ESSAYS ON VALUE CO-CREATION, CO-PRODUCTION, AND THE INTERFACE BETWEEN OPERATIONS AND RECOMMENDER SYSTEMS

A Dissertation

by

EMRE MUZAFFER DEMIREZEN

Submitted to the Office of Graduate Studies of
Texas A&M University
in partial fulfillment of the requirements for the degree of

DOCTOR OF PHILOSOPHY

Chair of Committee,	Subodha Kumar
Co-Chair of Committee,	Bala Shetty
Committee Members,	Sila Cetinkaya
	Neil Geismar
	Rogelio Oliva
	Chelliah Sriskandarajah
Head of the Department,	Rich Metters

August 2013

Major Subject: Information & Operations Management

ABSTRACT

In this dissertation, I study coordination or collaboration settings that are either within company or at inter-organizational levels in the form of three essays. In the first essay, I study the relationship between a client and a vendor in value co-creation environments such as knowledge intensive services. I consider that the client gets utility from the project throughout the development period. The output is contingent on the effort levels of each party and I allow these effort levels to be dynamic. Hence, the client needs to optimally decide the terms of the payment so as to maximize the project output and minimize its cost. In my second essay, I study another value co-creation environment. In this case, unlike the first essay, I assume that the effort levels are not observable but might be monitored. In both essays, I analyze the performance of different contracts and find the best one for the client in diverse settings. Among several other results, I derive the conditions under which the client chooses not to observe vendor's effort level and operates in a double moral hazard environment. In addition, I show that the remaining time of the project and the client's valuation of the project regulate the behavior of the effort levels and some other characteristics in the collaboration.

In the third essay, I consider a subscription based rental organization, such as Netflix and Blockbuster. In these environments, the satisfaction of customers depends on the availability of requested products. Hence, it is important for these firms to satisfy as much demand as possible. Recommender systems, in a DVD-rental context, are typically used to help customers in finding the right movies for them. However, recommendations can be utilized to shift demand among movies considering the inventory level and future demand to increase the number of satis-

fied customers or profitability. I address this issue by considering inventory in the optimization of recommender systems. I present several results that could be utilized by managers in order to make important tactical and operational decisions. Results suggest that the proposed approach may improve profitability of the firms substantially.

DEDICATION

Dedicated to my dear family: Tekin, Nadire, Emine, Tekin, Kerem, Begum, Lutfiye, and Recep...

ACKNOWLEDGEMENTS

I would never have been able to finish my dissertation without the guidance of my committee members, help from friends, and support from my family and my wife.

I would like to express my deepest gratitude to my co-advisor, Dr. Subodha Kumar, for his excellent guidance, caring, patience, and encouragement along the process. Simply put, I have learned a lot from him. I would like to also thank my co-advisor, Dr. Bala Shetty, for his never ending support, guidance, and help at various stages of my PhD that made my life at College Station even more enjoyable.

I wish to thank Dr. Chelliah Sriskandarajah not only because he guided me through several matters during my last year at Texas A&M, but also he served as a member in my committee and was a great co-author to work with. I would like to thank the rest of my committee: Dr. Sila Cetinkaya, Dr. Neil Geismar, and Dr. Rogelio Oliva for their encouragement, and insightful comments on my dissertation.

I want to thank Dr. Michael Ketzenberg, Dr. Rich Metters, Dr. Olga Perdikaki, Dr. Powel Robinson, Dr. Arun Sen, Dr. Gregory Heim, Ms. Donna Schumaker and the rest of the department, for the help I got so freely from them.

To my PhD mates at the department, thanks for being so helpful and supportive. Special thanks go to Howard for being a caring and encouraging friend.

Lastly, I would like to express my deep gratitude to my parents, who have taught me the values of life and education. My mother, father, and sister have always been a source of inspiration and motivation for me. Whenever and wherever I need, I always feel that they are with me and they never let me feel alone. Special thanks to my wife for her never ending help and moral support.

