Risk-Aversion and B2B Contracting under Asymmetric Information: Evidence from Managed Print Services

Jie Ning

Weatherhead School of Management, Case Western Reserve University, jie.ning@case.edu

Volodymyr Babich

McDonough School of Business, Georgetown University, vob2@georgetown.edu

John Handley

Xerox Innovation Group, john.handley@xerox.com

Jussi Keppo

NUS Business School, National University of Singapore, keppo@nus.edu.sg

June 3, 2017

Abstract

Managed print service (MPS) is a type of IT infrastructure service that provides centralized management of companies' printing device fleets. In this paper, we estimate the provider's risk preference in MPS using a proprietary data set from Xerox Corporation. We adopt a structural approach in our empirical analysis by modeling the contracting and usage processes of MPS as a two-stage screening game and building econometric models based on the equilibrium contracts and print volumes. Our econometric models have a unique hierarchical structure that allows clustering of printers with the same contracts in the same company, thereby capturing the B2B nature of MPS. We find that Xerox exhibits risk-aversion in MPS contracting, and provide institutional details of Xerox's commission holdback policy that may cause the observed risk-aversion. In the counterfactual analysis, we demonstrate the significance of the provider's risk-aversion and the implications of the commission holdback policy on equilibrium contracts, the expected earnings of Xerox and customer companies, and their preferences for printer models.

Key words: Risk-aversion, structural estimation, asymmetric information, Business-to-business

(B2B), IT infrastructure services

1 Introduction

Modeling contracting parties' risk preferences is crucial in various business-to-business (B2B) and business-to-consumer (B2C) settings. Risk preferences naturally shape theoretical models, their analysis, and managerial insights. However, although many papers have developed the theory of contracting, empirical studies validating the risk preference assumption remain scarce (Allen and Lueck 1995, Masten and Saussier 2000).

In this paper, we empirically estimate the service provider's risk preference in a particular B2B setting: the managed print services (MPS). MPS is the centralized management of companies and organizations' printer fleets. It covers supplies and the maintenance of printers in return for monthly payments. Because most MPS users (e.g., universities and manufacturers) are not in the professional printing industry, printing is an auxiliary operation to them. Thus, MPS belongs to the wide spectrum of IT infrastructure services, similar to utility computing and data center management. While these services have been developing rapidly in recent years, due to the growing trend towards servitization in IT infrastructure (Lacity et al. 1995, DiRomauldo and Gurbaxani 1998), there have been few empirical studies about their contracting and usage processes.

In this paper, we study the contract and print volume data from the MPS provided by Xerox, and estimate Xerox's risk-aversion. The contracting and usage of services in B2B context, such as MPS, differ from those in B2C contexts. We shall describe the details of the contracting process next. According to our collaborator at Xerox, contracting for MPS involves long negotiations between the service provider and a company, in which the provider designs contracts for given groups of printers to meet the company's needs. Each contract comprises three prices: price per black & white (BW) print, price per color print, and the fixed monthly payment. After the contracts are signed, the service starts, the company's employees use the printing service, and the company pays regularly. A unique feature of B2B context is that the decision how much to print is done by different individuals from the ones who negotiated the contract with the service provider.

Informed by MPS practice, we build a theoretical model of MPS contracting and usage. The theoretical model has two stages: contracting and usage. During the contracting stage, the provider chooses the variable and fixed prices to offer to a customer company for a group of its printers (called the *contracting group*). The customer company has private information over its willingness-to-pay for the print service in each contracting group. During the usage stage, the employees of the customer company choose print volumes within each contracting group to maximize their own utilities. When designing the contracts, the provider maximizes its mean-variance objective over the total payments from all time periods, and over all contracting groups and companies.

We solve for the equilibrium contracts and print volumes. Then we use these results as the basis for several structural econometric models. The econometric models have the distinguishing feature of having hierarchical errors. This hierarchical feature explicitly accounts for the clustering of printers within the same contracting group, and reflects the hierarchical structure of MPS. That is, the contracting of MPS is done on given groups of printers, whereas the usage of the service happens at each individual printer.

The main contributions of this paper are as follows. First, we provide empirical evidence for the provider's risk-aversion. Using a proprietary data set from Xerox Corporation, we show that, in contrast to the usual modeling assumptions in contracting literature, the provider behaves as if it were risk-averse when designing contracts for institutional customers. Because Xerox is the leading provider in MPS and has a big portfolio of companies and printers to diversify its risks, this result might appear counterintuitive. Finance theory offers four reasons for risk-aversion (Froot et al. 1993): bankruptcy costs, external financial costs, progressive tax schedule, and managerial risk-aversion. From our discussion with Xerox, the sales representatives are not concerned with bankruptcy costs, capital market imperfection, or corporate taxes, when they contract with new customers. With respect to managerial risk-aversion, a typical mechanism put forth in the literature is managers' non-diversifiable investment in their jobs. If their performance is drastically lower than expected, they will be punished, possibly even fired. The latter would adversely affect them in the labor market if the markets cannot differentiate between low managerial abilities and bad luck (DeMarzo and Duffie 1992). Again, from conversations with the sales team, being fired does not seem to be their concern either. However, we identified Xerox's corporate policy that could cause managerial risk-aversion: "commission holdback".

According to the sales team, they are given a target profitability when negotiating a contract. During the first year of the new contract, Xerox monitors the profitability from the new customer and compares it with the target. The sales representative collects her commission only if the target has been met. From Xerox's perspective, this policy creates incentives for the sales team to sign up profitable customers, so that the expected profitability is greater than the target. But for the sales team, this policy costs them their commissions if the customer is unprofitable. Consequently, they prefer to minimize likelihood of falling short of the target by having a high expected profitability and a low variability, which introduces risk-aversion in their decisions. Other firms in the MPS industry use similar policies with their sales teams. Therefore, it likely that our findings apply to them as well. Compensation holdback provisions have been shown to reduce the executives' risk-taking behavior in financial reporting (Hodge and Winn 2012). Our analysis implies that commission holdback may induce the sales team to be risk-averse in MPS contracting.

Our second contribution is demonstrating the significance of the provider's risk-aversion and the implications of the commission holdback policy. In §7.1, we show that if Xerox revokes its commission holdback policy and the sales team becomes less risk-averse, Xerox would extract higher surplus from the customers and generate higher expected earnings. Both fixed and variable prices would increase. In §7.2, we show that, if Xerox were able to affect the customer's printer model selection, then its preference for printer models may or may not be aligned with the customer's preference, depending on its level of risk-aversion and the particulars of the contracting group. Particularly, if Xerox reduces its risk-aversion by revoking the commission holdback policy, then both Xerox and the customer would prefer the same printer model.

The rest of the paper proceeds as the following. After describing the related literature in §2 and our data set in §3, we build the game-theoretic model in §4. In §5 we present the econometric model and discuss the estimation methods. Then in §6, we report the estimation results. We present the counterfactual analysis in §7 and conclude in §8. All proofs are in Appendix A.

2 Literature review

Modeling assumptions about the risk preferences of contracting parties in theoretical studies of B2B and B2C pricing vary. However, only a few papers in the vast contracting literature involves testing the risk-aversion assumptions empirically. Gopal et al. (2003) test the risk-aversion assumption of an offshore software outsourcing vendor and finds supportive evidence from the data of a leading Indian software developer. Lafontaine (1992) considers whether franchising can be explained by risk-sharing or one-sided moral hazard models where the franchisee is risk-averse, and reports that the data supports a two-sided moral hazard model. Both of these papers take a reduced-form approach and test whether the predicted relationship between certain variables and risk proxies is supported by the data. Our paper differs from these studies in that we build a structural model and estimate the provider's risk-aversion parameter from the data. A more recent empirical study is Barseghyan et al. (2013), where the authors estimate the risk-aversion of the insurance buyers from a structural model in a B2C setting. In contrast, in this paper we consider the risk preference of the principal who designs the contract in a B2B setting.

Given the characteristics of the MPS contracting process, the provider faces information asymmetry when designing contracts for a company. While theories about the provider's pricing strategy under information asymmetry are well-established, the empirical studies lag behind. One reason for the scarcity of such empirical analyses is the challenge of building econometric models under information asymmetry. This issue is particularly critical for studies following a structural approach. Indeed, as noted in Perrigne and Vuong (2011), many of the empirical studies that consider information asymmetry between supply and demand rely on reduced-form analysis (Chiappori and Salanié 2000). A few exceptions include Miravete (2002), Ivaldi and Martimort (1994), and Crawford and Shum (2007). We contribute to this small body of literature by proposing an econometric approach that significantly simplifies the estimation analysis.

Besides contributing to the contracting literature, our paper is also related to studies in IT outsourcing services. While there have been many papers in IT outsourcing in recent years (e.g., Dey et al. 2010, Fitoussi and Gurbaxani 2012), the rapidly developing area of IT infrastructure services has not received much attention. This paper fills in this gap by building a structural model on the contracting and service utilization in MPS.

Finally, many providers of MPS are large manufacturers of printers like Xerox and Hewlett Packard. Therefore, MPS can also be viewed as an example of after-sale services that is taking place in many manufacturing industries (Cohen et al. 2006, Oliva and Kallenberg 2003). Among studies on both IT outsourcing and after-sales services, particularly relevant to our work are the ones on contracts using agency theory. In IT outsourcing, these papers are concerned with the moral hazard of the service provider and achieving incentive alignments using different contract structures. For example, Fitoussi and Gurbaxani (2012) design performance measures of the provider using a multitask agent model, while Dey et al. (2010) examine the efficiency of the fixed-price, time-and-material, and performance-based contracts. Similar studies comparing performance-based contracts with other contract types in the Operations Management literature on after-sales services are Kim et al. (2007) and Guajardo et al. (2012). In MPS, however, there is only one contract structure adopted in practice—the multi-part tariff contract—and this paper focuses on identifying the optimal contract parameters under the multi-part tariff structure.

3 Data and contracting groups

In this section, we describe the data provided by Xerox on its MPS with a number of US companies. We explain the pre-processing steps and characterize the resulting data set. Our raw data comprises three files: one with print volumes in our sample during March 2010–April 2011 (the *volume file*); one with each device's contract in use during the same time period (the *contract file*); and one with the service cost (the *cost file*). The variables in the volume file are: printer ID, meter reads of cumulative print volume, meter type (BW print or color print), and date of the record. The contract file provides more information on printers and companies. The variables in the contract file are: printer ID, printer model, company name, manufacturer of the printer, contract in use, and date of the record. The cost file includes supplies' cost on a subset of devices. The variables are: printer ID, printer model, name of the consumables (e.g., toner, cartridges, fuser), cost of the consumables, and number of pages printed while the consumables last.

We carry out the following steps to clean the data. Because the service contracts are written based on monthly prints, we compute the monthly BW and color print volume of each device and remove inconsistent volume entries (e.g., negative monthly BW/color volume due to meter resets). We discard printers with fewer than 10 monthly BW or color volume records to obtain reliable estimates for the mean and variances of the monthly volume. We also remove abnormal contracts with zero prices or with multiple different contract records on the same date. Merging the cleaned volume and contract files generates a data set with 3,075 printers from 26 companies.

From the cost file, we compute each printer's consumables cost per BW and color print. Because we only observe cost information on a subset of printers from the contract and volume files, merging all three files generates a data set with 1,021 printers from 6 companies.

Our structural econometric model requires information on the service cost. Thus, we use the data set with 1,021 printers for structural estimation. While the requirement of cost information significantly reduces the sample size, as we show below, observations from the data set with 1,021 printers are representative for the sample of 3,075 printers. Henceforth we refer to the data set with 1,021 printers as the *cost data set*, and the one with 3,075 printers as the *expanded data set.*

Next we present descriptive statistics of the observed contracts and the print volume in the cost and expanded data sets. Each printer is covered by a contract with three components: a fixed monthly payment, a variable price per BW print, and a variable price per color print. A company can have hundreds of printers, but many of the printers share the same contract. In Table 1 we present the number of printers and the number of (unique) contracts observed in each company in both the cost and expanded data sets. The top panel of Table 1 describes the expanded data set, and the bottom panel the cost data set. For confidentiality reasons, we label the companies with numbers. The same company receives the same labels in the expanded and cost data sets.

Table 1. Number of unique contracts vs. Number of printers

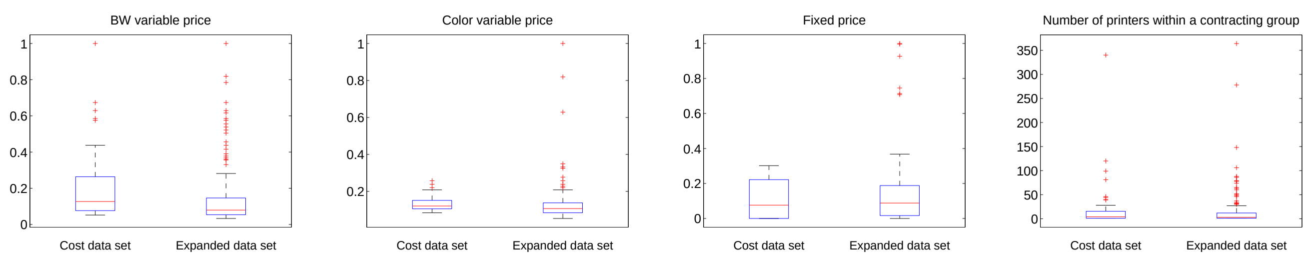
The top panel displays information for the expanded data set, with 3,075 printers and 210 contracts from 26 companies. The bottom panel displays information for the cost data set, with 1,021 printers and 44 contracts from 6 companies. The same company receives the same labels in the expanded and cost data sets.

Company	Num. of printers	Num. of contracts	Company	Num. of printers	Num. of contracts	Company	Num. of printers	Num. of contracts
1	80	16	*2*	387	17	*3*	677	72
4	86	1	*5*	235	18	*6*	364	1
7	46	1	*8*	52	1	*9*	116	5
10	217	5	*11*	21	1	*12*	30	1
13	50	8	*14*	10	3	*15*	16	3
16	163	4	*17*	46	4	*18*	27	5
19	34	2	*20*	63	1	*21*	103	17
22	13	3	*23*	48	1	*24*	18	2
25	145	8	*26*	28	10			
1	36	7	*2*	303	16	*3*	102	3
4	46	1	*5*	194	16	*6*	340	1

From Table 1, while some companies, e.g., Companies 4 and 6, have only one contract for their entire printing fleet, many others have multiple contracts. This observation reflects the B2B nature of MPS: although contracting negotiations happen between Xerox and the company, contracts are usually not designed for the entire company, due to the heterogeneity within the corporation. As we discuss in §4.1.1, this observation motivates our key assumption about the contracting process in MPS. Henceforth, we refer to the set of printers with the same contract as a *contracting group.*

Because we use the cost data set to do the structural estimation, it is important to understand how representative the cost data set is to the expanded data set. In Figures 4a and 3, we characterize and compare the contracts and print volumes in the cost and expanded data sets.

From the box plots on the cost data set in Figures 4a and 3, there is significant variation in the contract prices and print volumes. This allows us to estimate the structural models developed in §5. Further, comparing the box plots from different data sets shows that the cost data set on six companies is representative of the expanded data set on 26 companies. Thus, although our structural estimation is limited to the cost data set, our conclusion is likely to be applicable to the wider customer base of Xerox. Indeed, as we show in Appendix C.1, one of our main results that the provider is risk-averse is supported in the expanded data set as well.



(a) **(b)** **(c)** **(d)**

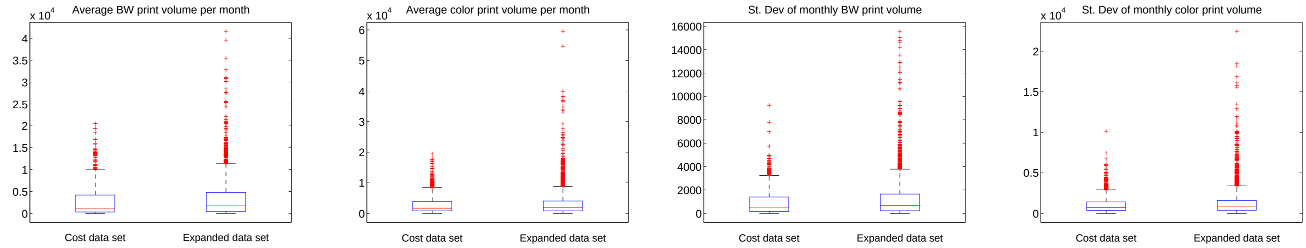
Contract prices in panels (a)–(c) are rescaled for confidentiality reasons.

Figure 1. Contract prices and size of contracting groups

4 Model, equilibrium contracts, and equilibrium print volume

Consider a service provider managing printers for a number of companies. Every company owns a fleet of printers. As discussed in §1, in MPS, the customer company first negotiates with the provider on service contracts. Then, once the contracts are signed, the service starts and the company's service usage, namely the print volumes, are decided period by period. We define one period as one month and the typical length of a service horizon is one year.

Due to the complexity of the B2B contracting problem, we make two key assumptions about



(a) **(b)** **(c)** **(d)**

We observe two streams of monthly print volumes: BW and color, for each printer. We characterize the printing pattern of each printer by computing its average BW and color print volumes per month, and the standard deviations of its monthly BW and color volumes over the observation horizon. Drawing box plots for each of these four values across all printers within a data set generates panels (a) – (d).

Figure 2. Print volumes

the contracting process and the customer's service usage in MPS. These assumptions are important to the formulation of our model, and are thus presented separately in §4.1.1. Building on these assumptions, in §4.1.2 we present the game chronology. In §4.2, we derive the customer's equilibrium print volume given a contract and model information asymmetry in the game. We formulate the provider's contract design problem in §4.3 and specify equilibrium contract prices in §4.4.

4.1 Key assumptions and chronology of the game

4.1.1 Key assumptions

As discussed in §3, in most cases, contracts are not designed for the entire company. Rather, they are designed for groups of printers within the company's fleet. According to Xerox, formation of these contracting groups is driven by exogenous factors such as the heterogeneous service needs and printer functionalities, and the company's own preference (e.g., request for a one-fit-all contract). As a result, when contracting with a company, Xerox takes these contracting groups as fixed and designs contracts for each group. This motivates us to make the following assumption about the contracting process of MPS.

