\right). \quad (11)$$

where X_j contains a column of one's as the intercept in (2). It is helpful to think of the reduced form equation (11) as a simple regression, where $\log\left(S_j\right)$ is the dependent variable and $1, X_j, \log\left(d_j\right)$ are independent variables. We have the following observations.

- α can be identified from the co-movement between $\log\left(S_j\right)$ and $\log\left(d_j\right)$.
- Co-movement between $\log\left(S_j\right)$ and the non-constant variables in X_j can identify part of the vector $\overline{\theta}$—the part that is not associated with the constant in X_j.
- $\sigma_\varepsilon^2, \sigma_\eta^2$, and intercept in $\overline{\theta}$ are not identifiable. This is obvious because there are only two useful moment conditions left in (11): an overall intercept and residual variance. They cannot separately identify these three parameters.

8 Conclusions

We propose and estimate a moral hazard model for contracts in business selling context, where manufacturers use rep firms to sell products to business customers. These rep firms are independent organizations specialized in selling. They receive commissions from manufacturers on realized sales.

The question we set out to address in this research is whether the observed commission rates are set at the optimal level, and if not, what are the economic consequences. The motivation behind our research question is an observation we obtained from interacting with industry experts. In particular, there seems to be a convention among manufacturers to pay their reps a 5% commission rate. Indeed, both the mean and median of observed commission rates are 5% in our data. The fact that commission rates seem to be anchored on a fixed number raises concern that manufacturers might not be entirely optimizing their commission rates.

Obviously, in order to investigate manufacturers' optimality in terms of choosing commission rates, it is necessary that our model does not *impose* such optimality in the first place. In that sense, this research is different from most previous empirical contracting work which tests comparative statics predictions derived from theoretical models that impose optimality on manufacturers.

The model assumes that rep firms are risk neutral, because the context of our study does not seem to fit into the standard theoretical risk aversion models, which include not only variable payment but also fixed salary. Presumably, the motivation for developing a model with risk averse agent is to investigate the relative importance between fixed salary and variable pay. However, the fact that rep firms do not receive any fixed payments invalidates this approach from the very beginning, and it strongly suggests that rep firms, as an organization, can absorb income shocks and are not behaving far from risk neutrality.

Our empirical results show that optimal commission rates are higher than observed commission rates in the data (11% vs 5%), and that manufacturers could improve profits (after paying commissions) by about 4% if they adopt the proposed optimal rates. In addition, we find support for the notion that manufacturers are willing to pay higher commission rates for more difficult to sell products.

References

[1] Albers, Sonke, Murali K. Mantrala and Shrihari Sridhar (2008). A Meta-analysis of Personal Selling Elasticities. Marketing Science Institute: Working Paper Series, Issue 1, 3-24.

[2] Anderson, Erin and Bob Trinkle (2005). Outsourcing the Sales Function: The Real Costs of Field Sales. Ohio: Thomson Higher Education.

[3] Anderson, James C. and James A. Narus (1999). Business Market Management: Understanding, Creating, and Delivering Value. New Jersey: Prentice-Hall, Inc.

[4] Armstrong, Christopher, David Larcker, and Che-Lin Su (2007). Stock Options and Chief Executive Officer Compensation. Working paper, Stanford University.

[5] Armstrong, J. Scott and Terry S. Overton (1977). Estimating Non-response Bias in Mail Surveys. Journal of Marketing Research, 14(3), 396-402.

[6] Basu, Amiya, Rajiv Lal, V. Srinivasan and Richard Staelin (1985). Salesforce Compensation Plans: An Agency Theoretic Perspective. Marketing Science, Vol. 4, No. 4, 267-291.

[7] Bergen, Mark, Shantanu Dutta, and Orville C. Walker, Jr. (1992). Agency Relationships in Marketing: A Review of the Implications and Applications of Agency and Related Theories. Journal of Marketing, 56(3), 1-24.

[8] Bolton, Ruth N., Katherine N. Lemon and Peter C. Verhoef (2004). The Theoretical Underpinnings of Customer Asset Management: A Framework and Propositions for Future Research. Journal of the Academy of Marketing Science. 32(3), 271-292

[9] Chiappori, P.A. and B. Salanié (2003). Testing Contract Theory: A Survey of Some Recent Work. In M. Dewatripont, L.P. Hansen, and S .J. Turnovsky, eds., Advances in Economics and Econometrics—Theory and Applications, Eighth World Congress. New York: Cambridge University Press.