TABLE OF CONTENTS

LIST OF FIGURES

FIGURE Page

LIST OF TABLES

TABLE Page

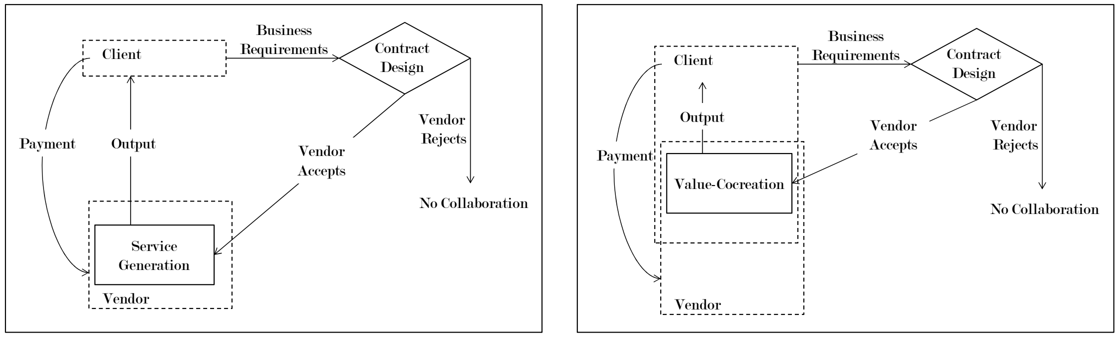
1. INTRODUCTION

1.1 Value Co-Creation

Management of services has increasingly become more sophisticated and challenging. Attaining high productivity and efficiency levels in service supply chains has been argued to be difficult because of many factors such as higher customer expectations for better service with higher quality and lower prices, loosely structured delivery processes, shorter product life cycles, tailoring of products or services to meet customer needs, and in particular, significant customer involvement throughout the process (Bettencourt et al. 2002). In the first two essays in my dissertation that are presented in Sections 2 and 3, I contribute to the literature by studying the involvement of customer firms in generating a service in a dynamic environment in different settings.

1.1.1 Traditional versus Collaborative Environments

I begin with illustrating the traditional vendor-client relationship in Figure 1.1(a) where the efforts of two parties are substitutes. In such settings, the client (referred to as *she*) offers a payment to the vendor (referred to as *he*) based on her business requirements and the vendor assumes the whole responsibility in the generation of the service if he accepts the offer. Then, based on the vendor's performance in the creation of the service, the client receives value and makes a payment to the vendor. In the operations management literature, the relationship between a client and a vendor so far has been mostly studied in such traditional settings (Cachon and Netessine 2004). However, the specifics of the client-vendor relationship is changing towards creating value through value co-creation in service environments (Toppin and Czerniawska 2005, Roels et al. 2010) that is illustrated in Figure 1.1(b).



(a) Traditional Vendor-Client Environment (b) Collaborative Vendor-Client Environment

Figure 1.1: Traditional versus Collaborative Vendor-Client Relationship

As shown in Figure 1.1(b), value co-creation implies that joint efforts are required from both the client and the vendor in order to create and deliver the service. Services are becoming more complex, unstructured, and tailored to fit unique needs of particular clients (Bettencourt et al. 2002). Therefore, in order to attain successful outcomes, clients must participate in the creation of the services. In effect, clients get involved in the service projects in (i) the problem definition stage, (ii) the selection of the solution, and (iii) the implementation of the solution. For example, in engineering and R&D projects, the client may provide critical input in the form of know-how or specialized equipment, and assist the vendor in all stages of the project (Plambeck and Taylor 2006). On the other hand, the clients often rely on the expertise of their vendors that may bring in knowledge or components that are required for the success of the research. The value co-creation environment presented in Figure 1.1(b) can be called a general collaborative service or knowledge intensive service. Some examples of such services are information technology (IT) outsourcing, software projects, auditing, consulting, financial planning, engineering projects, and R&D projects (Cohen et al. 1996, Cachon and Netessine 2004, Løwendahl 2005, Roels et al. 2010).

1.1.2 Contributions

Collaborative business environments have been studied with the application of contract theory, e.g., cross-functional coordination (Kouvelis and Lariviere 2000) and supply chain coordination (Cachon and Netessine 2004). The contracts which are related to my work are in the domains of general collaborative services (Roels et al. 2010), construction (Bajari and Tadelis 2001), legal services (Rubinfield and Scotchmer 1993), and call centers (Hasija et al. 2008). However, none of these studies focuses on the dynamics of collaboration between the parties. Most of the supply chain collaboration literature has focused on static settings where the effort levels are stationary (Cachon and Netessine 2004). However, it is not always necessary for the parties to keep their effort levels constant throughout the collaboration.