Assumption 1. *The contracting in MPS happens between the provider and the contracting groups of a customer company. The customer company's evaluation of the contract for a particular contracting group is based only on the expected utilities from printers within that contracting group.*

After contracts are signed, printers are used by different employees to satisfy their unique printing needs. Therefore, it is reasonable to assume that the print volume of each printer is individually decided, rather than by a centralized decision maker in a customer company. The motivates the second assumption about the customer company service usage over the contract horizon.

Assumption 2. *The print volume of each printer is individually decided by the employees.*

From Assumptions 1 and 2, we lay out a three-level hierarchical structure of MPS for our analysis. The levels, from the top down, are: the contracting group, the printer, and the period. The contracting between the customer company and the provider happens on the contracting group level, whereas the service usage happens on the printer-period level. This hierarchical structure in which contracting and usage decisions are separated and made by different entities highlights the B2B nature of MPS. It also distinguishes B2B contract design problems from the standard screening problems in a B2C context, where the same person makes contracting and usage decisions. This hierarchical structure of MPS leads to novel features in our theoretical and econometric models.

4.1.2 Game chronology

Let R be the total number of contracting groups, N_r be the number of printers within contracting group r, and τ_r be the contract horizon of group r ($r = 1, 2, \ldots, R$). In the remainder of the paper, we use index pair (r, j) to denote printer j of contracting group r ($r = 1, 2, \ldots, R$; $j = 1, 2, \ldots, N_r$).

Using Assumptions 1 and 2, we model the MPS contracting and usage process as a two-stage game with the following chronology.

Stage 1: the contracting stage. At $t = 0$, the provider offers a set of MPS contracts that screen the company based on its private information about the willingness-to-pay (WTP) for contracting group r. We model the complicated contracting negotiation as a screening problem similar to other economic studies (Ivaldi and Martimort 1994, Wolak 1994).

Stage 2: the usage stage. Every period $t = 1, 2, \ldots, \tau_r$, the employees at the company decide the utility-maximizing print volumes from device (r, j) ($r = 1, \ldots, R; j = 1, \ldots, N_r$).

In the following, let (f_r, p_{rb}, p_{rc}) denote the contract used for contracting group r. It consists of the monthly fixed payment for each printer, f_r, and the variable prices per BW and color print, p_{rb} and p_{rc}. Because the print volumes are two-dimensional in BW and color, it is convenient to work with vectors and matrices in the analysis. We use notation $\vec{x}$ for vectors and matrices. All vectors are column vectors by default, except vectors in the form of (x, y), which are 1-by-2 row vectors. Superscript T means transpose, and subscripts b and c denote the BW and color components of a vector, respectively. Henceforth, we use $\vec{p}_r = (p_{rb}, p_{rc})^T$ to denote the variable prices of the contract.

4.2 Usage-stage model, information asymmetry, and the optimal print volume

We model the provider's cash flows in §4.2.1, introduce the customer's utility and the optimal print volume in §4.2.2, and model the customer's WTP and information asymmetry in §4.2.3.

4.2.1 Provider's cash flow.

Let $\vec{q}_{rjt} = (q_{rjbt}, q_{rjct})^T$ denote the printer volume from device (r, j) in period t $(t = 1, 2, \ldots, \tau_r)$. Over the service horizon, the provider's cash flow from printer (r, j) in period t comprises the service cost and the revenue. According to Xerox, the service cost in month t, S_{rjt}, is approximately affine in the print volume $\vec{q}_{rjt}$:

$$S_{rjt} = \vec{s}_{rj}^{\,T} \vec{q}_{rjt} + \delta_{rjt}, \tag{1}$$

where $\vec{s}_{rj} = (s_{rjb}, s_{rjc})^T$ is the service cost per BW and color page, $\{\delta_{rjt} : t = 1, \ldots, \tau_r\}$ are identically and independently distributed (i.i.d.) random shocks over time and are independent across printers.

Given contract $(f_r, \vec{p}_r)$, the provider's revenue from printer (r, j) in period t is $\vec{p}_r^{\,T} \vec{q}_{rjt} + f_r$. Let $Y_{rjt}(\vec{q}_{rjt}, \vec{p}_r, f_r)$ be the profit received from printer (r, j) in month t under contract $(\vec{p}_r, f_r)$, then

$$Y_{rjt}(\vec{q}_{rjt}, \vec{p}_r, f_r) = (\vec{p}_r - \vec{s}_{rj})^T \, \vec{q}_{rjt} + f_r - \delta_{rjt}. \tag{2}$$

4.2.2 Customer company's utility from printer (r, j) and the optimal print volume.

Because each printer is used by the employees, the company's utility from printer (r, j) is interpreted as the utility of the employees who are using it. Assume that the company's gross surplus, $v_{rjt}(\vec{q}_{rjt})$, when printing $\vec{q}_{rjt}$ from printer (r, j) is

$$v_{rjt}(\vec{q}_{rjt}) = \vec{\chi}_{rjt}^{\,T}\vec{q}_{rjt} - \tfrac{1}{2}\vec{q}_{rjt}^{\,T}\vec{\Phi}^{-1}\vec{q}_{rjt}, \tag{3a}$$

$$\text{where} \qquad \vec{\chi}_{rjt} = \begin{pmatrix} \chi_{rjbt}, \\ \chi_{rjct} \end{pmatrix}, \quad \text{and} \quad \vec{\Phi} = \begin{pmatrix} \phi_b & \phi_{bc} \\ \phi_{bc} & \phi_c \end{pmatrix} \quad \big(\phi_b, \phi_c > 0 \text{ and } \phi_b\phi_c - \phi_{bc}^2 > 0\big). \tag{3b}$$

Equation (3) is the quadratic gross surplus commonly used in the literature (Ivaldi and Martimort 1994, Kim et al. 2010, Lambrecht et al. 2007), and the economic meaning of $\vec{\chi}_{rjt}$ and $\vec{\Phi}$ is explained later in this subsection.

Let $u_{rjt}(\vec{q}_{rjt}, \vec{p}_r, f_r)$ denote the customer company's utility when printing volume $\vec{q}_{rjt}$ from printer (r, j) in period t given contract $(f_r, \vec{p}_r)$. Using the gross surplus $v_{rjt}(\vec{q}_{rjt})$ in equation (3),

$$u_{rjt}(\vec{q}_{rjt}, \vec{p}_r, f_r) = v_{rjt}(\vec{q}_{rjt}) - \vec{p}_r^{\,T}\vec{q}_{rjt} - f_r. \tag{4}$$

In period t, the user of printer (r, j) decides the print volume $\vec{q}_{rjt}$ to maximize the utility $u_{rjt}(\vec{q}_{rjt}, \vec{p}_r, f_r)$, which is concave in $\vec{q}_{rjt}$ from equations (3) and (4). The optimal print volume, denoted by $\vec{Q}_{rjt}(\vec{p}_r, f_r)$, is derived from the first-order condition and satisfies

$$\vec{Q}_{rjt}(\vec{p}_r, f_r) = \vec{\Phi}(\vec{\chi}_{rjt} - \vec{p}_r). \tag{5}$$

From equation (5), the maximum variable prices the provider can charge on printer (r, j) in period t is $\vec{\chi}_{rjt}$. Henceforth, we shall refer to $\vec{\chi}_{rjt}$ as *the customer's WTP*. Furthermore, from equation (5), we observe that matrix $\vec{\Phi}$ captures the sensitivity of the company's print volume to price. Specifically, the diagonal components ϕ_b and ϕ_c of matrix $\vec{\Phi}$ (equation (3b)) are the sensitivities of the BW and color volumes to the BW and color prices, respectively. The off-diagonal component ϕ_{bc} specifies the sensitivities of the BW and color volumes to the color and BW prices, respectively. We shall refer to ϕ_{bc} as the cross-sensitivity.

The derivation above assumes that each printer's monthly demand depends on the prices. While the employees who use the printers are unlikely to be charged directly for their printing, the company may still try to curb the print volume by, say, setting the maximum print volume for each employee. Such quota may induce elastic demand from each device. We use this more general model of the relationship between volume and prices and let the data determine how elastic the demand is. As we show in Section 6, the customer companies indeed manifest elastic demand.

4.2.3 Model for the customer's WTP $\vec{\chi}_{rjt}$ and information asymmetry.

In this subsection, we model the customer's WTP $\vec{\chi}_{rjt}$ in detail. This is a key part of our model because WTP links the contracting process, in which the provider screens the customer based on the customer's (private) WTP, and the print volume given in (5). Screening happens on the contracting group level, but printing happens on the device-period level. Therefore, we need a hierarchical model for the WTP $\vec{\chi}_{rjt}$. We build a parsimonious model by decomposing the WTP $\vec{\chi}_{rjt}$ into group-, printer-, and period-level components, as follows

$$\vec{\chi}_{rjt} = \vec{\zeta}_r + \vec{\eta}_{rj} + \vec{\xi}_{rjt}, \quad \text{where} \quad \vec{\zeta}_r = \vec{\Psi}_r\theta_r + \epsilon_r\vec{1}_2, \quad \text{and} \quad \vec{\eta}_{rj} = \vec{\mu}_{rj} + \vec{\omega}_{rj}. \tag{6}$$

Here $\theta_r \in [0, 1]$ is a scalar, $\vec{1}_2$ is 2-by-1 vector of all one's, $\epsilon_r \overset{\text{iid}}{\sim} N(0, \sigma^2)$, $\vec{\omega}_{rj} \overset{\text{iid}}{\sim} N(0, \vec{\Sigma})$ where $\vec{\Sigma}$ is a diagonal matrix with σ_b^2 and σ_c^2 as the diagonal elements, $\vec{\xi}_{rjt}$ is a random variable with zero mean, and ϵ_r, $\vec{\omega}_{rj}$, and $\vec{\xi}_{rjt}$ are orthogonal to each other and to δ_{rjt} (equation (1)). Next we discuss each component in (6) in detail.

First, consider the contracting group-level WTP, $\vec{\zeta}_r = \vec{\Psi}_r\theta_r + \epsilon_r\vec{1}_2$. Here $\vec{\Psi}_r\theta_r$ is the mean value of WTP at the contracting group level and is known to the customer company at $t = 0$ when contracting takes place. The provider only knows the value of $\vec{\Psi}_r$ but not the value of θ_r. We shall call θ_r the *type* of group r. It captures the customer's private information. To simplify the analysis for the provider's screening problem, we constrain θ_r to $[0, 1]$ and use a contracting-group-specific vector $\vec{\Psi}_r$ as a scaling factor to allow the mean group-level WTP $\vec{\Psi}_r\theta_r$ to be higher than one.

The term $\epsilon_r\vec{1}_2$ in $\vec{\zeta}_r$ represents exogenous shocks to the mean group-level WTP $\vec{\Phi}_r\theta_r$. The value of shock ϵ_r realizes at $t = 1$ and captures the uncertainties of the customer and the provider about the future group-level WTP when they sign the contract at $t = 0$. As shown in §5, ϵ_r generates a group-level random effect in the econometric model, which allows clustering of printers within the same contracting group. To simplify the econometric model, we assume that the BW and color components of $\vec{\Psi}_r\theta_r$ are shifted by the same amount ϵ_r.

The second component $\vec{\eta}_{rj}$ in (6) is the printer-level deviation from the group-level WTP $\vec{\zeta}_r$.

The term $\vec{\eta}_{rj}$ comprises of two parts, a printer-specific mean vector $\vec{\mu}_{rj}$ that is known at $t = 0$ when contracting takes place, and an exogenous shock vector $\vec{\omega}_{rj}$ that realizes at $t = 1$.

The third component $\vec{\xi}_{rjt}$ in (6) represents the temporal shock to WTP. It has zero mean and its value is observed at the beginning of every period t. Shocks $\{\vec{\xi}_{rjt} : t = 1, \ldots, \tau_r\}$ are i.i.d..

Therefore, when viewed at $t = 0$ by the customer, WTP $\vec{\chi}_{rjt}$ fluctuates randomly around the mean value $\vec{\Psi}_r\theta_r + \vec{\mu}_{rj}$ with hierarchical shocks $\epsilon_r\vec{1}_2 + \vec{\omega}_{rj} + \vec{\xi}_{rjt}$. We shall refer to $\vec{\Psi}_r\theta_r + \vec{\mu}_{rj}$ at the *expected WTP of printer* (r, j). When viewed at $t = 0$ by the provider, the customer's WTP $\vec{\chi}_{rjt}$ has an unknown mean. To reflect the provider's uncertainty about θ_r, henceforth we use $\boldsymbol{\theta}_r$ for the random variable of group r's type from the provider's perspective. In contrast, θ_r is the (true) type of group r known to the customer.

Following the common practice in the literature (Miravete 2002, Wolak 1994, Ivaldi and Martimort 1994), we assume that $\boldsymbol{\theta}_r$ follows a beta distribution $\boldsymbol{\theta}_r \sim Beta(1, 1/\kappa)$ with $\kappa \in (0, 1]$. This offers both flexibility and analytical tractability. The value κ measures the level of information asymmetry; a higher κ value indicates stronger information asymmetry. When $\kappa = 1$, beta distribution reduces to a uniform distribution on $[0, 1]$, and a contracting group is equally likely to be of any type. As κ decreases, the beta distribution becomes more condensed, and the provider becomes more informed about the group's true type.

We assume that information asymmetry is one-dimensional. Other than θ_r, all parameters, including the price sensitivity matrix, are known both to the provider and the customer. This assumption is made for analytical tractability. A screening problem with one-dimensional information asymmetry is tractable, but one with multi-dimensional information asymmetry is generally very challenging to solve. The assumption that the price sensitivity is known to both contracting parties is commonly used in similar papers in the literature (Miravete 2002, Lambrecht et al. 2007). Our out-of-sample tests (see Appendix C.2) also provide empirical support for this assumption.

4.3 Contracting-stage model

We discuss the Revelation Principle in the screening problem and formulate the provider's incentive compatibility (IC) and individual rationality (IR) constraints in §4.3.1, and specify the provider's mean-variance objective and its contract design problem in §4.3.2.

4.3.1 The Revelation Principle and IC and IR constraints.

In practice, customers arrive over time and the provider negotiates contracts with them individually. However, as negotiations and contract offerings are private, there is no reputation building for the provider. Furthermore, there is no learning by the provider about the private information over time, because the WTP values across customers are independent. Therefore, the sequential arrival of the customers is inconsequential and the provider's decision-making process can be treated as a static portfolio screening problem, where the provider offers customized contracts to all customers at $t = 0$.

From the Revelation Principle (Laffont and Martimort 2001), it suffices to consider direct and truthful mechanisms in this game. Because the provider designs a contract for each contracting group individually (Assumption 1), the Revelation Principle states that the provider can design contracts based only on the contracting groups' private type, i.e., offer contract menu $\{(F_r(\theta), \vec{P}_r(\theta)) : \theta \in [0,1]\}$ to group r, where $F_r(\theta)$ and $\vec{P}_r(\theta)$ are the fixed and variable prices for group r if its type were θ, such that group r with type θ_r would select $(F_r(\theta_r), \vec{P}_r(\theta_r))$. We note that while the Revelation Principle is defined under a general nonlinear transfer between the provider and the contracting group, under the beta distribution assumption of the contracting group's private information, a general nonlinear payment function is equivalent to a two-part tariff (Tirole 1988).

To ensure that group r selects $\big(F_r(\theta_r), \vec{P}_r(\theta_r)\big)$, the contracts $\{(F_r(\theta), \vec{P}_r(\theta)) : \theta \in [0,1]\}$ must satisfy the following incentive compatibility (IC) and individual rationality (IR) constraints.

IC and IR constraints. At time 0 given contract $(\vec{p}_r, f_r)$, the customer company anticipates its optimal utility from device (r, j) in period t to be $u_{rjt}(\vec{Q}_{rjt}(\vec{p}_r, f_r), \vec{p}_r, f_r)$, where utility function u_{rjt} is defined in (4). Define $U_{rjt}(\vec{p}_r, f_r, \theta)$ to be the customer's optimal utility from printer (r, j) if its private type were $\theta \in [0,1]$. From equations (4)–(6),

$$U_{rjt}(\vec{p}_r, f_r, \theta) = \tfrac{1}{2}[(\vec{\Psi}_r\theta + \vec{\mu}_{rj} + \epsilon_r\vec{1}_2 + \vec{\omega}_{rj} + \vec{\xi}_{rjt}) - \vec{p}_r]^T \vec{\Phi}[(\vec{\Psi}_r\theta + \vec{\mu}_{rj} + \epsilon_r\vec{1}_2 + \vec{\omega}_{rj} + \vec{\xi}_{rjt}) - \vec{p}_r] - f_r. \quad (7)$$

From Assumption 1, when selecting contract for contracting group r, the customer company makes its decision based on the expected utility from all printers within that group. Let $U_r(\vec{p}_r, f_r, \theta)$ denote the customer's aggregate utility from contracting group r with type θ, i.e., $U_r(\vec{p}_r, f_r, \theta) =$

$\sum_{j=1}^{N_r} \sum_{t=1}^{\tau_r} U_{rjt}(\vec{p}_r, f_r, \theta)$. Because θ is known to the customer, the randomness of $U_r(\vec{p}_r, f_r, \theta)$ comes entirely from the shocks $\vec{\xi}_{rjt}$, $\vec{\omega}_{rj}$, and ϵ_r.

Therefore, the IC constraint for contracting group r is that for all $\theta', \theta \in [0,1]$, $\mathbb{E}_{(\epsilon,\vec{\omega},\vec{\xi})} \Big[U_r\big(\vec{P}_r(\theta), F_r(\theta), \theta\big)\Big] \geq \mathbb{E}_{(\epsilon,\vec{\omega},\vec{\xi})} \Big[U_r\big(\vec{P}_r(\theta'), F_r(\theta'), \theta\big)\Big]$, where the expectation is taken over the shocks $\vec{\xi}_{rjt}$, $\vec{\omega}_{rj}$, and ϵ_r. The IR constraint for contracting group r is that for all $\theta \in [0,1]$, $\mathbb{E}_{(\epsilon,\vec{\omega},\vec{\xi})} \Big[U_r\big(\vec{P}_r(\theta), F_r(\theta), \theta\big)\Big] \geq U_r^0$, where U_r^0 is the reservation utility of contracting group r.