[10] Coughlan, Anne T. and Chakravarthi Narasimhan (1992). An Empirical Analysis of Sales-Force Compensation Plans. Journal of Business, 65(1), 93-121.

[11] Dwyer, F. Robert and John F. Tanner (2009) Business Marketing: Connecting Strategy, Relationships, and Learning, 4th Edition, New York: McGraw-Hill/Irwin

[12] Holmstrom, Bengt (1979). Moral Hazard and Observability. Bell Journal of Economics. 10(Spring), 74-91.

[13] Hutt, Michael D. and Thomas W. Speh (2001) Business Marketing Management: A Strategic View of Industrial and Organizational Markets, 7th Edition, Harcourt Inc.

[14] Gander, Walter and Walter Gautschi (2000). "Adaptive Quadrature—Revisited", BIT Numerical Mathematics, 40(1), 84-101.

[15] Jackson, Donald W., Janet E. Keith and Richard K. Burdick (1987). The Relative Importance of Various Promotional Elements in Different Industrial Purchase Situations. Journal of Advertising, 16(4), 25-33.

[16] Joseph, Kissan and Manohar U. Kalwani (1995). The Impact of Environmental Uncertainty on the Design of Salesforce Compensation Plans. Marketing Letters, 6(3), 183-197.

[17] Lal, Rajiv and V. Srinivasan (1993). Compensation Plans for Single- and Multi-product Salesforces: An Application of the Holmstrom-Milgrom Model. Management Science, 39(7), 777-793.

[18] Lazear, Edward P. (2000). Performance Pay and Productivity. The American Economic Review, 90(5), 1346-1361.

[19] Mantrala, Murali, Prabhakant Sinha and Andris Zoltners (1994). Structuring a Multiproduct Sales Quota-Bonus Plan for a Heterogeneous Sales Force: A Practical Model-Based Approach. Marketing Science, 13(2), 121-144.

[20] Marschak, J. and W. Andrews (1944). Random simultaneous equations and the theory of production. Econometrica, 12 (3-4), 143-205.

[21] Misra, Sanjog, Anne Coughlan and Chakravarthi Narasimhan (2005). Salesforce Compensation: An Analytical and Empirical Examination of the Agency Theoretic Approach. Quantitative Marketing and Economics, 3, 5-39.

[22] Misra, Sanjog and Harikesh Nair (2008). Quota Dynamics and the Intertemporal Allocation of Sales-Force Effort. Working paper, Rochester University.

[23] Palmatier, Robert W., Lisa Scheer and Jan-Benedict Steenkamp (2007). Customer Loyalty to Whom? Managing the Benefits and Risks of Salesperson-Owned Loyalty, Journal of Marketing Research, 44(2), 185-199.

[24] Prendergast, Canice (1999). The Provision of Incentives in Firms. Journal of Economic Literature, 37(1), 7-63.

[25] Rao, Ram C. (1990). Compensating Heterogeneous Salesforces: Some Explicit Solutions. Marketing Science, 9(4), 319-341.

[26] Rossi, Peter E., Greg M. Allenby and Robert McCulloch (2005). Bayesian Statistics and Marketing. England: John Wiley & Sons Ltd.

[27] United States Census Bureau (2007), 2002 Economic Census: Wholesale Trade-Subject Series and Industry Series. Washington, DC: U.S. Department of Commerce Economics and Statistics Administration

[28] Vera-Hernández, Marcos (2003). Structural Estimation of a Principal-Agent model: Moral Hazard in Medical Insurance. RAND Journal of Economics, 34(4), 670-693.

[29] Villas-Boas, J. Miguel and Russell S. Winer (1999). Endogeneity in Brand Choice Models. Management Science, 45(10), 1324-1338.