Further, the literature generally focuses only on the final outcome and disregard any on-going value that is useful from the moment it is generated until the end of the collaboration. However, the relationship between a vendor and a client in a collaborative environment is usually an on-going process and dynamic in nature. Hence, any manager working with projects that are developed in useful increments should consider on-going value (Pigoski 1996). These increments might be different modules of a project or software that are useful to the client by themselves, or valuable knowledge co-created through the collaboration. Accordingly, other than some extension sections, I consider that the project's terminal value is less important compared to the on-going utility received from the collaboration, and can be ignored. This applies to cases where the parties work together until the end of the useful life of the service (Currie and Galliers 1999). For example, a system or application being developed in a collaborative project may become an essential part of the client's business during its development. One specific example is IT environments, where the systems are usually built and enhanced using the incremental development until the end of their useful lifetime, and the client can use the intermediate versions of the system (Pigoski 1996, Currie and Galliers 1999). Other examples include financial services, marketing, and continuous service improvement programs where the ongoing utility is essential and there is collaboration until the useful life of the service (Bhuiyan and Baghel 2005). Further, in some of the extension sections in the following sections, I consider scenarios where the client receives value both during and at the end of the collaboration. Note that a special case of such setting includes environments where the value is received only at the end of the collaboration.

In addition, collaboration, with a focus on dynamic environments, has not received considerable attention in the literature (Erhun and Keskinocak 2011). Karmarkar and Pitbladdo (1995) point out the importance of the involvement of customers in the service delivery process, and call for explicit modeling of service co-creation. According to Hopp et al. (2009) and Roels et al. (2010), there are many promising research opportunities pertaining to the management of service operations. With the first two essays in this dissertation, I respond to these calls by analyzing a dynamic setting that requires joint efforts from a client and a vendor for the generation of the output.

In the first essay that is presented in Section 2, I examine pure revenue sharing settings and the settings where the payment between the client firm and vendor firm depends on the effort spent by the vendor firm. In particular, I focus on the following contracts: (a) the effort dependent contract, (b) output dependent contract, and (c) the hybrid contract. In the second essay of my dissertation that is presented in Section 3, I further consider that the effort levels of both parties are not observable or verifiable. However, they can be monitored (with some cost) if the client chooses to do so. Based on whether the effort levels are observed or not, I study two contracts in Section 3. In the first case, effort levels are not monitored. Hence, a double moral hazard problem arises and the contract can be written only in terms of what is observable or verifiable. Therefore, in this case, an output dependent contract is utilized which might also be thought of as a revenue-sharing agreement. In the second scenario, the effort level of the vendor is monitored, and subsequently, an effort dependent contract is administered.

In Sections 2 and 3, I utilize a differential game approach in a single client and a single vendor setting. Differential games can be considered as a fusion of game theory and optimal control theory. Hence, they incorporate strategic decision making and continuous change simultaneously. Dockner et al. (2000) provides a general discussion of differential games. However, as Cachon and Netessine (2004) state, because of the inherent difficulty in solving differential games, there are a few studies that utilize differential games in operations and supply chain management literature. To the best of my knowledge, this is the first study that considers differential games in a value co-creation environment.

Based on the solutions of the models, I derive several useful managerial insights pertaining to many aspects of the value co-creation environments. Most importantly, I answer the questions regarding how the client should manage the collaborative relationship. Specifically, I identify which type of contract is better for the client under various scenarios. I find in Section 2 that, as long as the sensitivity of output to vendor's effort is not very high, the effort dependent contract is better than the output dependent contract for the client. On the other hand, if the output is very sensitive to the effort of the vendor, then the client should offer payments based on output in order to give enough incentive to the vendor to spend more effort and generate more value together. If the client's contract selection also includes the hybrid contract and if the output is moderately or highly sensitive to vendor's effort, it is better for the client to utilize a hybrid contract. This implies that the vendor should be offered a share of the output, as well as payments related to the effort he spends in the generation of the output. If the output is not much sensitive to vendor's effort, the client should prefer an effort dependent contract.