4.3.2 Provider's objective, risk-aversion, and the contract design problem.

The IC and IR constraints ensure that contracting group r with type θ chooses contract $\big(F_r(\theta), \vec{P}_r(\theta)\big)$. Recall that group r's type is $\boldsymbol{\theta}_r$ to the provider, the provider anticipates device (r,j) to print the optimal volume $\vec{Q}_{rjt}\big(\vec{P}_r(\boldsymbol{\theta}_r), F_r(\boldsymbol{\theta}_r)\big)$ in period t ($t = 1, 2, \ldots, \tau_r$). Therefore, the provider's period t profit is $Y_{rjt}\big(\vec{Q}_{rjt}(\vec{P}_r(\boldsymbol{\theta}_r), F_r(\boldsymbol{\theta}_r)), \vec{P}_r(\boldsymbol{\theta}_r), F_r(\boldsymbol{\theta}_r)\big)$ (equation (2)). The sources of randomness in this profit are the unknown type $\boldsymbol{\theta}_r$ and shocks ϵ_r, $\vec{\omega}_{rj}$, $\vec{\xi}_{rjt}$, and δ_{rjt}.

To capture risk-aversion of the provider, we assume that the provider has a mean-variance criterion with respect to the cumulative cash flows from all printers. Let $\vec{\mathbb{X}}$ denote the contract menus for all R contracting groups: $\vec{\mathbb{X}} = \{(F_r, \vec{P}_r) : r = 1, \ldots, R\}$, and define $\vec{\boldsymbol{\Theta}} = (\boldsymbol{\theta}_r : r = 1, \ldots, R)$ as the vector of R i.i.d. random variables. At $t = 0$, the provider's mean-variance criterion is

$$\begin{aligned} MV\big(\vec{\mathbb{X}}(\boldsymbol{\Theta})\big) = \mathbb{E}_{(\epsilon,\vec{\omega},\vec{\xi},\delta)} \left[\sum_{r=1}^{R} \sum_{j=1}^{N_r} \sum_{t=1}^{\tau_r} Y_{rjt} \left(\vec{Q}_{rjt}(\vec{P}_r(\boldsymbol{\theta}_r), F_r(\boldsymbol{\theta}_r)), \vec{P}_r(\boldsymbol{\theta}_r), F_r(\boldsymbol{\theta}_r) \right) \right] \\ - \lambda \mathrm{Var}_{(\epsilon,\vec{\omega},\vec{\xi},\delta)} \left(\sum_{r=1}^{R} \sum_{j=1}^{N_r} \sum_{t=1}^{\tau_r} Y_{rjt} \left(\vec{Q}_{rjt}(\vec{P}_r(\boldsymbol{\theta}_r), F_r(\boldsymbol{\theta}_r)), \vec{P}_r(\boldsymbol{\theta}_r), F_r(\boldsymbol{\theta}_r) \right) \right). \end{aligned} \tag{8}$$

The expectation and variance are taken over the random shocks $\vec{\xi}_{rjt}$, $\vec{\omega}_{rj}$, ϵ_r, and δ_{rjt} defined in (6). Variable $MV\big(\vec{\mathbb{X}}(\boldsymbol{\Theta})\big)$ is random due to the randomness in types $\boldsymbol{\Theta}$. Parameter λ reflects the service provider's attitude towards profit uncertainty—the greater the λ, the more risk-averse the provider. When $\lambda = 0$, the provider is risk-neutral.

The reasons for choosing mean-variance objective to capture risk-aversion are as follows. There are two parallel approaches for modeling risk-aversion: expected utility theory and risk measures. The two approaches strive towards the same goal (i.e., to capture decision maker's preferences towards uncertainty in outcomes), but are based on different assumptions and do not subsume each other. The expected utility theory was developed from the seminal work by von Neumann and Morgenstern (1947). The foundation of risk measures dates back to Markowitz (1952) and Artzner et al. (1999). In operations management, both approaches are represented. One of the simplest risk measures is variance in equation (8), which has been widely applied (see Kim et al. 2007, Van Mieghem 2007). We note that the mean-variance objective in (8) does have a connection with the utility theory. Specifically, if the utility function is quadratic and the uncertainty is Gaussian, then the expected utility maximization is equivalent to the maximization of mean-variance objective (8). In general, one can think of mean-variance as an approximation of more general utilities. In this paper, we use the mean-variance objective for its analytical simplicity.

The provider's objective is to find contracts $\vec{\mathbb{X}}$ that maximize its expected mean-variance criterion over $\vec{\boldsymbol{\Theta}}$: $\mathbb{E}_{\vec{\boldsymbol{\Theta}}}\left[MV(\vec{\mathbb{X}}(\vec{\boldsymbol{\Theta}}))\right]$. Therefore, the provider solves the following problem at $t = 0$.

$$\max_{\vec{\mathbb{X}}} \qquad \mathbb{E}_{\vec{\boldsymbol{\Theta}}}\left[MV\left(\vec{\mathbb{X}}(\vec{\boldsymbol{\Theta}})\right)\right] \tag{9a}$$

subject to, for $r = 1, 2, \ldots, R$,

$$\mathbb{E}_{(\epsilon,\vec{\omega},\vec{\xi})}\left[U_r\left(\vec{P}_r(\theta), F_r(\theta), \theta\right)\right] \geq \mathbb{E}_{(\epsilon,\vec{\omega},\vec{\xi})}\left[U_r\left(\vec{P}_r(\theta'), F_r(\theta'), \theta\right)\right] \quad \forall\, \theta', \theta \in [0,1], \tag{9b}$$

$$\mathbb{E}_{(\vec{\xi},\vec{\omega},\epsilon)}\left[U_r\left(\vec{P}_r(\theta), F_r(\theta), \theta\right)\right] \geq U_r^0 \qquad \forall \theta \in [0,1]. \tag{9c}$$

4.4 Equilibrium contract prices

Define $\vec{V}_{rj}$ as the variance of the monthly print volume from device (r, j) due to shock $\vec{\xi}_{rjt}$:

$$\vec{V}_{rj} = \begin{pmatrix} V_{rjb} & V_{rjbc} \\ V_{rjbc} & V_{rjc} \end{pmatrix} = \text{Var}_{\vec{\xi}}\left(\vec{Q}_{rjt}(\vec{p}_r, f_r)\middle|\, \epsilon_r, \vec{\omega}_{rj}\right) = \mathbb{E}_{\vec{\xi}}\left[\vec{\Phi}\vec{\xi}_{rjt}\vec{\xi}^{\,T}_{rjt}\vec{\Phi}\right], \tag{10}$$

where the last equality follows from equations (6) and (5).

Theorem 1 specifies the equilibrium fixed price $F_r^*(\theta)$ and variable prices $\vec{P}^{\,*}_r(\theta)$ for all $\theta \in [0,1]$.

Theorem 1. *Under the assumption that $\theta \sim Beta(1, 1/\kappa)$, the equilibrium contract prices for* *contracting group r are* ($r = 1, 2, \ldots, R$)*:*

$$\vec{P}^*_r(\theta) = \vec{s}^{\,0}_r + \left(2\lambda\vec{\Gamma}_r + \vec{\Phi}\right)^{-1}\left[\vec{\Phi}\vec{\Psi}_r(1-\theta)\kappa + 2\lambda(\vec{V}^s_r - \vec{V}^0_r\vec{s}^{\,0}_r)\right], \tag{11}$$

$$F^*_r(\theta) = \tfrac{1}{2}\left[\vec{\Psi}_r\theta + \vec{\mu}^0_r - \vec{P}^*_r(\theta)\right]^T\vec{\Phi}\left[\vec{\Psi}_r\theta + \vec{\mu}^0_r - \vec{P}^*_r(\theta)\right] - \vec{\Psi}^T_r\vec{\Phi}\int_0^\theta\left[z\vec{\Psi}_r - \vec{P}^*_r(z) + \vec{\mu}^0_r\right]dz - u_{0r}, \tag{12}$$

where we define,

$$\vec{s}^{\,0}_r = \frac{1}{N_r}\sum_{j=1}^{N_r}\vec{s}_{rj}, \qquad \vec{\mu}^0_r = \frac{1}{N_r}\sum_{j=1}^{N_r}\vec{\mu}_{rj}, \qquad \vec{V}^0_r = \frac{1}{N_r}\sum_{j=1}^{N_r}\vec{V}_{rj}, \qquad \vec{V}^s_r = \frac{1}{N_r}\sum_{j=1}^{N_r}\vec{V}_{rj}\vec{s}_{rj}, \tag{13a}$$

$$\vec{\Gamma}_r = \vec{V}^0_r + N_r\tau_r\sigma^2\vec{\Phi}\vec{1}_2\vec{1}^T_2\vec{\Phi} + \tau_r\vec{\Phi}\vec{\Sigma}\vec{\Phi}, \tag{13b}$$

$$u_{0r} = \frac{U_{0r}}{N_r\tau_r} - \frac{1}{2}\left(\sigma^2\vec{1}^T_2\vec{\Phi}\vec{1}_2 + \frac{1}{N_r}\sum_{j=1}^{N_r}\mathbb{E}_{\vec{\omega}}\left[\vec{\omega}^T_{rj}\vec{\Phi}\vec{\omega}_{rj}\right] - \frac{1}{N_r\tau_r}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\mathbb{E}_{\vec{\xi}}\left[\vec{\xi}^T_{rjt}\vec{\Phi}\vec{\xi}_{rjt}\right]\right). \tag{13c}$$

From Theorem 1, when the provider is risk-neutral ($\lambda = 0$), the equilibrium variable prices are

$$\vec{P}^*_{RN}(\theta) = \vec{s}^{\,0}_r + (1-\theta)\kappa\vec{\Psi}_r, \tag{14}$$

where $\vec{s}^{\,0}_r$ is defined in equation (13a) and it represents the average service cost of contracting group r; $\kappa \in (0, 1]$ is a parameter in the distribution $Beta(1, 1/\kappa)$ representing the presence of information asymmetry; and $\vec{\Psi}_r$ is the scaling factor in the group-level WTP (equation (6)). Thus, from equation (14), when the provider is risk-neutral, the variable prices are only affected by the service cost and information asymmetry. Note that the variable prices in (11) and (14) depend only on $\vec{\Psi}_r$ but not $\vec{\mu}_{rj}$, although both affect the expected group-level WTP $\vec{\Psi}_r\theta_r + \vec{\mu}_{rj}$ (equation (6)). The intuition is that under information asymmetry, the provider sets the variable prices to screen the customer's private type θ_r, and use the fixed price to extract the customer's utility. From equation (6), $\vec{\Psi}_r$ multiplies the private type θ_r, whereas $\vec{\mu}_{rj}$ provides a known shift to the customer's WTP. Thus, the provider extracts the shift caused by $\vec{\mu}_{rj}$ through the fixed prices, and screens the private information associated with $\vec{\Psi}_r$.

Comparing equation (11) with equation (14) reveals how a risk-averse provider ($\lambda > 0$) accounts for the profit variability when choosing the variable prices. The variable prices in equation (11) are affected by the variance of the print volume $\vec{V}_{rj}$ (equation (10)), in addition to service cost and asymmetric information. As shown in Appendix C.1, this difference in contract prices under the RA and RN assumptions leads to a simple way of testing for the provider's risk preference.

To understand how various shocks in the model translate into the contract prices of a risk-averse provider, consider the term $\vec{\Gamma}_r$ that multiplies the risk-aversion parameter λ in equation (11). From equation (13b), $\vec{\Gamma}_r$ is the sum of three components. The first one $\vec{V}_r^0$ is the average variance of the print volume in contracting group r (equations (13a) and (10)). It represents the first type of shocks that contribute to a variable cash flow: the temporal fluctuations in the monthly print volume. The second component $N_r \tau_r \sigma^2 \vec{\Phi} \vec{1}_2 \vec{1}_2^T \vec{\Phi}$ represents the group-level shock to the WTP of all printers within contracting group r. In this component, τ_r is the service horizon of contracting group r, σ^2 is the variance of the group-level shock ϵ_r (equation (6)), and $\vec{\Phi}$ is the company's price sensitivity matrix (equation (3b)). Recall that the shock ϵ_r applies to all N_r printers in the contracting group and throughout the service horizon τ_r. Finally, the third component $\tau_r \vec{\Phi} \vec{\Sigma} \vec{\Phi}$ represents the printer-specific shock to the WTP of device (r, j). In this component, $\vec{\Sigma}$ is the covariance matrix of the printer-specific WTP shock $\vec{\omega}_{rj}$ (equation (6)), which applies throughout the service horizon τ_r. When the provider is risk-averse ($\lambda > 0$), it considers the uncertainties in all three levels when designing the contract; and when the provider is risk-neutral ($\lambda = 0$), its optimal contract design does not depend on the uncertainties.

Finally, in deriving Theorem 1, we assumed that the equilibrium prices always satisfy the nonnegativity constraints. This assumption is consistent with similar studies in the literature (Miravete 2002) and it holds in our counterfactual analysis.

5 Econometric models and estimation method

In this section, we develop econometric models and describe our estimation method. The challenge of structural estimation under asymmetric information is that the predicted contracts (Theorem 1) and print volumes (equation (5)) depend on the company's private information θ_r, which is unobservable. As a result, we cannot estimate the model parameters by directly fitting the predicted contracts and volumes to the observations. While there are papers that do structural estimation under information asymmetry (Ivaldi and Martimort 1994, Wolak 1994), they all assume a risk-neutral principal. In this paper, however, we allow the provider to be risk-averse. This modeling flexibility significantly complicates the estimation approach used in earlier studies, rendering them computationally infeasible. Below we estimate model parameters following the "inversion" idea of Laffont and Martimort (2001) in Appendix 2.1 , as described next.

When the provider designs contracts to screen a contracting group based on its private information, the contract menu is such that the contracting group would self-select into the terms that are designed specifically for its private WTP. For a two-part tariff contract used in MPS, this is reflected in the one-to-one correspondence between the private information θ_r and the variable prices (equation (11)). Thus, observing the variable prices is as good as observing the private value information itself: from the realized variable prices of contracting group r, we can invert equation (11) to find its private WTP. This is the basic idea underlying our econometric models.

Let $\vec{p}_r^* = (p_{rb}^*, p_{rc}^*)^T$ denote the observed variable prices of contracting group r ($r = 1, 2, \ldots, R$). The following lemma finds $\vec{\Psi}_r\theta_r$ from the observed variable prices $\vec{p}_r^*$ by inverting equation (11).

Lemma 1. *The expected group-level WTP of contracting group r, $\vec{\Psi}_r\theta_r$, satisfies*

$$\vec{\Psi}_r\theta_r = \vec{\Psi}_r - \tfrac{1}{\kappa}\vec{\Phi}^{-1}\big(2\lambda\vec{\Gamma}_r + \vec{\Phi}\big)\big(\vec{p}_r^* - \vec{s}_r^{\,0}\big) + \tfrac{2\lambda}{\kappa}\vec{\Phi}^{-1}\big(\vec{V}_r^s - \vec{V}_r^0\vec{s}_r^{\,0}\big). \tag{15}$$

Using $\vec{\Psi}_r\theta_r$ from Lemma 1, we build three econometric models in §5.1. We discuss the identification issues and estimation methods in §5.2.

5.1 Econometric models

Using $\vec{\Psi}_r\theta_r$ from Lemma 1, the expected monthly print volume from printer (r, j) under contract $\left(f_r^*, \vec{p}_r^*\right)$ is given by (equations (5)–(6))

$$\mathbb{E}_{\vec{\xi}}\left[\vec{Q}_{rjt}\left(\vec{p}_r^*, f_r^*\right)\right] = \vec{\Phi}\big(\vec{\Psi}_r\theta_r + \vec{\mu}_{rj} - \vec{p}_r^*\big) + \vec{\Phi}\vec{1}_2\epsilon_r + \vec{\Phi}\vec{\omega}_{rj}. \tag{16}$$

Let $\vec{q}_{rj}^* = \left(q_{rjb}^*, q_{rjc}^*\right)^T$ denote the observed average monthly BW and color print volume from printer (r, j) ($r = 1, \ldots, R; j = 1, \ldots, N_r$). Then equation (16) offers theoretical predictions for

$\vec{q}^{\,*}_{rj}$, from which we build the following econometric models:

$$\vec{q}^{\,*}_{rj} = \vec{\Phi}\big(\vec{\Psi}_r\theta_r + \vec{\mu}_{rj} - \vec{p}^{\,*}_r\big) + \vec{\Phi}\vec{1}_2\epsilon_r + \vec{\Phi}\vec{\omega}_{rj}. \tag{17}$$

Because the print volume is two-dimensional, equation (17) yields two econometric models, with the average BW and color print volumes as the response variables, respectively. Note that econometric models from (17) have hierarchical errors: the group-level error $\vec{\Phi}\vec{1}_2\epsilon_r$, and the printer-level error $\vec{\Phi}\vec{\omega}_{rj}$. The group-level error $\vec{\Phi}\vec{1}_2\epsilon_r$ comes from the group-level shock $\epsilon_r \stackrel{iid}{\sim} N(0,\sigma^2)$. Because ϵ_r is shared by all printers within the contracting group, it introduces clustering among the printers and, thus, correlated print volumes. The printer-level error $\vec{\Phi}\vec{\omega}_{rj}$ comes from the printer-specific random shock $\vec{\omega}_{rj} \stackrel{iid}{\sim} N\big(0,\vec{\Sigma}\big)$. Putting together, hierarchical shocks ϵ_r and $\vec{\omega}_{rj}$ in the WTP translate into hierarchical econometric models that capture the unique hierarchical feature of MPS.