Table 1. Descriptive Statistics

Variables	Definition	Mean	Stdev
Sales	Annual sales revenue at a customer (K$)	322.6	506.6
Effort	Annual monetary cost of effort put into a customer (K$)	1.85	1.62
Commission rate	Average commission rate on sales at a customer (%)	5.01	1.92
Exper	Salesperson's total selling experience (Yr)	11.94	9.20
Tenure	Average tenure of salesforce in a rep firm (Yr)	8.89	3.48
Buyer size	The customer company's annual sales (M$)	322.6	491.5
Advertising	Rep firm's annual advertising expenditure (K$)	21.45	16.18
Duration	Relationship duration between a salesperson and customer (Yr)	6.46	5.47

Table 2. Tobit Analysis of Effort

Variables	Model 1	Model 2
Intercept	1.124***	0.309
	(0.276)	(0.397)
Commission rate×100	0.134***	0.116**
	(0.051)	(0.051)
Exper		0.017
		(0.011)
Tenure		0.080***
		(0.028)
Buyersize/100		0.036*
		(0.020)
Advertising/10		−0.091
		(0.069)
Duration		0.009
		(0.019)
Stdev of shock	1.686***	1.634***
	(0.073)	(0.070)
# of obs	295	295
# of params	3	8
-logL	556.46	546.79

Note: standard errors in parentheses. ***: p<0.01, **: p<0.05, *: p<0.10. Some variables are scaled as noted to achieve similar scale on all variables.

Table 3A. Model Parameters

		Basic model	Basic model +proxy	Basic model +proxy+corr
		$f(S,e\|d)$	$f(S,e\|d)$	$f(S,e,d)$
$\overline{\theta}$:	Constant	5.147***	5.112***	4.909***
		(0.176)	(0.211)	(0.184)
	$D_{\text{electricals}}$	−0.424***	−0.420***	−0.318***
		(0.153)	(0.153)	(0.117)
	$D_{\text{electronics}}$	−0.421***	−0.410***	−0.219*
		(0.161)	(0.165)	(0.129)
α		0.075***	0.076***	0.115***
		(0.013)	(0.013)	(0.020)
$\underline{e}$:	Constant	−0.600***	−0.599***	0.227
		(0.188)	(0.188)	(0.202)
	Adver/10	0.094**	0.094**	0.054
		(0.047)	(0.047)	(0.034)
σ_ε	(Sales shock)	1.706***	1.706***	1.688***
		(0.072)	(0.072)	(0.071)
σ_η	(Agent's private info)	0.764***		
		(0.048)		
β^η			0.042	0.059
			(0.138)	(0.106)
σ_υ	(Agent's private info)		0.764***	0.579***
			(0.048)	(0.043)
σ_ξ				1.683***
				(0.074)
$\sigma_{\upsilon\xi}$	(Covariance)			−0.579***
				(0.070)
β^d				−0.633*
				(0.327)
# of obs		295	295	295
# of params		8	9	45
logL		−2479.80	−2479.75	−3004.99

Note: standard errors in parentheses. ***:p<0.01,**:p<0.05,*:p<0.10. In the last column "Basic model +proxy+corr", rep firm fixed effects are included, although not reported, in commission rate eq. (8). Some variables are scaled as noted in the table to achieve similar scale on all variables.

Table 3B. Model Parameters (with More Variables in $\widetilde{\theta}$)

		Basic model	Basic model +proxy	Basic model +proxy+corr
		$f(S,e\|d)$	$f(S,e\|d)$	$f(S,e,d)$
$\bar{\theta}$:	Constant	4.762***	4.729***	4.655***
		(0.260)	(0.244)	(0.203)
	Tenure	0.022	0.023*	0.016*
		(0.016)	(0.012)	(0.010)
	$\mathrm{D}_{\mathrm{electricals}}$	−0.498***	−0.494***	−0.375***
		(0.099)	(0.152)	(0.120)
	$\mathrm{D}_{\mathrm{electronics}}$	−0.387***	−0.377**	−0.211*
		(0.004)	(0.162)	(0.129)
	Exper	0.008	0.008	0.006*
		(0.005)	(0.005)	(0.003)
	Buyersize/100	0.032***	0.032***	0.018***
		(0.009)	(0.009)	(0.006)
	Duration	0.004	0.004	0.003
		(0.008)	(0.008)	(0.005)
α		0.074***	0.074***	0.110***
		(0.001)	(0.013)	(0.020)
$\underline{e}$:	Constant	−0.649***	−0.648***	0.131
		(0.007)	(0.189)	(0.203)
	Adver/10	0.104***	0.105**	0.065*
		(0.023)	(0.047)	(0.035)
σ_ε	(Sales shock)	1.708***	1.708***	1.688***
		(0.059)	(0.072)	(0.071)
σ_η	(Agent's private info)	0.744***		
		(0.016)		
β^η			0.038	0.059
			(0.135)	(0.106)
σ_υ	(Agent's private info)		0.743***	0.577***
			(0.046)	(0.042)
σ_ξ				1.673***
				(0.073)
$\sigma_{\upsilon\xi}$	(Covariance)			−0.555***
				(0.069)
β^d				−0.662**
				(0.325)
# of obs		295	295	295
# of params		12	13	49
logL		−2468.58	−2468.54	−2995.51

Note: standard errors in parentheses. ***:p<0.01,**:p<0.05,*:p<0.10. In the last column "Basic model +proxy+corr", rep firm fixed effects are included, although not reported, in the commission rate equation (8). Some variables are scaled as noted in the table to achieve similar scale on all variables.