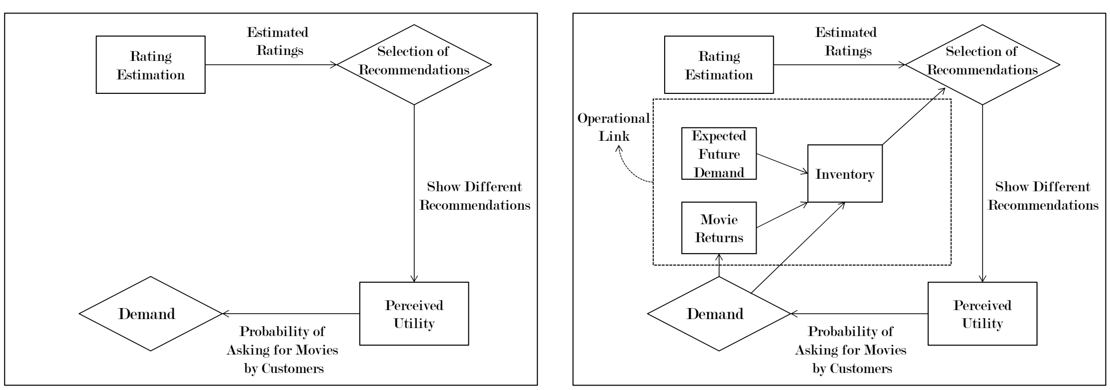
In addition, my analysis in Section 3 reveals whether the client should monitor vendor's effort or not, and how the severity of the double moral hazard problem changes across scenarios. I find that the vendor's effort should not be monitored when the participation cost of the vendor is very low. Further, my results show that the client should not monitor vendor's effort if the sensitivity of output to vendor's effort is relatively higher than the sensitivity to client's effort. I also derive the optimal payment terms for these contracts and the evolution of the equilibrium effort levels. Furthermore, I explore some extensions of the models.

1.2 Recommender Systems in DVD Rental Firms

The explosive growth of Internet has enabled easy access to a vast amount of information and led to the e-everything phenomena. E-commerce businesses grew into multi-billion firms like e-marketplace Amazon.com, or e-entertainment firms Netflix and Blockbuster. Amazon.com offers many different items, including millions of e-books for its kindle customers (Amazon.com 2011). Likewise, Netflix carries more than 140,000 different movie titles (Liedtke 2012). This is an overabundance of information for both firms and their customers (or users). Hence, finalizing a decision to buy or rent from the huge collection of these firms is a rather challenging task for the customers.

These firms can increase customer satisfaction by providing proper information to abridge the decision processes of their customers (Brynjolfsson et al. 2003, Murthi and Sarkar 2003). To this end, recommender systems are increasingly being used by electronic retailers to help customers find the products they are looking for (Senecal and Nantel 2004). Recommender systems can reduce the information overload and search complexity of the customers, and improve their decision quality (Xiao and Benbasat 2007). The tailored recommendations are also shown to increase both sales and customers' shopping experiences (De et al. 2010, Pathak et al. 2010). Hence, recommender systems are regarded as vital tools in sustaining Internet economy (Shapiro and Varian 1999).

E-entertainment firms, such as Netflix and Blockbuster, offer subscription based plans to their customers through DVD by mail or online streaming and use recommendations in both services. As I discuss later, DVD by mail business is vital to these firms because of the operating losses in online streaming, especially in international markets (Netflix 2012). Accordingly, the CEO of Netflix points out that they want to keep their DVD businesses as healthy as possible for many years ahead (Grover and Edwards 2011). Therefore, in this paper, I mainly focus on the DVD by mail businesses of these firms.



(a) Conceptual Working of Recommender System without Consideration of Inventory

(b) Conceptual Working of Recommender System with Consideration of Inventory

Figure 1.2: Significance of Considering Inventory in Recommender Systems

1.2.1 Significance of Considering Inventory in Recommender Systems

In Figure 1.2, I illustrate the conceptual working of recommender systems with and without the consideration of inventory. In this study, inventory accounts for not only the current status but also the expected future rentals and returns. When the inventory is not considered in the recommendation decision, then, as shown in Figure 1.2(a), first the recommender estimates customers' ratings for each movie. Next, the movies with high estimated ratings are recommended to the customers. This conceptual recommendation environment is user centric, i.e., it is designed to find out the movies that each customer will value the most (Schafer et al. 2001). Past studies have suggested that recommendations help the firm in shifting demand among movies – as if they were advertisements (Fleder and Hosanagar 2009). Hence, the recommender systems that are solely based on estimated ratings (without explicit consideration of inventory and its future pattern) do not fully utilize the effectiveness of recommendations.