To ensure the robustness of our estimation results, we set up the following econometric model on the observed average variable payment, $VP^*_{rj} = (\vec{p}^{\,*}_r)^T\vec{q}^{\,*}_{rj}$, from equation (17):

$$VP^*_{rj} = (p^{\,*}_r)^T\big[\vec{\Phi}\big(\vec{\Psi}_r\theta_r + \vec{\mu}_{rj} - \vec{p}^{\,*}_r\big) + \vec{\Phi}\vec{1}_2\epsilon_r + \vec{\Phi}\vec{\omega}_{rj}\big]. \tag{18}$$

As in (16), the right side of model (18) is random with hierarchical errors $\vec{\Phi}\vec{1}_2\epsilon_r$ and $\vec{\Phi}\vec{\omega}_{rj}$.

Equations (17) and (18) provide three econometric models. Henceforth, we shall refer to BW and color parts of model (17) as the *BW model* and *Color model*, and refer to model (18) as the *Payment model*. These three models are built on the same assumptions, but provide different ways of connecting model parameters with data. Thus, estimating structural parameters from these econometric models allows a test on robustness of our estimation results.

Next we parameterize the group-level scaling factor $\vec{\Psi}_r$ and the printer-level deviation $\vec{\mu}_{rj}$ with the demographic information in the data. Let $\vec{W}_r$ and $\vec{Z}_{rj}$ denote the vectors of group- and printer-level demographics, respectively. To simplify the econometric models and for parsimony, let

$$\vec{\Psi}_r = (\psi_b,\psi_c)^T + \big(\vec{\beta}^T\vec{W}_r\big)\vec{1}_2, \quad \text{and} \quad \vec{\mu}_{rj} = \big(\vec{\gamma}^T\vec{Z}_{rj}\big)\vec{1}_2, \tag{19}$$

where ψ_b, ψ_c, $\vec{\beta}$ and $\vec{\gamma}$ are coefficients to be estimated. Thus, the BW and color components of $\vec{\Psi}_r$

differ in the intercept ψ_b and ψ_c, and the BW and color components of $\vec{\mu}_{rj}$ are the same. We do not include an intercept in $\vec{\mu}_{rj}$ to avoid identification issues because $\vec{\Psi}_r\theta_r$ already has a constant term. We do not differentiate $\vec{\beta}$ and $\vec{\gamma}$ in BW and color because that will significantly increase the number of parameters to estimate, thus weakening the statistical power of our estimation results.

For ease of interpretation, it is useful to rewrite the BW, Color, and Payment models in standard statistical model format with mean and additive error terms. In the following we first summarize the model parameters and data, then rewrite the econometric models (17) and (18) in Lemma 2.

Let $\vec{\Lambda} = (\phi_b, \phi_c, \phi_{bc}, \kappa, \lambda, \psi_b, \psi_c, \vec{\beta}, \vec{\gamma}, \sigma^2, \sigma_b^2, \sigma_c^2)$ denote the set of model parameters, where ϕ_b, ϕ_c, and ϕ_{bc} are the company's price sensitivities (equation (3b)), κ is the distribution parameter of the unknown type $\boldsymbol{\theta}_r$, λ is the provider's risk-aversion parameter (equation (8)), ψ_b, ψ_c, $\vec{\beta}$ and $\vec{\gamma}$ are the coefficients in equation (19), σ^2 is the variance of the group-level shock ϵ_r, and σ_b^2 and σ_c^2 are the diagonals in the covariance matrix $\vec{\Sigma}$ of the printer-specific shock $\vec{\omega}_{rj}$. Let $\vec{K}_{rj} = (f_r^*, \vec{p}_r^*, \vec{s}_{rj}, \vec{V}_{rj}, \vec{W}_r, \vec{Z}_{rj}, N_r, \tau_r)$ denote the set of known quantities of printer (r, j), where $(f_r^*, \vec{p}_r^*)$ is the observed contract, $\vec{s}_{rj}$ is the provider's service cost (equation (1)), $\vec{V}_{rj}$ is the observed covariance of the monthly print volume (equation (10)), $\vec{W}_r$ and $\vec{Z}_{rj}$ are the characteristics of the contracting group and the printer (equation (19)), and N_r and τ_r are the fleet size and service horizon.

Lemma 2. *The BW model, the Color model, and the Payment model can be rewritten as follows:* $q_{rjb}^* = G^1(\vec{K}_{rj}; \vec{\Lambda}) + u_r^1 + e_{rj}^1$, $q_{rjc}^* = G^2(\vec{K}_{rj}; \vec{\Lambda}) + u_r^2 + e_{rj}^2$, *and* $VP_{rj}^* = G^3(\vec{K}_{rj}; \vec{\Lambda}) + u_r^3 + e_{rj}^3$, *where* $G^i(\vec{K}_{rj}; \vec{\Lambda})$ $(i = 1, 2, 3)$ *are deterministic nonlinear functions defined in equations* (58), (60) *and* (63) *in Appendix A.2, and* u_r^i *and* e_{rj}^i *are group- and printer-level errors defined as follows:* $u_r^1 = (\phi_b + \phi_{bc})\epsilon_r$, $e_{rj}^1 = \phi_b\omega_{rjb} + \phi_{bc}\omega_{rjc}$, $u_r^2 = (\phi_c + \phi_{bc})\epsilon_r$, $e_{rj}^2 = \phi_{bc}\omega_{rjb} + \phi_c\omega_{rjc}$, $u_r^3 = [(\phi_b + \phi_{bc})p_{rb}^* + (\phi_c + \phi_{bc})p_{rb}^*]\,\epsilon_r$, *and* $e_{rj}^3 = (p_{rb}^*\phi_b + p_{rc}^*\phi_{bc})\,\omega_{rjb} + (p_{rb}^*\phi_{bc} + p_{rc}^*\phi_c)\,\omega_{rjc}$.

From Lemma 2, the BW, Color, and Payment models are nonlinear random-effects models. They differ not only in the response variables, but also in terms of the nonlinear function G^i (equations (58), (60) and (63) in Appendix A.2) and the distribution of their group- and printer-level errors u_r^i and e_{rj}^i. Specifically, unlike the BW and Color models, the Payment model has heteroskedastic errors across different contracting groups.

5.2 Identification and estimation method

All model parameters are identified in the Payment model. In the BW and Color model, however, parameters ψ_b and ψ_c in equation (19) cannot be individually identified. As a result, we cannot estimate ψ_b and ψ_c in BW and Color models.

For all three econometric models, we obtain the point estimates for the parameters using maximum log-likelihood estimation (MLE). We derive the log-likelihood functions following the approach in Hsiao (2014), and obtain the point estimates of model parameters using Matlab's optimization package. We use the parametric bootstrapping method to find the 95% confidence interval of the point estimates. All results are based on 300 bootstrap samples (Efron and Tibshirani 1994). Refer to Appendix B for more details on the estimation method.

6 Estimation results

In this section we present our estimation results from the three econometric models. We report that the estimate for the risk-aversion parameter λ is positive in all the three econometric models, supporting the RA assumption. We note that although the results presented in this section is based on the cost data set with 1,021 printers from 6 companies, in Appendix C.1 we provide alternative evidence for the provider's risk-aversion in the expanded data set with 3,075 printers from 26 companies. In Appendix C.2, we present out-of-sample tests of our model.

The econometric models make use of the demographic information on the group level, $\vec{W}_r$, and the printer level, $\vec{Z}_{rj}$. The demographic information contained in the data set includes each printer's product model and the company it belongs to. A total of 35 printer models are observed among the 1,021 printers, with the three most widely installed models accounting for 62.1% of the population. We label the printer models by their prevalence, with model 1 being the most popular and model 35 the least. Because only 12 printer models have more than 10 printers observed in the data set, the printer-level characteristics $\vec{Z}_{rj}$ comprises of 13 dummy variables. Dummy variables 1 through 12 correspond to printer models 1 through 12 with at least 10 observations, and dummy variable 13 corresponds to all other printer models.

The group-level vector $\vec{W}_r$ has several parts. First, it contains the respective numbers of models 1, 2, and 3 installed in group r. Let n_r^k denote the number of model k printers in group r ($k = 1, 2, 3$).

To avoid extreme values in the covariates, we use $\log(n_r^k + 1)$ in $\vec{W}_r$.

Second, $\vec{W}_r$ includes dummy variables to indicate the company that owns group r. There are six companies in the data set, but due to the intercept for $\vec{\Psi}_r$ in equation (19), $\vec{W}_r$ has only five company dummy variables, which are labeled company 1 through 5. The company with the most number of printers (company 6) is used as the baseline to estimate ψ_b and ψ_c.

Third, based on discussions with Xerox, a key distinction among the printers is whether it is a solid-ink printer or a laser printer. Thus, $\vec{W}_r$ contains the proportion of solid-ink printers within group r. Finally, the number of printers within a group could play an important role in the contracting stage. Thus, a component of $\vec{W}_r$ is $\log(N_r)$, where N_r is the fleet size of group r.

We present estimation results for the BW, Color, and Payment models in Table 2. All the models generate consistent estimates. Specifically, the estimate for the provider's risk-averse parameter λ is positive and is statistically significant, implying that the provider is risk-averse in MPS.

The estimates for the price sensitivity of BW, ϕ_b, and color, ϕ_c, printing are around 1, and the cross-sensitivity ϕ_{bc} is around -1 in all models. From equation (5), this implies that, on average, each cent increase in the BW and color variable price would cause the BW and color print volume to decrease by ten pages, and that BW and color printing are substitutes to each other.

Another noteworthy result is the estimates for β. From equation (19), the higher $\vec{\beta}$ is, the higher the scaling factor $\vec{\Psi}_r$, and, thus, the higher the expected group-level WTP for the printing service, $\vec{\Psi}_r\theta_r$ (equation (6)). From Table 2, the coefficient of $\log(N_r)$ is negative in the payment model, meaning that the company is willing to pay less as the number of printers increases, consistent with the intuition of quantity discount in the contracting stage. Furthermore, the coefficient of $\log(n_r^1 + 1)$ is negative in the payment model, and the coefficient of $\log(n_r^2 + 1)$ is positive in the BW and payment models. This implies that the fewer Model 1 printers and the more Model 2 printers are installed in a group, the more the company is willing to pay for the service. As we shall discuss in §7, this result has significant operational implications.

Further, there is significant information asymmetry between Xerox and its customers. Recall from §4.3.1 that $\kappa \in (0, 1]$ indicates the level of information asymmetry. The greater κ is, the greater the information asymmetry. The estimate of κ is close to 1, indicating that Xerox has no knowledge of the private information associated with group r, θ_r, and assumes that θ_r is equally likely to take any value on $[0, 1]$. In §7, we discuss the effects of information asymmetry on the contract prices when the provider changes from risk-neutral to highly risk-averse.

Table 2. Estimation Results

C.I. stands for confidence interval, which are computed from 300 bootstrap samples. n_r^k $(k = 1, 2, 3)$ is the number of printer model k installed in group r. Company 6 is used as baseline in the estimation. Parameters ψ_b and ψ_c cannot be identified in the BW and color model.
Note: †: $\bar{\kappa} \in (-\infty, \infty)$ is related to κ through a logistic transformation: $\kappa = 1/\left(1 + e^{-\bar{\kappa}}\right)$. In the estimation, we estimate $\bar{\kappa}$, find its confidence intervals, and then use the logistic function to find the estimate and confidence intervals for κ. This transformation of the confidence interval is valid because the logistic function is monotone and maps from $[0, 1]$ to $(-\infty, \infty)$ (see, e.g., Section 5.2.1 of Agresti (2002) for the use of this idea in logistic regression). The estimates for κ are as follows: BW model: 0.99991 with confidence interval [0.99967, 0.99997]; Color model: 0.99994 with confidence interval [0.99988, 0.99997]; Payment model: 0.9998 with confidence interval [0.9997, 0.9999].

	BW model		Color model		Payment model	
	Estimate	95% C.I.	Estimate	95% C.I.	Estimate	95% C.I.
λ	0.6	$[0.4, 0.8]$	0.8	$[0.3, 1.9]$	0.7	$[0.5, 0.9]$
ϕ_b	1.6	$[1.2, 2.2]$	0.4	$[0.1, 1.4]$	0.7	$[0.5, 1.0]$
ϕ_c	1.3	$[0.8, 2.1]$	0.6	$[0.5, 0.7]$	1.0	$[0.7, 1.4]$
ϕ_{bc}	-1.3	[-1.7, -0.9]	-0.8	[-1.0, -0.6]	-0.8	[-1, -0.5]
ψ_b	-	-	-	-	14	$[6, 29]$
ψ_c	-	-	-	-	23	$[18, 29]$
$\bar{\kappa}^\dagger$	9.3	$[8.0, 10.5]$	9.8	$[9.0, 10.5]$	8.5	$[8.0, 9.1]$
$\kappa^\dagger$	$1.00^\dagger$	$[1.00, 1.00]^\dagger$	$1.00^\dagger$	$[1.00, 1.00]^\dagger$	$1.00^\dagger$	$[1.00, 1.00]^\dagger$
$\log(N_r)$: β	-0.3	[-1.7, 1.2]	0.1	[-2.6, 2.8]	-1.4	[-2.3, -0.5]
Solid-ink%: β	0.4	[-1.6, 2.5]	-1.0	[-2.6, 0.5]	0.0	[-1.9, 1.9]
$\log(n_r^1 + 1)$: β	-0.5	[-2, 1]	-0.7	[-3, 2]	-1.2	[-2, -0.4]
$\log(n_r^2 + 1)$: β	1.5	$[0.1, 2.8]$	1.2	[-0.6, 2.9]	2.0	$[0.8, 3.2]$
$\log(n_r^3 + 1)$: β	-2	[-5, 1]	0.7	[-4, 6]	-0.5	[-1,0.2]
Company 1: β	-0.2	[-2.9, 2.6]	-1.2	[-2.5, 0.2]	-1.4	[-2.7, -0.1]
Company 2: β	-0.1	[-1.3, 1.2]	-0.3	[-1.2, 0.6]	0.5	[-1.1, 2.0]
Company 3: β	-0.6	[-1.8, 0.7]	2.0	[1.0, 2.9]	1.9	$[1.3, 2.6]$
Company 4: β	-0.4	[-1.8, 1.0]	1.4	[-0.3, 3.1]	1.6	$[0.4, 2.8]$
Company 5: β	0.9	[-1.3, 3.1]	1.6	$[0.4, 2.8]$	0.7	$[0.1, 1.2]$
Model 1: γ	0.9	[-0.9,2.7]	0.2	[-1.0, 1.4]	1.8	$[1.2, 2.4]$
Model 2: γ	-0.9	[-2.2,0.4]	0.6	[-0.3, 1.5]	-0.5	[-1.3, 0.2]
Model 3: γ	-0.9	[-2.0,0.2]	0.9	[-1.1, 2.9]	-0.8	[-3.5, 1.9]
Model 4: γ	1.6	[-0.01,3.2]	-0.8	[-1.7, 0.01]	0.2	[-0.9, 1.3]
Model 5: γ	0.9	[-0.9,2.7]	1.1	[0.03, 2.2]	-0.6	[-1.8, 0.7]
Model 6: γ	-1.7	[-4.1,0.8]	-0.5	[-1.7, 0.7]	2.0	$[0.9, 3.0]$
Model 7: γ	0.3	[-1.2,1.7]	-1.3	[-2.2, -0.4]	1.7	[-0.9, 4.2]
Model 8: γ	-0.8	[-2.7,1.1]	-1.7	[-2.9, -0.5]	1.8	$[1.0, 2.6]$
Model 9: γ	-0.3	[-2.2,1.6]	-0.6	[-1.6, 0.3]	1.1	[-1.6, 3.8]
Model 10: γ	-1.7	$[-3.3, -0.1]$	0.8	$[-0.3, 2.0]$	1.9	$[-0.1, 3.9]$
Model 11: γ	0.3	$[-1.6, 2.1]$	0.5	$[-1.0, 1.9]$	0.0	$[-2.4, 2.4]$
Model 12: γ	0.5	$[-1.5, 2.4]$	-1.8	$[-3.3, -0.4]$	-1.3	$[-3.5, 1]$
Other models: γ	-1.8	$[-3.8, 0.2]$	-1.7	$[-2.8, -0.5]$	0.5	$[-2.9, 2.0]$
σ^2	0.7	$[0.2, 3.3]$	0.3	$[0.0, 16.6]$	0.2	$[0.1, 0.5]$
σ_b^2	0.9	$[0.4, 2.1]$	0.9	$[0.5, 1.7]$	1.3	$[0.7, 2.4]$
σ_c^2	3.0	$[1.2, 7.8]$	1.2	$[0.7, 2.2]$	3.1	$[1.3, 7.7]$

Finally, we note that the variance of the group-level shock, σ^2, is comparable to the variance of the printer-level shocks, σ_b^2 and σ_c^2, which are responsible to the inter-printer variation. This implies that there is some clustering among printers within the same contracting group, thus lending support to our hierarchical model and the modeling assumption about contracting groups.

7 Counterfactual analyses

7.1 Implications of the commission policy

According to Xerox, an important motivation for the commission holdback policy is to avoid inferior performance when sales team signs contracts. But, this policy might also cause the sales representatives to behave in a risk-averse manner. How does this risk-averse behavior affect Xerox's expected earnings? What would happen if Xerox could make its sales managers less risk-averse, for example, by revoking its "commission holdback" policy? In this subsection, we answer these questions by using the risk-aversion parameter λ as a proxy for Xerox's commission policy.

Two factors are essential for the equilibrium solution in this paper: the provider's risk-aversion, and information asymmetry. To separate their effects, we use the first-best contracts under symmetric information as a benchmark. The following lemma specifies the first-best contracts.