Table 4A. Counterfactuals Based on Estimates in Table 3A

		Basic model	Basic model +proxy	Basic model +proxy+corr
Optimal Commission Rates				
	Mean	7.52%	7.53%	11.30%
	Median	7.50%	7.55%	11.45%
Increase in Manufacturer Profit under Optimal Commission Rates	Mean	1.17%	1.17%	4.73%
	Median	0.62%	0.63%	3.73%

Note: mean and median are calculated over 295 customers.

Table 4B. Counterfactuals Based on Estimates in Table 3B

		Basic model	Basic model +proxy	Basic model +proxy+corr
Optimal Commission Rates				
	Mean	7.40%	7.41%	10.79%
	Median	7.40%	7.40%	11.00%
Increase in Manufacturer Profit under Optimal Commission Rates	Mean	1.11%	1.11%	4.19%
	Median	0.58%	0.58%	3.25%

Note: mean and median are calculated over 295 customers.

A Appendix: About the Identification of Covariance $\sigma_{\upsilon\xi}$

The identification for most of the model parameters are relatively straightforward except $\sigma_{\upsilon\xi}$. In this appendix, we provide an unrealistically simple example (in terms of its functional form) to illustrate the idea behind the identification of $\sigma_{\upsilon\xi}$.

Consider a system

$$\begin{cases} e = \beta_1 d + \upsilon \\ d = \xi \end{cases}$$

where the first equation is the main equation, and the second equation is added in an attempt to address the potential endogenous d by allowing for the two shocks to be correlated via $\sigma_{\upsilon\xi}$. As shown in Rossi et al. (2005), $\sigma_{\upsilon\xi}$ is not identified because the system has three moments but four parameters:

$$\begin{aligned} var\left(e\right) &= \beta_1^2\sigma_\xi^2 + \sigma_\upsilon^2 + 2\beta_1\sigma_{\upsilon\xi} \\ var\left(d\right) &= \sigma_\xi^2 \\ cov\left(e,d\right) &= \beta_1\sigma_\xi^2 + \sigma_{\upsilon\xi} \end{aligned}$$

In our case, help comes from a third equation—sales equation:

$$S = \beta_2 e + \upsilon + \varepsilon,$$

which can be further written as $S = \beta_2\left(\beta_1\xi + \upsilon\right) + \upsilon + \varepsilon = \beta_1\beta_2\xi + \left(1+\beta_2\right)\upsilon + \varepsilon$. Now there are three more moments with only two more parameters, β_2 and σ_ε^2:

$$\begin{aligned} var\left(S\right) &= \beta_1^2\beta_2^2\sigma_\xi^2 + \left(1+\beta_2\right)^2\sigma_\upsilon^2 + 2\beta_1\beta_2\left(1+\beta_2\right)\sigma_{\upsilon\xi} + \sigma_\varepsilon^2 \\ cov\left(e,S\right) &= \beta_1^2\beta_2\sigma_\xi^2 + \beta_1\left(1+2\beta_2\right)\sigma_{\upsilon\xi} + \left(1+\beta_2\right)\sigma_\upsilon^2 \\ cov\left(d,S\right) &= \beta_1\beta_2\sigma_\xi^2 + \left(1+\beta_2\right)\sigma_{\upsilon\xi} \end{aligned}$$

Thus all parameters are identified in the three-equation-system, even without excluded instrumental variables.

In fact, in our case, there is even a deterministic relationship between β_1 and β_2, because β_1 in the effort equation is determined by the agents' optimizing behavior which takes into account the sales parameter β_2. That is to say, there are actually only 5 parameters in our case and the system might even be over-identified. But this is by no means a *proof* of identification. We intend to use this simple example only to provide some intuition for why we can identify $\sigma_{\upsilon\xi}$ without traditional instrumental variables.