Therefore, the conceptual model presented in Figure 1.2(a) does not result in higher customer satisfaction if the DVD-rental firm is not able to ship the requested movies to its customers. On the other hand, if the firm considers the expected request and return patterns and better utilizes its inventory information in the generation of the recommendations, it can shift the current demand among movies such that future shortages (and, therefore customer dissatisfaction) are reduced more effectively. To this end, Figure 1.2(b) presents the second conceptual framework that incorporates a feedback loop between the recommender system and the inventory. In this framework, inventory is affected by the recommendations. In turn, this inventory information is utilized, along with expected request and return patterns and estimated ratings, in the selection of recommendations that are shown to users.

1.2.2 Prevalent Industry Practice

The prevalent industry practice is to offer movies that (i) are estimated to be liked by customers (Hastings et al. 2008), and (ii) have immediate availability in the inventories without consideration of expected rentals and returns (Shih et al. 2009). Because all available movies with estimated ratings above a threshold level are recommended in the prevalent industry practice, I also refer to it as the *all-inclusive policy*. The rationale for using such a threshold is to maintain the trust of users on the recommender system (Tintarev and Masthoff 2007). In the long-run, users would request the recommended movies less frequently if they perceive that the recommender system offers movies that they do not like (Chen and Pu 2005, Kim et al. 2009). The recommendation threshold is a good practice per se, however, it does not necessarily maximize the satisfaction of customers because of the primitive use of inventory information.

The DVD-rental firms should utilize the operational link presented in Figure 1.2(b) and shift demand among movies effectively such that user dissatisfaction is reduced. However, the focus of the industry is not on better utilizing the recommendations. Rather, the focus of recommender systems has been on making better predictions, running fast and efficient, being able to function with sparse datasets, or sometimes offering a wider variety of products (Herlocker et al. 2004, Adomavicius and Tuzhilin 2005). Similarly, in DVD-rental firms the focus is on better prediction accuracy, e.g., see Netlix Prize discussed in Bennett and Lanning (2007). Academicians are also increasingly recognizing the fact that the recommender systems are not utilized in the most effective manner (Schafer et al. 2001, Bodapati 2008). I provide more details in Section 4.1 regarding how DVD-rental firms can better utilize recommendations. However, before that, I briefly outline my contributions in the following subsection.

1.2.3 Contributions

My analysis is based on the policies observed from DVD-rental firms and the discussions with managers of one such firm. In this paper, I answer several questions that are of managerial interest in this industry. I first ask and answer the following question: Does the usage of inventory information in recommender systems have a significant impact on customer satisfaction? The answer is yes. The major drawback of all-inclusive policy is that it is unable to shift demand from high-demand movies (that have current or expected future shortages) to low-demand movies (that have current or expected future excess supplies). In effect, I address the problem of finding the best set of recommendations while explicitly considering inventory and derive the conditions where the prevalent industry practice is not optimal. Besides, seemingly reasonable or greedy heuristics are shown to be non-optimal. For some special cases of the problem, I derive the optimal recommendation decisions that help managers in understanding the structure of the optimal solution.

The second issue is about the quality of the inventory information. Here, I ask the following question: Does the performance of my proposed method deteriorate when the inventory information is inaccurate? The answer is not necessarily yes. I find that the quality of the solution, i.e., expected number of unsatisfied customers, is not very sensitive to underestimation errors. However, when the inventory is overestimated, the quality of the solution is sensitive to the error. Interestingly, in a critically balanced system, where demand and inventory for most movies are close, my proposed solution approach may perform even worse than the prevalent industry practice when the inventory is overestimated. Hence, if there is uncertainty about the inventory levels, the firm should be conservative in the estimation, i.e., it should underestimate its inventory rather than overestimate.

The third issue I study is the trade-off between user trust in the recommender system and the short-term profitability of the firm. In this context, I answer the following question: How should the firms set the threshold in order to balance short-term profitability and long-term trust if they utilize my proposed approach? The recommendation threshold affects both short-term profitability (lower threshold is better) and long-term customer trust (higher threshold is better). My analysis reveals that the answer to this question depends on other characteristics of the firms. Therefore, the managers should be careful in setting the threshold level.