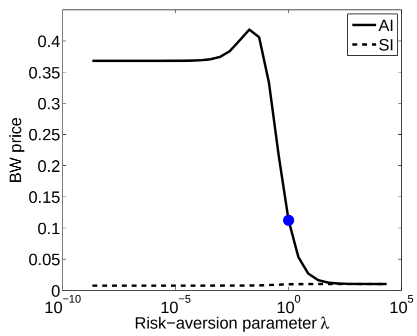
Lemma 3. *The first-best variable prices and fixed price of contracting group r are*

$$\vec{P}_r^{FB}(\theta_r) = \vec{s}_r^{\,0} + 2\lambda\left(2\lambda\vec{\Gamma}_r + \vec{\Phi}\right)^{-1}\left(\vec{V}_r^s - \vec{V}_r^0\vec{s}_r^{\,0}\right), \tag{20}$$

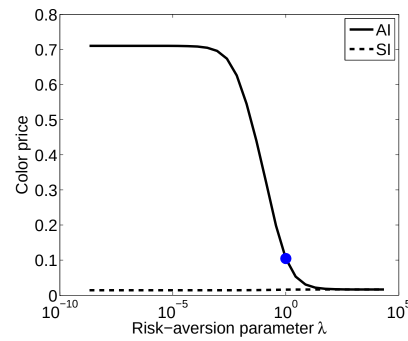
$$F_r^{FB}(\theta_r) = \tfrac{1}{2}\big[\vec{\Psi}_r\theta + \vec{\mu}_r^0 - \vec{P}_r^{FB}(\theta)\big]^T\vec{\Phi}\big[\vec{\Psi}_r\theta + \vec{\mu}_r^0 - \vec{P}_r^{FB}(\theta)\big] - u_{0r}. \tag{21}$$

Comparing the first-best fixed price $F_r^{FB}(\theta_r)$ in equation (21) with the equilibrium fixed price $F_r^*(\theta_r)$ under asymmetric information (equation (12)), we confirm that, under symmetric information, the provider extracts all surplus from the customer company. However, under asymmetric information, in equilibrium, the customer retains positive information rent $\vec{\Psi}_r\vec{\Phi}\int_0^{\theta_r}\left[z\vec{\Psi}_r + \vec{\mu}_r^0 - \vec{P}_r^*(z)\right]dz$ from contracting group r.

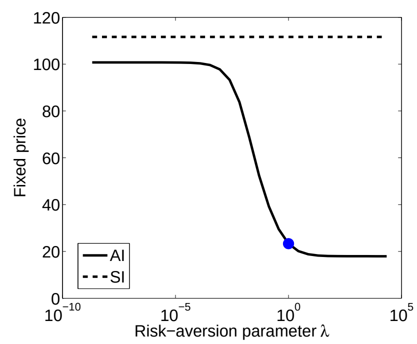
Figure 7.1 presents the effects of the risk-aversion parameter λ on the asymmetric information equilibrium, the first-best contract prices, and the customer's information rent. Because Xerox can influence the sales team's risk attitude via its commission policy, this figure provides insights on what would happen if the commission holdback policy were changed. In panels (a)–(c) of Figure 7.1, the dashed lines are the first-best prices from Lemma 3, and the solid lines are the equilibrium prices under asymmetric information from Theorem 1. The filled circle corresponds to the estimated values of λ for Xerox. We note that Figure 7.1 does not account for consequences that are outside the scope of the paper. For example, holdback policy may affect the sales managers' effort, which may subsequently affect Xerox's expected earnings and contract prices.



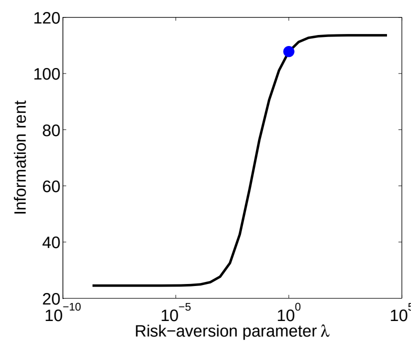
(a) BW price



(b) Color price



(c) Fixed price



(d) Information rent

Curves labeled "SI" and "AI" are the equilibrium first-best prices and asymmetric information prices, respectively. The filled circle corresponds to the estimated value of λ for Xerox. For confidentiality reasons, all prices are scaled. BW and color prices are scaled by the same factor. The fixed price and information rent are scaled differently.

Figure 3. Effects of risk-aversion

We start by discussing the first-best contract prices. From Figures 3a–3c, the first-best contract prices are hardly affected by the provider's risk-aversion. Intuitively, this is because under risk-neutrality ($\lambda = 0$), the first-best variable prices equal the variable service costs $\vec{s}_r^{\,0}$ (equation (20)). Because the monthly profit variability is proportional to the variable price net of the service cost, when the provider becomes risk-averse and aims to reduce profit variability, the variable prices should still be very close to the variable service costs. Furthermore, when the variable prices do not change, the customer's surplus remains constant. Because the first-best fixed price equals to the customer's surplus, the fixed-price is also insensitive to λ. To sum up, under symmetric information, although Xerox's commission policy might affect the risk attitude of the sales teams, it does not affect the equilibrium contracts significantly.

This is no longer the case under asymmetric information. From Figures 3a–3c, the equilibrium contract prices under asymmetric information depend on the provider's risk-aversion in complex ways, implying that changes in the commission policy might translate into changes in contracts. Below we discuss these results in detail.

First, consider the effects of risk-aversion on variable prices. From Figures 3a–3b, the variable prices under asymmetric information are higher than the first-best prices. This confirms the standard result that information asymmetry distorts the contract prices upward as the provider attempts to reduce the information rent.

Second, when the provider is either risk-neutral or highly risk-averse, the variable prices are insensitive to local changes in the provider's risk-aversion (left and right ends of dark lines in panels (a) and (b)). Specifically, when the provider is highly risk-averse, the variable prices converge to their first-best values close to the service cost (equation (11)). Intuitively, in this case, the provider's primary goal becomes minimizing the profit variability. Thus, it sets the variable prices equal to the variable service costs.

Third, when the provider is moderately risk-averse (middle parts of dark lines in panels (a) and (b)), the BW price first rises and then falls in λ, whereas the color price decreases monotonically. This difference stems from the portfolio effects due to the customer's two-dimensional demand in BW and color printing. If the customer's demand is uni-dimensional, the variable price always decreases as the provider becomes more risk-averse. But when the customer's demand is two-dimensional, the provider can play with the two variable prices when maximizing its objective, thus leading to a different dependence on λ.

Next, consider the effects of risk-aversion on the fixed price and information rent. From Figures 3c and 3d, the fixed price under asymmetric information falls in λ, and the customer's information rent increases with it. Intuitively, this is because the fixed price is the customer's surplus less its information rent. When the provider becomes increasingly risk-averse, it puts more weight in minimizing the variability of the profit rather than minimizing the customer's information rent. Thus, the customer retains more information rent, and the fixed price decreases.

To summarize, for a counterfactual where Xerox revokes its commission holdback policy and the sales team becomes less risk averse, Xerox would extract higher surplus from the customers and enjoy higher expected earnings. Both fixed and variable prices to the customers would increase. As we noted earlier, these results do not account for other consequences of revoking the holdback policy that are outside the scope of the analysis.

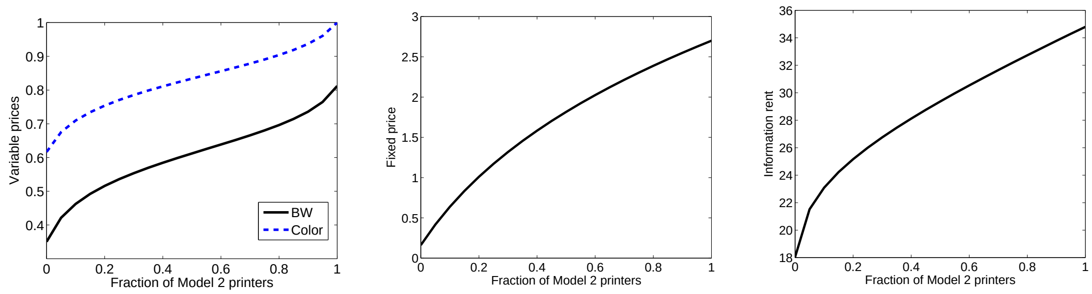
7.2 Implications of the customer's printer selection

The printers at a customer are installed before contracting with Xerox MPS, and are taken as fixed when Xerox sales team do the pricing of the service. Suppose there was a way that Xerox could change the customer's model selection and, thus, integrate MPS with product offering. How would the optimal model selection depend on the provider's risk aversion? Would the changed model selection benefit or hurt the customers?

Before carrying out the analysis, it is useful to establish some intuition about the effects of model selection on Xerox's objective and the customer's information rent. Intuitively, different printer models cost differently in services. From equations (6) and (19) and Table 2, printer model selection also affects the customer's WTP for the service. By affecting service cost and WTP, model selection influences the contract design problem, which depends on Xerox's risk-aversion. In the following, we shed light on the net effect of model selection on equilibrium contracts, Xerox's objective, and the customer's information rent by conducting counterfactual analysis.

Because Xerox negotiates on each contracting group individually, it is reasonable to analyze each contracting group individually. There are 44 contracting groups in the data set. For each one, we fix the total number of printers as observed and vary its model selection. For illustration purposes, we consider the effects of changing the fraction of Model 2 printers. When decreasing the fraction of Model 2, we replace it with Model 1; when increasing the fraction of Model 2, we use it to replace Model 1 printers, and then Model 3, 4, ... printers. For all model parameters, we use the estimates from the Payment model in Table 2.

To study the role of risk-aversion, we allow the risk-aversion parameter to take three values: $\lambda = 0$, which represents a risk-neutral provider, $\lambda = 0.66$, which is the provider's estimated level of risk-aversion, and $\lambda = 65.5$, which represents a highly risk-averse provider.



(a) Variable prices **(b)** Fixed price **(c)** Information rent

Panels (a) and (b) present the results for variable and fixed prices, respectively. The prices are rescaled for confidentiality reasons. Panel (c) shows the customer's information rent. This figure is representative for all contracting groups and for $\lambda = 0,\ 0.66,\ 65.5$.

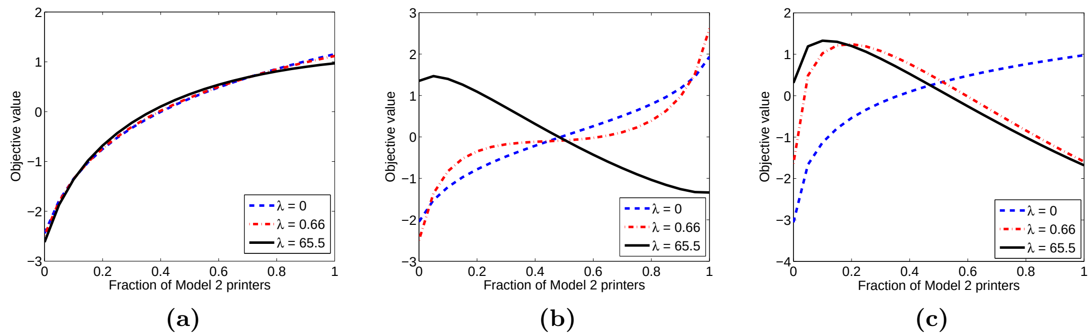
Figure 4. Effects of the fraction of Model 2 printers on contract prices and a customer's information rent

Figure 7.2 shows the effects of model selection on contract prices and the customer's information rent from a contracting group. We found that all contracting groups exhibit the same dependence on their Model 2 fractions, and this dependence does not change qualitatively with risk-aversion levels $\lambda = 0,\ 0.66,\ 65.5$. Thus, Figure 7.2 is representative of all contracting groups and of different λ values. The x-axis in the three panels is the fraction of Model 2 printers in a contracting group.

From Figures 4a–4b, for any risk-aversion and any contracting group, by increasing the fraction of Model 2 printers in a contracting group, the provider is able to charge higher prices for the same service. Interestingly, the higher prices do not make the customer worse off. From Figure 4c, the customer's information rent also increases.

It is important to realize that Figure 4c does not mean that the customer company is "leaving money on the table" by having a low fraction of Model 2 printers. Figure 4c is drawn for Xerox MPS, but the printers of a contracting group may have been installed long before the customer company contracted with Xerox. Therefore, the observed printer selection can be a historical artifact. At the time the customer had purchased printers (which may even be non-Xerox manufactured), it did not anticipate buying Xerox service in the future.

Would Xerox prefer a high fraction of Model 2 printers given the higher contract prices? Interestingly, the answer depends on the level of its risk-aversion and contracting group particulars. We compute the provider's objective for each contracting group under different λ values and observe three types of dependence among the 44 contracting groups. Figure 7.2 shows these three types of dependence by plotting three representative contracting groups, one for each type, in panels (a)–(c). The x-axis is the Model 2 fraction in a contracting group.



Panels (a)–(c) show the three types of dependence observed in the data. Figures are constructed from three particular contracting groups out of 44 in total. Lines in panels (a)–(c) correspond to the mean-variance objective under different λ, and are centered and rescaled for ease of comparison and exposition.

Figure 5. Effects of the fraction of Model 2 printers on the provider's objective

From Figure 7.2, when the provider is risk-neutral ($\lambda = 0$), its objective value strictly increases with Model 2 fraction in all contracting groups (dashed lines in panels (a)–(c)). When the provider is moderately risk-averse ($\lambda = 0.66$), it prefers a high Model 2 fraction for contracting groups in panels (a) and (b). But for contracting groups in panel (c), it prefers a low Model 2 fraction instead (dash-dot lines). When the provider is highly risk-averse ($\lambda = 65.5$), the provider prefers a high Model 2 fraction for contracting groups in panel (a), but prefers low Model 2 fraction for contracting groups in panels (b) and (c) (solid lines in Figure 7.2).

Figure 7.2 reflects the tension between rent extraction and risk-aversion when the provider designs contracts under asymmetric information. A risk-neutral provider has the only goal of minimizing the customer's information rent and, thus, maximizing its expected earnings. When the provider is risk-averse, it has a dual objective of increasing the expected earnings, while controlling its variability. The relative importance of the two depends on the risk-aversion parameter λ. The higher the value of λ, the more emphasis the provider puts on controlling the earnings variability. From Figure 7.2, for contracting groups in panel (a), risk-aversion does not cause the provider's preference to deviate from its risk-neutral counterpart. But for contracting groups in panels (b) and (c), the provider's preference changes when the provider is sufficiently risk-averse.

The conclusions about the effects of risk-aversion on the provider and customer's printer model preferences are as follows. If the provider is risk-averse, for some contracting groups, the preferences of the provider and the customer for Model 2 differ. If the provider is risk-neutral, then the preferences of the provider and the customers for Model 2 are aligned. Therefore, if Xerox reduces its risk aversion by revoking the commission holdback policy, then we expect to see more Model 2 being used, and, by our model, this will be a win-win for both the provider and the customer.

8 Concluding Remarks

In this paper, we studied an important example of B2B infrastructural services—managed print services. Even though we based our conclusions on Xerox data, our results and insights are relevant for other B2B service industries. The importance of these results will grow as the trend towards servitization continues.

We find evidence that Xerox's sales team is risk-averse. This is an important finding for both theory and practice. For theory, this informs the choice of modeling assumptions in academic research on service contracting. Assuming a provider's risk-neutrality is preferable for tractability. However, we have now provided empirical evidence that this simplifying assumption is false in practice. Therefore, modelers must consider the tradeoff between model simplicity and model realism for B2B service contracts more carefully.

We reviewed traditional explanations of firms' risk-aversion with Xerox managers and concluded that these explanations do not apply well in their setting. Instead, we have identified a company policy that can explain the apparent risk-aversion of the sales team—the commission-holdback policy. Using Xerox data, we conducted counterfactual analyses to determine the consequences of rescinding the commission-holdback policy and reducing risk-aversion of the sales team. The consequences include service provider's valuation, contract pricing, and printer model selection.

Specifically, a risk-averse provider uses prices to both extract informational rent from its customers and to control the variance of its cash flows. Therefore, a risk-averse provider does not extract as much value from the customers as a risk-neutral provider would. Thus, reducing risk-aversion of the sales team could increase Xerox's valuation.

For contract pricing, lower risk-aversion of the sales team translates into higher fixed prices.

Whether variable prices increase or decrease is not clear, because there are two dimensions to variables prices (BW and color). We show that, due to portfolio effects, it can happen that with lower risk-aversion the prices for BW and color print services will move in opposite directions.

We showed that although a risk-neutral provider and its customers share preferences for a printer model, a risk-averse sales team prefers differently. Therefore, if the sales team become less risk-averse, customers and the provider may switch to the mutually preferred printer model over time.

To derive these results, we built a theoretical model of MPS contracting under asymmetric information, where the provider has mean-variance objective. A salient feature of this model, which also applies to other B2B service contracts, is that there is an agency problem at the customer level. The representatives of the customer company who negotiate contracts with the service provider are not the same as the employees of the customer company who later use the service. This feature is not present in typical B2C contracting settings and the corresponding academic literature. We solved the theoretical model for the optimal contracts and used the results to build several structural econometric models. These models possess a hierarchical structure due to the customer agency problem mentioned above. Using these econometric models, we estimated risk-aversion, degree of asymmetric information, and other parameters, based on Xerox data. Our modeling and structural econometric analyses illustrate the use of B2B service-contracting data and will help to stimulate additional research in this important area.

Acknowledgments

This project is funded by a Xerox Foundation grant.

References

Agresti, A. 2002. *Categorical Data Analysis*. 2nd ed. New York: John Wiley & Sons.

Allen, D.W., D. Lueck. 1995. Risk preferences and the economics of contracts. *Am. Econ. Rev.* **85(2)** 447–451.

Artzner, P., F. Delbaen, J.M. Eber, D. Heath. 1999. Coherent measures of risk. *Math. Fin.* **9(3)** 203–228.

Barseghyan, L., F. Molinari, T. O'Donoghue, J.C. Teitelbaum. 2013. The nature of risk preferences: Evidence from insurance choices. *Am. Econ. Rev.* **103(6)** 2466–2529.

Chiappori, P.A., B. Salanié. 2000. Testing for asymmetric information in insurance markets. *J. Polit. Econ.* **108** 56–78.

Cohen, M.A., N. Agrawal, V. Agrawal. 2006. Winning in the aftermarket. *Harv. Bus. Rev.* **May** 129–138.

Crawford, S.G., M. Shum. 2007. Monopoly quality degradation and regulation in cable television. *J. Law Econ.* **50** 181–219.

DeMarzo, P., D. Duffie. 1992. Corporate financial hedging with long-term linear managerial compensation. Tech. rep., Stanford Graduate School of Business.