The fourth issue deals with the solution approach DVD-rental firms should employ. The problem I study is a stochastic problem. However, I carefully justify the validity of a deterministic approximation in the form of an integer program and use it for further analyses. The deterministic problem can be solved with two different approaches. The first approach uses estimated values for rentals and returns, and solves the optimization problem based on that. Secondly, I propose a practicable dynamic solution approach where the future decisions are based on the observation of actual rentals and returns in current and past periods. The question I ask is: Does the firm have to solve the problem dynamically? The answer is "not always." My analysis reveals that it is beneficial for the firm to adopt dynamically in a critically balanced system. However, if this is not the case, the firm may consider the solution of the static approach. This might help them in saving the costs related to updating parameter values and solving the problem in each period.

As discussed earlier, the focus of the recommendation literature and business practice is mostly on achieving better prediction accuracy. Hence, Schafer et al. (2001, p. 145) argue that the recommender systems should be made useful for the firms in other ways and calls for research that better integrates recommendations with other forms of techniques. In this direction, Bodapati (2008), Hosanagar et al.

(2008), and Garfinkel et al. (2008) state that the recommendations may not be the products that have the highest utility expectations. However, none of these studies consider the impact of recommendations on demand, which is the focus of my paper.

In summary, my research is motivated from both industry and academic perspectives. My study brings a fresh perspective to the recommender system research because I argue and show that recommendations should be tailored according to both the tastes of the users and the inventory status, along with expected future demand and returns.

2. PURE REVENUE SHARING AND PAYMENT PER HOUR MODELS IN VALUE CO-CREATION

2.1 Problem Definition

I study in this section a collaborative setting where a client and a vendor participate. In this section, I limit the discussion to pure revenue sharing contracts and the contracts that solely depend on the effort level of the vendor as these contracts are observed in business settings (Gil and Lafontaine 2012). In this setting, the client tries to find the contract that will maximize the value she receives from the collaboration, i.e., the difference between the value of the output and all costs related to the collaboration. On the other hand, the vendor tries to receive the maximum value he can receive from the collaboration. His value is the difference between the transfer payment he receives from the client and the costs related to the collaboration. As explained in the introduction section, the client or the vendor can receive on-going value from the project. Therefore, I model the problem in the forms of differential games that differ in payment terms. In the next section, I introduce different components of the models.

2.1.1 Input Parameters

The client's and vendor's effort levels are denoted by $u(t)$ and $v(t)$, respectively. These can be regarded as resources, such as labor-hours, exerted by each party. The effort levels are continuous variables, hence they are defined as functions of time. In this paper, t denotes the time instance, and T represents the total time horizon. The cost for client spending the effort level is modeled with a general power term structure, i.e., $c_c u(t)^\gamma$. The commonly used quadratic cost structure is a special case of this modeling approach. In this formulation, c_c is the cost multiplier for the effort the client spends. I assume that γ is more than 1 in order to reflect the fact that the marginal cost of effort increases with the level of effort. This means that the client chooses her least costly options first. In a similar way, the cost for the vendor is $c_v v(t)^\delta$ with cost multiplier term c_v and the cost elasticity term $\delta > 1$. In addition, γ and δ could be referred to as the costliness of client and vendor, respectively.

In this section, I assume that the effort levels are verifiable or observable. IT enabled services and cross-located personnel mainly help in monitoring these efforts. More specifically, Internet has facilitated real-time access to information across supply chains that enables decision models and software to take actions for streamlining supply chain operations (Swaminathan and Tayur 2003). RosettaNet (www.rosettanet.org) and the GS1 (www.gs1.org) are global standards organizations that enable collaboration and automation of transactions across industries and in global supply chains (Erhun and Keskinocak 2011). These standards and other IT enabled services have benefits such as real-time data transfer and automated communication that enable firms to reduce contract costs, planning times, number of manual transactions, and administrative costs (Erhun and Keskinocak 2011). Hence, IT creates an environment that fosters timely reporting, interaction, and visibility that are required in a co-value creation environment (Mentzer et al. 2000, Erhun and Keskinocak 2011).

The settings that I focus in this section, i.e., collaborative or knowledge intensive services, require even more than just IT enablers. The integrated parties in the collaboration probably even have personnel located in the other party's sites, they may hold regular meetings, might have presentations, etc. Because of this high level of integration between the client and the vendor, it is reasonable to consider that the effort level of each party is observable. I assume in this section that the variable portion of the monitoring costs are negligible. In addition, the fixed part of this cost is sunk, hence I do not consider monitoring costs explicitly in this section. However, in Section 3, I explicitly consider monitoring costs in modeling the problem. Clearly, the cost parameters c_c and c_v can be easily estimated using the cost structures of the firms. The cost elasticities of client's and vendor's effort level (i.e., γ and δ, respectively) can also be estimated based on firms' overtime policies, hire-fire policies, etc. In the next section, I discuss the output parameters of the model.

2.1.2 Output Parameters

As discussed earlier, the client firms in many business settings get utility from the output during the collaboration, not just at the end. However, I do not rule out the possibility that the parties receive value also at the end of the collaboration as this case is analyzed in Section 2.4. Therefore, I model the output that is denoted by $q(t)$ as continuous, doubly differentiable, strictly concave for positive effort levels, and nondecreasing in effort levels. In addition, the instantaneous increase in output is because of the collaborative work between the client and the vendor. Hence, both the client and the vendor need to exert effort in order to increase the output. I consider the output as a Cobb-Douglas function similar to other studies (Cohen et al. 1996, Roels et al. 2010, Kim and Nettessine 2012). The Cobb-Douglas functional form ensures that if one party exerts higher levels of effort, then the other party will have incentive to do so as well. In effect, I model the instantaneous increase in the output as $\dot{q}(t) = u(t)^{\alpha} v(t)^{\beta}$ with $\alpha, \beta \geq 0$,[1] and $\alpha + \beta < 1$. Here, $\alpha + \beta < 1$ is actually related to the concavity of the output function. In effect, I do not need to assume $\alpha + \beta < 1$ if a set of more relaxed conditions, i.e., $\alpha < \gamma$ and $\beta < \delta$ holds. If these relaxed conditions are also not satisfied, then it would mean that there is increasing

[1] Please note that the Cobb-Douglas functional form is flexible in the sense that it can represent a setting in which the vendor assumes the whole responsibility in the creation of the output. This is possible by setting α to 0.

return to scale with respect to the effort levels. Clearly, it does not make practical sense, and therefore this constraint is imposed.

Cobb-Douglass functional form can represent the speed of improvement in output in different settings, e.g., in food and automotive industries (Cohen et al. 1996). The parameters α and β represent the output elasticity or sensitivity to client's and vendor's effort levels respectively. These parameters can be estimated using the data from the past projects. It should be expected that the relative weight of the output elasticity parameters will be different across different business settings. In a setting where the vendor assumes almost all responsibility in the generation of the output, it should be expected that α is close to 0. On the other hand, if it is a value co-creation setting, then both of these parameters should be expected to be greater than 0. Based on these conditions on the parameters, now I present the following remark. All proofs are provided in Appendix A.

Remark 1 *The following condition satisfies:* $(\gamma - \alpha)\delta - \beta\gamma > 0$.

Since the client gets value from the project while it is being developed, the value can be defined as

$$\int_0^T kq(t)dt. \tag{2.1}$$

In the above equation, k is used to convert the output to a utility measure. One possible interpretation of k is dollar value per unit output. Hence, I also refer to k as the valuation of the project. I would like to note that I do not discount the net values of the parties. This is because, I consider the planning horizon as a short or medium term. However, it is easy to modify the model for long term planning horizons by including the discount factor. I find that the consideration of the discount factor does not affect the key insights of this section. From the discussion above, $q(t)$ can be written as:

$$q(t) = \int_0^t u(s)^\alpha v(s)^\beta ds; \; q(0) = 0. \tag{2.2}$$

2.1.3 Payment Structures

As discussed earlier, the payment between the client and the vendor can be based on the effort level of the vendor. In this section, I also examine the business settings where the client firm offers a share of the output to the vendor firm. Specifically, in this case, a pure-revenue sharing contract is used that is based on the output (Gil and Lafontaine 2012). In addition, I also study a contract that is a combination of both the effort level of the vendor and the output. I analyze these cases in three different models. Since the planning horizon is usually short to medium term, I do not discount the net values of the parties in the analyses. However, it is easy to modify the model for long term planning horizon by including the discount factor. I find that the inclusion of discount factor does not affect the key insights of the paper.