Dey, D., M. Fan, C. Zhang. 2010. Design and analysis of contracts for software outsourcing. *Inf. Syst. Res.* **21(1)** 93–114.

DiRomauldo, A., V. Gurbaxani. 1998. Strategic intent for IT outsourcing. *Sloan Manage. Rev.* **39(4)** 67–80.

Efron, B., R. J. Tibshirani. 1994. *An introduction to the boostrap*. CRC Press.

Fitoussi, D., V. Gurbaxani. 2012. IT outsourcing contracts and performance measurement. *Inf. Syst. Res.* **23(1)** 129–143.

Froot, K.A., D.S. Scharfstein, J.C. Stein. 1993. Risk management: Coordinating corporate investment and financing policies. *J. Fin.* (48(5)) 1629–1658.

Gopal, A., K. Sivaramakrishnan, M.S. Krishnan, T. Mukhopadhyay. 2003. Contracts in offshore software development: An empirical analysis. *Manag. Sci.* **16(12)** 1671–1683.

Guajardo, J.A., M.A. Cohen, S. Kim, S. Netessine. 2012. Impact of performance-based contracting on product reliability: An empirical analysis. *Manag. Sci.* **58(5)** 961–979.

Hodge, F., A. Winn. 2012. Do compensation clawback and holdback provisions change executive reporting choices? SSRN. URL `https://ssrn.com/abstract=2104205`.

Hsiao, C. 2014. *Analysis of panel data*, vol. 54. Cambridge University Press.

Ivaldi, M., D. Martimort. 1994. Competition under nonlinear pricing. *Annales d'Economie et de Statistique* **34** 71–114.

Kim, S., M.A. Cohen, S. Netessine. 2007. Performance contracting in after-sales service supply chains. *Manag. Sci.* **53(12)** 1843–1858.

Kim, Y., R. Telang, W.B. Vogt, R. Krishnan. 2010. An empirical analysis of mobile voice service and sms: A structural model. *Manag. Sci.* **56** 234–252.

Lacity, M.C., L.P. Wilcocks, D.F. Feeny. 1995. IT outsourcing: Maximize flexibility and control. *Harv. Bus. Rev.* **May-June** 84–93.

Laffont, J.J., D. Martimort. 2001. *The Theory of Incentives: The Principal-Agent Model*. Princeton University Press.

Lafontaine, F. 1992. Agency theory and franchising: Some empirical results. *Rand J. Econ.* **23(2)** 263–283.

Lambrecht, A., K. Seim, B. Skiera. 2007. Does uncertainty matter? Consumer behaviour under three-part tariffs. *Market Sci.* **26(5)** 698–710.

Markowitz, H.M. 1952. Portfolio selection. *J. Fin.* **7(1)** 77–91.

Masten, S.E., S. Saussier. 2000. Econometrics of contracts: An assessment of developments in the empirical literature on contracting. *Revue d'économie industrielle* **92(1)** 215–236.

Miravete, E.J. 2002. Estimating demand for local telephone service with asymmetric information and optional calling plans. *Rev. Econ. Stud.* **69** 943–971.

Oliva, R., R. Kallenberg. 2003. Managing the transition from products to services. *Int. J. Serv. Ind. Manag.* **14(2)** 160–172.

Perrigne, I., Q. Vuong. 2011. Nonparametric identification of a contract model with adverse selection and moral hazard. *Econometrica* **79** 1499–1539.

Tirole, J. 1988. *The theory of industrial organization*. MIT Press.

Van Mieghem, J.A. 2007. Risk mitigation in newsvendor networks: Resource diversification, flexibility, sharing, and hedging. *Manag. Sci.* **53(8)** 1269–1288.

von Neumann, J., O. Morgenstern. 1947. *Theory of games and economic behavior*. Princeton University Press.

Wolak, F.A. 1994. An econometric analysis of the asymmetric information, regulator-utility interaction. *Annales d'Economie et de Statistique* **34** 13–69.

Appendix

A Proofs

A.1 Proof of Theorem 1.

To prove Theorem 1, we first present and prove two lemmas. Lemma 4 states that the portfolio optimization problem (9) with R contracting groups is equivalent to R individual screening problems. Lemma 5 further simplifies the optimization problem of each contracting group. In the analysis, we use $h(\cdot)$ to denote the probability density function of the Beta distribution $Beta(1, 1/\kappa)$. (Here we assume that no exclusion of printers is allowed. This is consistent with practice where when a MPS provider contracts with an companies company, all the company's printers are covered.)

Lemma 4. *The provider's portfolio optimization with all R contracting groups reduces to a set of optimization problems, one per contracting group. The optimal contract design for contracting group r ($r = 1, 2, \ldots, R$) solves the following optimization:*

$$\max_{\{(F_r(\theta), \vec{P}_r(\theta)): \theta \in [0,1]\}} \quad \mathbb{E}_{\boldsymbol{\theta}_r} \left[MV_r(F_r(\boldsymbol{\theta}_r), \vec{P}_r(\boldsymbol{\theta}_r)) \right] \tag{22a}$$

$$s.t. \qquad \mathbb{E}_{(\vec{\xi}, \vec{\omega}, \epsilon)} \left[U_r\big(\vec{P}_r(\theta'), F_r(\theta'), \theta\big) \right] \leq \mathbb{E}_{(\vec{\xi}, \vec{\omega}, \epsilon)} \left[U_r\big(\vec{P}_r(\theta), F_r(\theta), \theta\big) \right] \quad \forall\, \theta, \theta' \in [0, 1], \tag{22b}$$

$$\mathbb{E}_{(\vec{\xi}, \vec{\omega}, \epsilon)} \left[U_r\big(\vec{P}_r(\theta), F_r(\theta), \theta\big) \right] \geq U_{0r} \qquad \forall \theta \in [0, 1], \tag{22c}$$

where

$$\begin{aligned} MV_r(F_r(\theta), \vec{P}_r(\theta)) = {} & N_r \tau_r \big[\big(\vec{P}_r(\theta) - \vec{s}^{\,0}_r\big)^T \vec{\Phi} \big(\vec{\Psi}_r \theta - \vec{P}_r(\theta) + \vec{\mu}^0_r + F_r(\theta) \big) \big] \\ & - \lambda N_r \tau_r \Bigg\{ N_r \tau_r \left(\left[\vec{P}_r(\theta) - \vec{s}^{\,0}_r \right]^T \vec{\Phi} \vec{1}_2 \right)^2 \sigma^2 + \tau_r \left(\vec{P}^{\,T}_r(\theta) \vec{\Phi} \vec{\Sigma} \vec{\Phi} \vec{P}_r(\theta) - 2 \vec{P}^{\,T}_r(\theta) \vec{\Phi} \vec{\Sigma} \vec{\Phi} \vec{s}^{\,0}_r \right) \\ & + \left(\vec{P}^{\,T}_r(\theta) \vec{V}^0_r \vec{P}_r(\theta) - 2 \vec{P}^{\,T}_r(\theta) \vec{V}^s_r \right) \Bigg\}, \end{aligned} \tag{23}$$

and

$$\vec{s}_r^{\,0} = \frac{1}{N_r}\sum_{j=1}^{N_r}\vec{s}_{rj}, \qquad \vec{\mu}_r^{\,0} = \frac{1}{N_r}\sum_{j=1}^{N_r}\vec{\mu}_{rj}, \qquad \vec{V}_r^{\,0} = \frac{1}{N_r}\sum_{j=1}^{N_r}\vec{V}_{rj}, \qquad \vec{V}_r^{\,s} = \frac{1}{N_r}\sum_{j=1}^{N_r}\vec{V}_{rj}\vec{s}_{rj}, \tag{24a}$$

$$\vec{V}_{rj} = \, Var\left(\vec{Q}_{rjt}(\vec{p}_r, f_r)\Big|\, \epsilon_r, \vec{\omega}_{rj}\right) = \mathbb{E}\left[\vec{\Phi}\vec{\xi}_{rjt}\vec{\xi}_{rjt}^{\,T}\vec{\Phi}\right]. \tag{24b}$$

Proof. In optimization (9), constraints of different contracting groups are independent of each other. Therefore, proving Lemma 4 reduces to proving that the objective function of (9), namely, the expectation $\mathbb{E}_\Theta[\cdot]$ over the entire portfolio, can be written as the sum of expectations over each contracting group.

The remainder of the proof establishes this result by first showing that the mean-variance term within the expectation $\mathbb{E}_\Theta[\cdot]$ can be written as the sum of R mean-variance terms, each of which is associated with one contracting group.

Start by simplifying the objective function (9). The summation in the variance and expectation terms satisfies

$$\begin{aligned}
&\sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\Big[Y_{rjt}(\vec{Q}_{rjt}(\vec{p}_r, f_r), \vec{p}_r, f_r)\Big] \\
&= \sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\left\{\Big[\vec{P}_r(\theta) - \vec{s}_{rj}\Big]^T \vec{Q}_{rjt}\big(\vec{P}_r(\theta), F_r(\theta)\big) + F_r(\theta) - \delta_{rjt}\right\} \\
&= \sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\Big[\vec{P}_r(\theta) - \vec{s}_{rj}\Big]^T \Big[\vec{\Phi}\left(\vec{\Psi}_r\theta + \vec{\mu}_{rj} - \vec{P}_r(\theta)\right) + \vec{\Phi}\left(\epsilon_r\vec{1}_2 + \vec{\omega}_{rj} + \vec{\xi}_{rjt}\right)\Big] \qquad (25a) \\
&\qquad + \sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\left[F_r(\theta) - \delta_{rjt}\right], \qquad (25b)
\end{aligned}$$

where (25a) follows from (5).

Next consider the variance of (25). Because $\vec{P}_r(\theta) - \vec{s}_{rj}$ and $\vec{\Phi}\big(\vec{\Psi}_r\theta + \vec{\mu}_{rj} - \vec{P}_r(\theta)\big)$ in (25a), and

$F_r(\theta)$ in (25b) are constants, variance of (25) is equivalent to

$$\text{Var}\left(\sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]^T\vec{\Phi}\left(\epsilon_r\vec{1}_2+\vec{\omega}_{rj}+\vec{\xi}_{rjt}\right)-\sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\delta_{rjt}\right)$$

$$= \text{Var}\left(\sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]^T\vec{\Phi}\vec{1}_2\epsilon_r\right)+ \tag{26a}$$

$$\text{Var}\left(\sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]^T\vec{\Phi}\vec{\omega}_{rj}\right)+ \tag{26b}$$

$$\text{Var}\left(\sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]^T\vec{\Phi}\vec{\xi}_{rjt}\right)+ \tag{26c}$$

$$\text{Var}\left(\sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\delta_{rjt}\right), \tag{26d}$$

where the decomposition follows from the fact that random variables ϵ, ω, ξ, and δ are orthogonal to each other.

Recall that the contracting group-level random effects $\epsilon_r \overset{iid}{\sim} N(0,\sigma^2)$. Using $\vec{s}^{\,0}_r = \frac{1}{N_r}\sum_{j=1}^{N_r}\vec{s}_{rj}$, the variance term in (26a) satisfies

$$\text{Var}\left(\sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]^T\vec{\Phi}\vec{1}_2\epsilon_r\right) = \text{Var}\left(\sum_{r=1}^{R}\sum_{j=1}^{N_r}\tau_r\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]^T\vec{\Phi}\vec{1}_2\epsilon_r\right)$$
$$= \text{Var}\left(\sum_{r=1}^{R}\tau_r\left[N_r\vec{P}_r(\theta)-\sum_{j=1}^{N_r}\vec{s}_{rj}\right]^T\vec{\Phi}\vec{1}_2\epsilon_r\right) = \sum_{r=1}^{R}\left(N_r\tau_r\left[\vec{P}_r(\theta)-\vec{s}^{\,0}_r\right]^T\vec{\Phi}\vec{1}_2\right)^2\sigma^2. \tag{27}$$

Recall that $\vec{\omega}_{rj} \overset{iid}{\sim} N\big(0,\vec{\Sigma}\big)$. The variance term in (26b) is

$$\text{Var}\left(\sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]^T\vec{\Phi}\vec{\omega}_{rj}\right) = \text{Var}\left(\sum_{r=1}^{R}\sum_{j=1}^{N_r}\tau_r\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]^T\vec{\Phi}\vec{\omega}_{rj}\right)$$
$$= \sum_{r=1}^{R}\sum_{j=1}^{N_r}\tau_r^2\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]^T\vec{\Phi}\vec{\Sigma}\vec{\Phi}\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]. \tag{28}$$

Recall that $\vec{\xi}_{rjt}$ have zero mean and are independent across printers. Using $\vec{V}_{rj}$ in (24b), the variance term in (26c) is

$$\mathrm{Var}\left(\sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]^T\vec{\Phi}\vec{\xi}_{rj}(t)\right)=\sum_{r=1}^{R}\sum_{j=1}^{N_r}\tau_r\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]^T\vec{V}_{rj}\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]. \tag{29}$$

Finally, because (26d) is independent of the contract terms, it can be dropped from the objective function. Therefore, combining (27)–(29) and dropping terms that are independent of the contract, the variance term in the objective function can be simplified as follows

$$\sum_{r=1}^{R}\left[N_r\tau_r\big(\vec{P}_r(\theta)-\vec{s}_r^{\,0}\big)^T\vec{\Phi}\vec{1}_2\right]^2\sigma^2+$$
$$\sum_{r=1}^{R}N_r\tau_r^2\left[\vec{P}_r^{\,T}(\theta)\vec{\Phi}\vec{\Sigma}\vec{\Phi}\vec{P}_r(\theta)-2\vec{P}_r^{\,T}(\theta)\vec{\Phi}\vec{\Sigma}\vec{\Phi}\vec{s}_r^{\,0}\right]+ \tag{30a}$$
$$\sum_{r=1}^{R}N_r\tau_r\left[\vec{P}_r^{\,T}(\theta)\vec{V}_r^0\vec{P}_r(\theta)-2\vec{P}_r^{\,T}(\theta)\vec{V}_r^s\right]. \tag{30b}$$

where $\vec{V}_r^0=\frac{1}{N_r}\sum_{j=1}^{N_r}\vec{V}_{rj}$, $\vec{V}_r^s=\frac{1}{N_r}\sum_{j=1}^{N_r}\vec{V}_{rj}\vec{s}_{rj}$, (30a) corresponds to (28) and (30b) to (29).

Now consider the expectation term in the objective function (9). Using $\mathbb{E}[\epsilon_r]=0$, $\mathbb{E}[\vec{\omega}_{rj}]=0$, and $\mathbb{E}[\vec{\xi}_{rj}(t)]=0$, equation (25) yields

$$\mathbb{E}\left[\sum_{r=1}^{R}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\left[\vec{P}_r(\theta)-\vec{s}_{rj}\right]^T\vec{\Phi}\left(\vec{\Psi}_r\theta-\vec{P}_r(\theta)+\vec{\mu}_{rj}\right)+F_r(\theta)\right]$$
$$=\sum_{r=1}^{R}N_r\tau_r\left[\big(\vec{P}_r(\theta)-\vec{s}_r^{\,0}\big)^T\vec{\Phi}\big(\vec{\Psi}_r\theta-\vec{P}_r(\theta)+\vec{\mu}_r^0\big)+F_r(\theta)\right]+\sum_{r=1}^{R}\tau_r\left[N_r(\vec{s}_r^{\,0})^T\vec{\Phi}\vec{\mu}_r^{\,0}-\sum_{j=1}^{N_r}\vec{s}_{rj}^{\,T}\vec{\Phi}\vec{\mu}_{rj}\right]. \tag{31}$$

The last term in (31) is independent of the contract and can be dropped.

Therefore, conditional on the value of θ, the mean-variance objective on the overall cumulative cash flow can be written as a summation of R mean-variance terms $\mathrm{MV}_r(F_r(\theta),\vec{P}_r(\theta))$ $(r=1,2,\ldots,R)$, where $\mathrm{MV}_r(F_r(\theta),\vec{P}_r(\theta))$ is as defined in equation (23).

Because $\boldsymbol{\theta}_r\perp\epsilon_r\perp\vec{\omega}_{rj}$ and $\boldsymbol{\theta}_r$ is independent across contracting groups, the expectation $\mathbb{E}_{\boldsymbol{\Theta}}[\cdot]$ can be written as the summation of $\mathbb{E}_{\boldsymbol{\theta}_r}[MV_r(F_r(\boldsymbol{\theta}_r),\vec{P}_r(\boldsymbol{\theta}_r))]$, establishing the lemma. □

Lemma 5. *The provider's optimization* (22) *is equivalent to the following problem.*

$$\max_{\{(F_r(\theta),\vec{P}_r(\theta)):\theta\in[0,1]\}} \qquad \int_0^1 MV_r\big(F_r(\theta),\vec{P}_r(\theta)\big)h(\theta)\,d\theta, \tag{32a}$$

$$s.t. \qquad \frac{du_r^{**}(\theta)}{d\theta} = \vec{\Psi}_r^T\vec{\Phi}\big[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}\,_r^0\big] \tag{32b}$$

$$\vec{\Psi}_r^T\vec{\Phi}\frac{d\vec{P}_r(\theta)}{d\theta} \le 0, \tag{32c}$$

$$u_r^{**}(0) \ge u_{0r}, \tag{32d}$$

where $MV_r(F_r(\theta),\vec{P}_r(\theta))$ *is defined in* (23)*, and*

$$u_r^{**}(\theta) = \frac{1}{2}\left[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}\,_r^0\right]^T\vec{\Phi}\left[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}\,_r^0\right] - F_r(\theta), \tag{33a}$$

$$u_{0r} = \frac{U_{0r}}{N_r\tau_r} - \frac{1}{2}\left(\sigma^2\vec{1}_2^T\vec{\Phi}\vec{1}_2 + \frac{1}{N_r}\sum_{j=1}^{N_r}\mathbb{E}[\vec{\omega}_{rj}^T\vec{\Phi}\vec{\omega}_{rj}] - \frac{1}{N_r\tau_r}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\mathbb{E}\left[\vec{\xi}_{rjt}^T\vec{\Phi}\vec{\xi}_{rjt}\right]\right). \tag{33b}$$

Proof. First consider the IC constraint (22b). From equation (7),

$$U_r\left(\vec{P}_r(\theta),F_r(\theta),\theta\right) = \sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}U_{rjt}\left(\vec{P}_r(\theta),F_r(\theta),\theta\right)$$
$$=\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\frac{1}{2}\big(\vec{\Psi}_r\theta + \vec{\mu}_{rj} - \vec{P}_r(\theta) + \epsilon_r\vec{1}_2 + \vec{\omega}_{rj} + \vec{\xi}_{rjt}\big)^T\vec{\Phi}$$
$$\times\left(\vec{\Psi}_r\theta + \vec{\mu}_{rj} - \vec{P}_r(\theta) + \epsilon_r\vec{1}_2 + \vec{\omega}_{rj} + \vec{\xi}_{rjt}\right) - F_r(\theta).$$

Given that $\mathbb{E}[\epsilon_r]=0$, $\mathbb{E}[\vec{\omega}_{rj}]=0$, $\mathbb{E}[\vec{\xi}_{rj}(t)]=0$, and $\epsilon_r \perp \vec{\omega}_{rj} \perp \vec{\xi}_{rj}(t)$, and using $\vec{\mu}_r^0 = \frac{1}{N_r}\sum_{j=1}^{N_r}\vec{\mu}_{rj}$, $\mathbb{E}_{(\vec{\xi},\vec{\omega},\epsilon)}\left[U_r\big(\vec{P}(\theta),F(\theta),\theta\big)\right]$ satisfies

$$\mathbb{E}_{(\vec{\xi},\vec{\omega},\epsilon)}\left[U_r\left(\vec{P}_r(\theta),F_r(\theta),\theta\right)\right]$$
$$= N_r\tau_r\left[\frac{1}{2}\big(\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}\,_r^0\big)^T\vec{\Phi}\left(\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}\,_r^0\right) - F_r(\theta)\right] \tag{34a}$$

$$+\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r}\frac{1}{2}\,\mathbb{E}\left[\sigma^2\vec{1}_2^T\vec{\Phi}\vec{1}_2 + \vec{\omega}_{rj}^T\vec{\Phi}\vec{\omega}_{rj} + \vec{\xi}\,_{rjt}^T\vec{\Phi}\vec{\xi}_{rjt}\right]. \tag{34b}$$

Because (34b) is invariant in θ, the IC constraint (22b) reduces to

$$\frac{1}{2}\left[\vec{\Psi}_r\theta - \vec{P}_r(\theta') + \vec{\mu}_r^{\,0}\right]^T \vec{\Phi}\left[\vec{\Psi}_r\theta - \vec{P}_r(\theta') + \vec{\mu}_r^{\,0}\right] - F_r(\theta')$$
$$\leq \frac{1}{2}\left[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\right]^T \vec{\Phi}\left[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\right] - F_r(\theta) \tag{35}$$

That is, the contract designed for contracting group r of type-θ then solves the following problem:

$$\theta = \arg\max_{\theta'} \; \frac{1}{2}\left[\vec{\Psi}_r\theta - \vec{P}_r(\theta') + \vec{\mu}_r^{\,0}\right]^T \vec{\Phi}\left[\vec{\Psi}_r\theta - \vec{P}_r(\theta') + \vec{\mu}_r^{\,0}\right] - F_r(\theta'). \tag{36}$$

The first- and second-order local conditions for the contract $(f_r(\theta), \vec{p}_r(\theta))$ to be incentive compatible are, respectively:

$$\left(\frac{\mathrm{d}\vec{P}_r(\theta)}{\mathrm{d}\theta}\right)^T \vec{\Phi}\left[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\right] + \frac{\mathrm{d}F_r(\theta)}{\mathrm{d}\theta} = 0, \tag{37a}$$

$$\vec{\Psi}_r^T\vec{\Phi}\frac{\mathrm{d}\vec{P}_r(\theta)}{\mathrm{d}\theta} \leq 0. \tag{37b}$$

Because the Spence-Mirrlees condition is satisfied by the utility associated with each contracting group, the local first and second order conditions are necessary and sufficient for incentive compatibility (Laffont and Martimort 2001).

Let

$$u_r^{**}(\theta) = \frac{1}{2}\left[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\right]^T \vec{\Phi}\left[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\right] - F_r(\theta). \tag{38}$$

Then the first-order IC constraint (37a) is equivalent to the following constraint on $u^{**}(\theta)$:

$$\frac{\mathrm{d}u_r^{**}(\theta)}{\mathrm{d}\theta} = \vec{\boldsymbol{\Psi}}^T\vec{\Phi}\big[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\big]. \tag{39}$$

Finally consider the participation constraint (22c). Combining (34) and (38) yields

$$u_r^{**}(\theta) \geq u_{0r} \qquad \forall\theta \in [0, 1], \tag{40}$$

where

$$u_{0r} = \frac{U_{0r}}{N_r \tau_r} - \frac{1}{2}\left(\sigma^2 \vec{1}_2^T \vec{\Phi} \vec{1}_2 + \frac{1}{N_r}\sum_{j=1}^{N_r} \mathbb{E}[\vec{\omega}_{rj}^T \vec{\Phi} \vec{\omega}_{rj}] - \frac{1}{N_r \tau_r}\sum_{j=1}^{N_r}\sum_{t=1}^{\tau_r} \mathbb{E}\left[\vec{\xi}_{rjt}^T \vec{\Phi} \vec{\xi}_{rjt}\right]\right). \tag{41}$$

Furthermore, because $\vec{\Phi}\big[\vec{\Psi}_r\theta - \vec{p}_r + \vec{\mu}_r^{\,0}\big]$ is the expected monthly volume (equation (5)), which is positive by definition, and $\vec{\Psi}_r > 0$, $\mathrm{d}u_r^{**}(\theta)/\mathrm{d}\theta > 0$ in (39). Therefore, (40) reduces to a constraint on the lowest type $\theta = 0$:

$$u_r^{**}(0) \geq u_{0r}, \tag{42}$$

completing the proof. □

Proof of Theorem 1. From Lemma 5, optimization (22) for contracting group r is equivalent to optimization (32), which we shall solve below. We first ignore constraint (32c).

Let the participation constraint of the lowest type be binding

$$u_r^{**}(0) = u_{0r}. \tag{43}$$

Then by equation (32b), we have

$$u_r^{**}(\theta) = u_{0r} + \vec{\Psi}_r^T \vec{\Phi} \int_0^\theta \big[z\vec{\Psi}_r - \vec{P}_r(z) + \vec{\mu}_r^{\,0}\big]\, dz. \tag{44}$$

Using the definition of $u_r^{**}(\theta)$, equation (44) yields

$$F_r(\theta) = \frac{1}{2}\big[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\big]^T \vec{\Phi}\big[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\big] - u_{0r} - \vec{\Psi}_r^T \vec{\Phi} \int_0^\theta \big[z\vec{\Psi}_r - \vec{P}_r(z) + \vec{\mu}_r^{\,0}\big]\, dz. \tag{45}$$

Then in the objective function of (32), the integral over $f_r(\theta)$ becomes

$$\begin{aligned}\int_0^1 F_r(\theta)h(\theta)\, d\theta = &\int_0^1 \frac{1}{2}\big[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\big]^T \vec{\Phi}\big[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\big]h(\theta)d\theta \\ &- \vec{\Psi}_r^T \vec{\Phi} \int_0^1 \left[\int_0^\theta \big[z\vec{\Psi}_r - \vec{p}_r(z) + \vec{\mu}_r^{\,0}\big]\, dz\right] h(\theta)\, d\theta - u_{0r}.\end{aligned} \tag{46}$$

Using integration by parts, the double integral on the right side of equation (46) becomes

$$
\begin{aligned}
&\int_0^1 \int_0^\theta \left[z\vec{\Psi}_r - \vec{P}_r(z) + \vec{\mu}_r^{\,0}\right] h(\theta)\, dz\, d\theta \\
&= H(\theta) \int_0^\theta \left[z\vec{\Psi}_r - \vec{P}_r(z) + \vec{\mu}_r^{\,0}\right] dz \Big|_0^1 - \int_0^1 H(\theta) \left[\theta\vec{\Psi}_r - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\right] d\theta \\
&= \int_0^1 \bar{H}(\theta) \left[\theta\vec{\Psi}_r - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\right] d\theta,
\end{aligned}
$$

where $H(\theta)$ is the cumulative distribution function of θ, $\bar{H}(\theta) = 1 - H(\theta)$. Thus, the integral over $f_r(\theta)$ becomes

$$
\int_0^1 \left(\frac{1}{2}\left[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\right] - \frac{\bar{H}(\theta)}{h(\theta)}\vec{\Psi}_r\right)^T \vec{\Phi}\left[\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\right] h(\theta) d\theta - u_{0r}. \tag{47}
$$

Plugging (47) into the first term of $\mathrm{MV}_r\big(F_r(\theta), \vec{P}_r(\theta)\big)$ in equation (23) yields

$$
\begin{aligned}
&\int_0^1 N_r\tau_r \left[\left(\vec{P}_r(\theta) - \vec{s}_r^{\,0}\right)^T \vec{\Phi}\left(\vec{\Psi}_r\theta - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\right) + F_r(\theta)\right] h(\theta) d\theta \\
&= N_r\tau_r \int_0^1 \left(\frac{\vec{P}_r(\theta) + \theta\vec{\Psi}_r + \vec{\mu}_r^{\,0}}{2} - \vec{s}_r^{\,0} - \frac{\bar{H}(\theta)}{h(\theta)}\vec{\Psi}_r\right)^T \vec{\Phi}\left[\theta\vec{\Psi}_r - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\right] h(\theta)\, d\theta - N_r\tau_r u_{0r}. \qquad (48)
\end{aligned}
$$

Combining (48) with the rest three terms of $\mathrm{MV}_r\big(F_r(\theta), \vec{P}_r(\theta)\big)$ in (23), the integral in (32a) is

$$
\begin{aligned}
&\int_0^1 MV_r\big(F_r(\theta), \vec{P}_r(\theta)\big) h(\theta) d\theta \\
&= N_r\tau_r \int_0^1 \left(\frac{\vec{P}_r(\theta) + \theta\vec{\Psi}_r + \vec{\mu}_r^{\,0}}{2} - \vec{s}_r^{\,0} - \frac{\bar{H}(\theta)}{h(\theta)}\vec{\Psi}_r\right)^T \vec{\Phi}\left[\theta\vec{\Psi}_r - \vec{P}_r(\theta) + \vec{\mu}_r^{\,0}\right] h(\theta)\, d\theta \qquad (49a) \\
&\quad - N_r\tau_r u_{0r} - \lambda N_r^2\tau_r^2\sigma^2 \int_0^1 \left(\left[\vec{P}_r(\theta) - \vec{s}_r^{\,0}\right]^T \vec{\Phi}\vec{1}_2\right)^2 h(\theta)\, d\theta \qquad (49b) \\
&\quad - \lambda N_r\tau_r^2 \int_0^1 \left(\vec{P}_r^{\,T}(\theta)\vec{\Phi}\vec{\Sigma}\vec{\Phi}\vec{P}_r(\theta) - 2\vec{P}_r^{\,T}(\theta)\vec{\Phi}\vec{\Sigma}\vec{\Phi}\vec{s}_r^{\,0}\right) h(\theta)\, d\theta \qquad (49c) \\
&\quad - \lambda N_r\tau_r \int_0^1 \left(\vec{P}_r^{\,T}(\theta)\vec{V}_r^0\vec{P}_r(\theta) - 2\vec{P}_r^{\,T}(\theta)\vec{V}_r^s\right) h(\theta)\, d\theta. \qquad (49d)
\end{aligned}
$$

Therefore, finding the function $\vec{P}_r(\theta)$ that optimizes equation (49) is equivalent to the following pointwise maximization

$$\max_{\vec{p}_r} \quad \left(\frac{\vec{p}_r + \theta\vec{\Psi}_r + \vec{\mu}_r^{\,0}}{2} - \vec{s}_r^{\,0} - \frac{\bar{H}(\theta)}{h(\theta)}\vec{\Psi}_r\right)^T \vec{\Phi}\left[\theta\vec{\Psi}_r - \vec{p}_r + \vec{\mu}_r^{\,0}\right] \tag{50a}$$

$$- \lambda N_r \tau_r \sigma^2 \left(\left[\vec{p}_r - \vec{s}_r^{\,0}\right]^T \vec{\Phi}\vec{1}_2\right)^2 - \lambda\tau_r \left(\vec{p}_r^{\,T}\vec{\Phi}\vec{\Sigma}\vec{\Phi}\vec{p}_r - 2\vec{p}_r^{\,T}\vec{\Phi}\vec{\Sigma}\vec{\Phi}\vec{s}_r^{\,0}\right) \tag{50b}$$

$$- \lambda\left(\vec{p}_r^{\,T}\vec{V}_r^0\vec{p}_r - 2\vec{p}_r^{\,T}\vec{V}_r^s\right). \tag{50c}$$

Maximization (50) is a concave function of $\vec{p}_r$. First-order condition yields

$$\vec{P}_r^{\,*}(\theta) = \vec{s}_r^{\,0} + \left(2\lambda\vec{\Gamma}_r + \vec{\Phi}\right)^{-1}\left[\vec{\Phi}\frac{\bar{H}(\theta)}{h(\theta)}\vec{\Psi}_r + 2\lambda(\vec{V}_r^s - \vec{V}_r^0\vec{s}_r^{\,0})\right], \tag{51}$$

where $\vec{\Gamma}_r$ is given in Theorem 1.

When θ follows beta distribution $Beta(1, 1/\kappa)$, $\bar{H}(\theta)/h(\theta) = (1-\theta)\kappa$. It is easy to check that $\vec{P}_r^{\,*}(\theta)$ satisfies the earlier ignored constraint (32c). Therefore, $\vec{P}_r^{\,*}(\theta)$ in equation (51) is the equilibrium variable price.

The equilibrium fixed price $F_r^*(\theta)$ follows by combining equations (51) and (45). □

A.2 Proofs of statements in §5

Proof of Lemma 1. From equation (11), the observed variable prices $\vec{p}_r^{\,*}$ satisfies

$$\vec{p}_r^{\,*} = \vec{P}_r^{\,*}(\theta_r) = \vec{s}_r^{\,0} + \left(2\lambda\vec{\Gamma}_r + \vec{\Phi}\right)^{-1}\left[\vec{\Phi}(1-\theta_r)\kappa\vec{\Psi}_r + 2\lambda\left(\vec{V}_r^s - \vec{V}_r^0\vec{s}_r^{\,0}\right)\right]. \tag{52}$$

This gives

$$\vec{\Psi}_r(1-\theta_r) = \frac{1}{\kappa}\vec{\Phi}\left(2\lambda\vec{\Gamma}_r + \vec{\Phi}\right)\left(\vec{p}_r^{\,*} - \vec{s}_r^{\,0}\right) - \frac{2\lambda}{\kappa}\vec{\Phi}\left(\vec{V}_r^s - \vec{V}_r^0\vec{s}_r^{\,0}\right),$$

which yields equation (15). □

Proof of Lemma 2. Substituting equations (15), (16) and (19) into equation (17) yields

$$\vec{q}_{rj}^{\,*} = \vec{\Phi}\vec{\Psi}_r - \frac{1}{\kappa}\left(2\lambda\vec{\Gamma}_r + \vec{\Phi}\right)\left(\vec{p}_r^{\,*} - \vec{s}_r^{\,0}\right) + \frac{2\lambda}{\kappa}\left(\vec{V}_r^{\,s} - \vec{V}_r^{\,0}\vec{s}_r^{\,0}\right) - \vec{\Phi}\vec{p}_r^{\,*} + \vec{\Phi}\vec{1}_2\vec{\gamma}^T\vec{Z}_{rj} + \vec{\Phi}\vec{1}_2\epsilon_r + \vec{\Phi}\vec{\omega}_{rj}. \tag{53}$$

In equation (53), term $\vec{\Gamma}_r$ is defined in equation (13b), which is repeated below for convenience:

$$\vec{\Gamma}_r = \vec{V}_r^{\,0} + N_r \tau_r \sigma^2 \vec{\Phi} \vec{1}_2 \vec{1}_2^T \vec{\Phi} + \tau_r \vec{\Phi} \vec{\Sigma} \vec{\Phi}.$$

Substituting equation (3b) for $\vec{\Phi}$ and recalling that $\vec{\Sigma}$ is the diagonal covariance matrix of shock $\vec{\omega}_{rj}$ with diagonals σ_b^2 and σ_c^2, we can write $\vec{\Gamma}_r$ as follows

$$\vec{\Gamma}_r = \vec{V}_r^{\,0} + N_r \tau_r \begin{pmatrix} g_1 & g_2 \\ g_2 & g_3 \end{pmatrix} + \tau_r \begin{pmatrix} h_1 & h_2 \\ h_2 & h_3 \end{pmatrix} \tag{54}$$

where

$$\vec{V}_r^{\,0} = \begin{pmatrix} V_{rb}^0 & V_{rbc}^0 \\ V_{rbc}^0 & V_{rc}^0 \end{pmatrix} \tag{55}$$

is the observed average covariance matrix of the monthly print volume from all devices in contracting group r (equation (13a)), and g_i and h_i $(i = 1, 2, 3)$ are nonlinear functions of model parameters:

$$g_1 = \sigma^2(\phi_b + \phi_{bc})^2, \tag{56a}$$

$$g_2 = \sigma^2(\phi_b + \phi_{bc})(\phi_c + \phi_{bc}), \tag{56b}$$

$$g_3 = \sigma^2(\phi_c + \phi_{bc})^2, \tag{56c}$$

$$h_1 = \phi_b^2 \sigma_b^2 + \phi_{bc}^2 \sigma_c^2, \tag{56d}$$

$$h_2 = \phi_{bc}(\phi_b \sigma_b^2 + \phi_c \sigma_c^2), \tag{56e}$$

$$h_3 = \phi_{bc}^2 \sigma_b^2 + \phi_c^2 \sigma_c^2. \tag{56f}$$

Substituting equation (54) into equation (53) and writing all vectors in terms of their BW and color components, we obtain the following two nonlinear random-effects models on the observed BW and color monthly print volume, q_{rjb}^* and q_{rjc}^*. The model on the average BW volume q_{rjb}^* is

$$q_{rjb}^* = G^1(\vec{K}_{rj}; \vec{\Lambda}) + u_r^1 + e_{rj}^1 \tag{57}$$

where errors u_r^1 and e_{rj}^1 are as defined in the lemma and $G^1(\vec{K}_{rj};\vec{\Lambda})$ is

$$G^1(\vec{K}_{rj};\vec{\Lambda}) = (\phi_b\psi_b + \phi_{bc}\psi_c) - \left(\frac{2\lambda}{\kappa}\right) K_r^1 - \left(\frac{2\lambda}{\kappa}g_1\right) K_r^2 - \left(\frac{2\lambda}{\kappa}h_1\right) K_r^3 - \left(\frac{2\lambda}{\kappa}g_2\right) K_r^4$$
$$- \left(\frac{2\lambda}{\kappa}h_2\right) K_r^5 + (\phi_b + \phi_{bc})\vec{\beta}^T \vec{K}_r^6 - \left(\frac{\phi_b}{\kappa}\right) K_r^7 - \left(\frac{\phi_{bc}}{\kappa}\right) K_r^8$$
$$+ \left(\frac{2\lambda}{\kappa}\right) K_r^9 - (\phi_b)\, K_r^{10} - (\phi_{bc})\, K_r^{11} + (\phi_b + \phi_{bc}) \left(\vec{\gamma}^T\right) \vec{K}_{rj}^{12}, \tag{58}$$

The model on the average color volume q_{rjc}^* is

$$q_{rjc}^* = G^2(\vec{K}_{rj};\vec{\Lambda}) + u_r^2 + e_{rj}^2, \tag{59}$$

where errors u_r^2 and e_{rj}^2 are defined in the lemma and $G^2(\vec{Y};\vec{\Lambda})$ is

$$G^2(\vec{K}_{rj};\vec{\Lambda}) = (\phi_{bc}\psi_b + \phi_c\psi_c) - \left(\frac{2\lambda}{\kappa}\right) K_r^{13} - \left(\frac{2\lambda}{\kappa}g_2\right) K_r^2 - \left(\frac{2\lambda}{\kappa}h_2\right) K_r^3 - \left(\frac{2\lambda}{\kappa}g_3\right) K_r^4$$
$$- \left(\frac{2\lambda}{\kappa}h_3\right) K_r^5 + (\phi_c + \phi_{bc})\vec{\beta}^T \vec{K}_r^6 - \left(\frac{\phi_{bc}}{\kappa}\right) K_r^7 - \left(\frac{\phi_c}{\kappa}\right) K_r^8$$
$$+ \left(\frac{2\lambda}{\kappa}\right) K_r^{14} - (\phi_{bc})\, K_r^{10} - (\phi_c)\, K_r^{11} + (\phi_b + \phi_{bc}) \left(\vec{\gamma}^T\right) \vec{K}_{rj}^{12}. \tag{60}$$

In equations (58) and (60), Y^i $(i = 1, \ldots, 14)$ are explanatory variables defined as follows

$$K_r^1 = V_{rb}^0\left(p_{rb}^* - s_{rb}^0\right) + V_{rbc}^0\left(p_{rc}^* - s_{rc}^0\right), \qquad K_r^2 = N_r\tau_r\left(p_{rb}^* - s_{rb}^0\right), \qquad K_r^3 = \tau_r\left(p_{rb}^* - s_{rb}^0\right), \tag{61a}$$

$$K_r^4 = N_r\tau_r\left(p_{rc}^* - s_{rc}^0\right), \qquad K_r^5 = \tau_r\left(p_{rc}^* - s_{rc}^0\right), \qquad \vec{K}_r^6 = \vec{W}_r, \quad K_r^7 = p_{rb}^* - s_{rb}^0, \tag{61b}$$

$$K_r^8 = p_{rc}^* - s_{rc}^0, \qquad K_r^9 = V_{rb}^s - V_{rb}^0 s_{rb}^0 - V_{rbc}^0 s_{rc}^0, \qquad K_r^{10} = p_{rb}^*, \qquad K_r^{11} = p_{rc}^*, \tag{61c}$$

$$\vec{K}_{rj}^{12} = \vec{Z}_{rj}, \qquad K_r^{13} = V^{\,0}_{\,rbc}\left(p_{rb}^* - s_{rb}^0\right) + V^{\,0}_{\,rc}\left(p_{rc}^* - s_{rc}^0\right), \qquad K_r^{14} = V_{rc}^s - V_{rbc}^0 s_{rb}^0 - V_{rc}^0 s_{rc}^0. \tag{61d}$$

By definition, $VP_{rj}^* = p_{rb}^* q_{rjb}^* + p_{rc}^* q_{rjc}^*$. Thus, using models (57) and (59), the econometric model with the variable payment as the response variable is:

$$VP_{rj}^* = G^3(\vec{K}_{rj};\vec{\Lambda}) + u_r^3 + e_{rj}^3, \tag{62}$$

where u_r^3 and e_{rj}^3 are defined in the lemma and $G^3(\vec{K}_{rj};\vec{\Lambda})$ is the following nonlinear function of model parameters

$$G^3(\vec{K}_{rj};\vec{\Lambda}) = p_{rb}^* G^1(\vec{K}_{rj};\vec{\Lambda}) + p_{rc}^* G^2(\vec{K}_{rj};\vec{\Lambda}). \tag{63}$$

□

A.3 Proof of Lemma 3

Proof of Lemma 3. When the provider knows the type of each contracting group, it maximizes the mean-variance objective $MV_r\big(F(\theta),\vec{P}(\theta)\big)$ defined in equation (23), which is repeated below for the reader's convenience:

$$\begin{aligned}\mathrm{MV}_r(F_r(\theta),\vec{P}_r(\theta)) &= N_r\tau_r\big[\big(\vec{P}_r(\theta)-\vec{s}_r^{\,0}\big)^T\vec{\Phi}\big(\vec{\Psi}_r\theta-\vec{P}_r(\theta)+\vec{\mu}_r^{\,0}\big)+F_r(\theta)\big]\\ &-\lambda N_r\tau_r\Bigg\{N_r\tau_r\left(\left[\vec{P}_r(\theta)-\vec{s}_r^{\,0}\right]^T\vec{\Phi}\vec{1}_2\right)^2\sigma^2+\tau_r\left(\vec{P}_r^{\,T}(\theta)\vec{\Phi}\vec{\Sigma}\vec{\Phi}\vec{P}_r(\theta)-2\vec{P}_r^{\,T}(\theta)\vec{\Phi}\vec{\Sigma}\vec{\Phi}\vec{s}_r^{\,0}\right)\\ &+\left(\vec{P}_r^{\,T}(\theta)\vec{V}_r^0\vec{P}_r(\theta)-2\vec{P}_r^{\,T}(\theta)\vec{V}_r^s\right)\Bigg\}.\end{aligned}$$

Because the company has no private information, the provider maximizes $MV_r\big(F(\theta),\vec{P}(\theta)\big)$ of contracting group r ($r=1,2,\ldots,R$) subject to the participation constraint: $\mathbb{E}_{(\vec{\xi},\vec{\omega},\epsilon)}\left[U_r(\vec{P}_r(\theta),F_r(\theta),\theta)\right]\geq U_{0r}$, where $\mathbb{E}_{(\vec{\xi},\vec{\omega},\epsilon)}$ is the expected utility over the contracting horizon, and U_{0r} is the reservation utility of contracting group r as defined in Lemma 4.

From equation (34) and using $u_r^{**}(\theta)$ and u_{0r} defined in equations (38) and (41), the (IR) constraint can be rewritten as $u_r^{**}(\theta)\geq u_{0r}$. Thus, the provider's contract design problem under symmetric information is to maximize $MV_r\big(F(\theta),\vec{P}(\theta)\big)$ under the (IR) constraint $u_r^{**}(\theta)\geq u_{0r}$. At the optimal solution, the (IR) constraint is binding and thus

$$F_r(\theta)=\frac{1}{2}\big[\vec{\Psi}_r\theta+\vec{\mu}_r^{\,0}-\vec{P}_r(\theta)\big]^T\vec{\Phi}\big[\vec{\Psi}_r\theta+\vec{\mu}_r^{\,0}-\vec{P}_r(\theta)\big]-u_{0r}. \tag{64}$$

Substituting equation (64) into the mean-variance objective yields a concave function in the variable price $\vec{P}_r(\theta)$. Thus, the first-order condition yields that the optimal variable prices satisfy equation (20). From equations (64) and (12), the optimal fixed price satisfies (21), completing the proof. □

B Computational details of estimating the econometric model parameters

We used Matlab for the maximum likelihood estimation (MLE) of the BW model, the Color model, and the Payment model in Lemma 2. For each model, the built-in unconstrained optimization functions *fminsearch* and *fminunc* with quasi-Newton algorithm is used (we apply logistic transformation to $\kappa \in [0, 1]$ and logarithmic transformations to λ, ϕ_b, ϕ_c, ψ_b, ψ_c, σ^2, σ_b^2 and σ_c^2 so that all parameters being estimated are defined on $(-\infty, +\infty)$). The tolerance levels are 10^{-8} for both the objective function value and the variables. The maximum iteration is set to be 5000 and the maximum number of function evaluations is set to be 80000.

In solving the optimization problem, we randomly generate 200 initial points for the solver. Thus, the solver solves the same optimization problem 200 times from 200 initial points using two algorithms *fminsearch* and *fminunc*. We then use the output that generates the highest log-likelihood value as the point estimates. This way we can reduce the possibility of being trapped in a local optimum. For a given set of initial points, solving the optimization takes about 20 seconds using *fminunc* and about 3 mins using *fminsearch* on a High Performance Computing node with 12 Intel processing cores and 48GB of memory.

Before working with real data, we carry out simulation and confirm that the best output can recover the true parameter values used to generate the data. Furthermore, when varying the random seeds and ranges for generating the initial points, the optimization procedure provides very similar outputs, thus providing computational evidence for the numerical stability of our solution.

The confidence intervals are computed using parametric bootstrapping. Take the BW model as an example. Let $\hat{\vec{\Lambda}}$ denote the estimated value of the parameters from the BW model. With $\hat{\vec{\Lambda}}$, we can compute $G^1(\vec{K}_{rj}; \hat{\vec{\Lambda}})$ and fully charaterize the distributions of u_r^1 and e_{rj}^1 in Lemma 2. We obtain 44 samples from the estimated distribution of u_r^1 as the group-level errors, and obtain 1021 samples from the estimated distribution of e_{rj}^2 as the printer-level errors. Adding these errors to the mean $G^1(\vec{K}_{rj}; \hat{\vec{\Lambda}})$, we obtain a bootstrap data set. We carry out the same procedure as explained above on this bootstrap data set to get a set of parameter estimates. We generate a total of 300 bootstrap data sets, and obtain 300 estimates of the parameters. We use the standard deviation of these 300 estimates and the estimate from the actual data set to construct the confidence intervals.

(For more on parametric bootstrapping, see e.g., Ch. 6.5 in Efron and Tibshirani 1994).

C Alternative evidence of risk-aversion and out-of-sample tests

C.1 Alternative evidence of provider's risk-aversion

Table 2 shows that the RA assumption holds under different model specifications. However, given the complexity of the nonlinear random-effects models in Lemma 2, concerns on the robustness of the result may still arise. In this subsection, we show that our result is robust by testing the following hypothesis using a reduced-form approach.

Hypothesis 1.

(H_0) *The provider is risk-neutral.*

(H_1) *The provider is risk-averse.*

From Theorem 1, the contract prices depend on the covariance matrix of the monthly print volume, $\vec{V}_{rj}$ (equation (10)), only when the provider is risk-averse. Consequently, the company's print volume and monthly payment depend on $\vec{V}_{rj}$ only when the provider is risk-averse. This means that under the null hypothesis (H_0), the observed average BW print volume q^*_{rjb}, average color print volume q^*_{rjc}, and the average variable payment VP^*_{rj}, would not depend on the covariance matrix V_{rj}. Under the alternative hypothesis (H_1), however, q^*_{rjb}, q^*_{rjc}, and VP^*_{rj} would depend on V_{rj}. This observation allows us to test Hypothesis 1 by doing likelihood ratio tests, as explained below.

Use y^i_{rj} as the generic notation for the response variable in the model, where $i \in \{B, C, P\}$ indicates whether y^i_{rj} equals the average BW print volume q^*_{rjb}, the average color print volume q^*_{rjc}, or the average variable payment VP^*_{rj}. Then the null hypothese (H_0) predicts the following linear random-effects model, henceforth referred to as model (RN$_i$):

$$(RN_i) \qquad y^i_{rj} = \alpha^i_0 + \alpha^i_1 f^*_r + \alpha^i_2 p^*_{rb} + \alpha^i_3 p^*_{rc} + \left(\vec{\alpha}^i_4\right)^T \vec{W}_r + \alpha^i_5 s_{rjb} + \alpha^i_6 s_{rjc} + \alpha^i_7 Z_{rj} + o^i_r + z^i_{rj}, \quad (65)$$

where $(\hat{f}_r, \hat{p}_{rb}, \hat{p}_{rc})$ are the contract prices, (s_{rjb}, s_{rjc}) are the variable service costs, $\vec{W}_r$ and Z_{rj} are demographic characteristics, and o^i_r and z^i_{rj} are hierarchical errors.

Hypothesis (H_1) predicts the following model, henceforth referred to as model (RA_i):

$$(RA_i) \qquad y^i_{rj} = \alpha^i_0 + \alpha^i_1 f^*_r + \alpha^i_2 p^*_{rb} + \alpha^i_3 p^*_{rc} + \left(\vec{\alpha}^i_4\right)^T \vec{W}_r + \alpha^i_5 s_{rjb} + \alpha^i_6 s_{rjc} + \alpha^i_7 Z_{rj} + \left(\alpha^i_8 V_{rjb} + \alpha^i_9 V_{rjc} + \alpha^i_{10} V_{rjbc}\right) + o^i_r + z^i_{rj}, \tag{66}$$

where $(V_{rjb}, V_{rjc}, V_{rjbc})$ are the variance and covariance of the monthly print volume (equation (10)).

Equations (RN_i) and (RA_i) are nested random-effects models. Thus, the task of testing Hypothesis 1 reduces to a likelihood ratio test between (RN_i) and (RA_i). If (H_0) holds, then (RA_i) would not provide a better fit to the data than (RN_i), and the coefficients of the variance terms V_{rjb}, V_{rjc} and V_{rjbc} would not be significant. The top panel of Table 3 presents the results when testing (RN_i) and (RA_i) models using the 1,021 printers with cost data. We observe that, for all three choices of response variable, the RN assumption is rejected and the variance terms V_{rjc}, V_{rjc}, and V_{rjbc} are statistically significant.

Finally, to be able to use all 3,065 printers in the test, we remove the variable service costs (s_{rjb}, s_{rjc}) from the list of explanatory variables on the right sides of models (RN_i) and (RA_i). Re-running the estimation and the likelihood ratio tests generates results in the bottom panel of Table 3. We report that these expanded tests still reject the null hypothesis that the provider is risk-neutral, with statistically significant variance terms V_{rjc}, V_{rjc}, and V_{rjbc}.

Table 3. Reduced-form analysis results

L.R.T. stands for "likelihood ratio test."

		t-values of coefficients		
Response	*p*-values of L.R.T.	V_{rjb}	V_{rjc}	V_{rjbc}
1021 printers with cost data				
q^*_{rjb}	< 2.2E-16	15	2	2
q^*_{rjc}	< 2.2E-16	-0.5	18	4
VP^*_{rj}	< 2.2E-16	3	17	0.3
All 3075 printers				
q^*_{rjb}	< 2.2E-16	33	3	5
q^*_{rjc}	< 2.2E-16	0.5	42	4
VP^*_{rj}	< 2.2E-16	6	27	4

C.2 Out-of-sample tests

In this subsection, we perform out-of-sample tests to check whether the contract prices and print volumes predicted by our model agree with the observations. Due to the hierarchical structure of our model, we do the tests on the contracting group level. Specifically, we split the printers within a contracting group into two sub-groups: in-sample (90% of the printers) and out-of-sample (10% of the printers). By our assumption, printers in these two sub-groups have the same private type θ and thus the same group-level WTP $\vec{\zeta}$. Hence, we can use the in-sample printers to find the true average group-level WTP $\vec{\Psi}_r\theta_r$ (Lemma 1), and then use these estimated WTPs to predict the contract prices (Theorem 1) and average print volumes (equation (5)) for the out-of-sample printers. If our contracting model is a good approximation of practice, then our WTP estimate, $\vec{\Psi}_r\theta_r$, should be close to the true baseline WTP of the test printers. Further, the predicted contract prices and volumes should be able to explain the observations from the out-of-sample printers.

Next we describe how we implement this test. There are 44 contracting groups in our data set and, among the 44 contracting groups, 18 groups contain at least ten printers. To have adequate data for the in-sample estimation, we only use these 18 groups for out-of-sample tests. Within each contracting group, we randomly select 90% of the printers as the in-sample printers, and the remaining 10% as the test printers. We pool the observations from all in-sample printers together, estimate the model parameters, and use Lemma 1 to find the average group-level WTPs $\vec{\Psi}_r\theta_r$ of each contracting group. By Lemma 1, we need to know $\vec{\Psi}_r$ to estimate $\vec{\Psi}_r\theta_r$. Thus, we only use the Payment model in Lemma 2 for the in-sample estimation.

We use Theorem 1 with the estimated WTP $\vec{\Psi}_r\theta_r$ to predict the contracts and print volumes for the out-of-sample printers. Comparing the predicted variable prices with the true contract prices, we report that the R-squares are 0.33 and 0.21 for the BW and color variable prices, and is 0.51 for the fixed price. Comparing the predicted average print volumes with the observations, we report that the R-squares are 0.11 and 0.24 for the BW and color volumes, respectively.