In the first model, the payment is based on the effort level of the vendor. In this case, the payment between the parties is $\int_0^T pv(t)dt$ where $p \geq 0$. Here, the parameter p denotes the payment per unit effort of the vendor. As will be explained later, I optimize the parameter p in order to maximize the value for the client since we are taking it as the dominant party in the collaboration. In the second model, the client transfers a portion of the revenues or savings to the vendor. Hence, the transfer between the parties is modeled to be $\int_0^T lq(t)dt$ with $l \geq 0$. I also optimize the payment parameter l. Finally, in the third scenario, I consider that the payment is due to both the effort level and the output, and define it as $\int_0^T pv(t)dt + \int_0^T lq(t)dt$ with p, $l \geq 0$. I limit the discussion to these prominent payment structures or contracts. I focus on analyzing their sensitivity with respect to different characteristics of the parties, and compare and contrast these contracts. Table 2.1 summarizes the

Symbol	Definition
Parameters	
c_c	Cost multiplier for client's effort
c_v	Cost multiplier for vendor's effort
k	Valuation of the project (i.e., client's value per unit output)
α	Output elasticity (or sensitivity) to the client's effort, $\alpha \geq 0$
β	Output elasticity (or sensitivity) to the vendor's effort, $\beta > 0$
γ	Cost elasticity (costliness) of the client's effort, $\gamma > 1$
δ	Cost elasticity (costliness) of the vendor's effort, $\delta > 1$
T	Length of planning horizon
Variables	
$q(t)$	Output at time t **(state variable)**
$u(t)$	Level of client's effort at time t **(control variable)**
$v(t)$	Level of vendor's effort at time t **(control variable)**
p	Transfer payment per unit vendor's effort **(decision variable)**
l	Transfer payment per unit output **(decision variable)**

Table 2.1: List of Parameters and Variables in Section 2

variables and the parameters used in the model. Next, I focus on the details of the model and the solution.

2.2 Model and the Solution

In the settings I analyze, the client is the principal and the vendor is the agent. Building the discussion in the reverse setting is also possible without any further complication. The client offers a business relationship to the vendor and the vendor accepts it if the gain is more than his reservation utility. For simplicity, I consider the reservation utility to be zero. However, with slight modifications, the results apply to the cases where the reservation utility is positive. If the offer is accepted by the vendor, both parties start working together and select and adjust their effort levels dynamically. Now, I analyze each of the three models discussed earlier.

2.2.1 Effort Dependent Contract

The client maximizes her value by optimizing the effort trajectory $u(t)$ throughout the planning horizon. Similarly, the vendor optimizes his effort trajectory $v(t)$ for all t. Hence, based on the earlier discussion, the objective functions of the client and the vendor, and the constraints can be written as:

$$\begin{aligned}
&\max_{u(t)} \quad \left\{ \int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - \int_0^T pv(t)dt \right\}, \\
&\max_{v(t)} \quad \left\{ \int_0^T pv(t)dt - \int_0^T c_v v(t)^\delta dt \right\},
\end{aligned}$$

$$\begin{aligned}
\text{subject to} \quad & \dot{q}(t) = u(t)^\alpha v(t)^\beta; \\
& \int_0^T pv(t)dt - \int_0^T c_v v(t)^\delta dt \geq 0; \\
& u(t) \geq 0;\ v(t) \geq 0.
\end{aligned}$$

The Hamiltonians for the client and the vendor (i.e., $H_c(t)$ and $H_v(t)$, respectively) can be written as below (Arrow and Kurz 1970):

$$\begin{aligned}
H_c(t) &= kq(t) - c_c u(t)^\gamma - pv(t) + \lambda_1(t) u(t)^\alpha v(t)^\beta, \\
H_v(t) &= pv(t) - c_v v(t)^\delta + \lambda_2(t) u(t)^\alpha v(t)^\beta.
\end{aligned}$$

Here, $\lambda_1(t)$ and $\lambda_2(t)$ are the adjoint variables for the client and the vendor respectively. These variables can be interpreted as the change in the corresponding objective functions of the client and the vendor for a small change in the state variable $q(t)$, i.e., the output (Sethi and Thompson 2000). In effect, the adjoint variables are the marginal value of changes in the output for the client and the vendor.

The Hamiltonian for the client (i.e., $H_c(t)$) consists of two parts. The first part is the integrand of the objective function for the client, and the second part is the adjoint variable $\lambda_1(t)$ times the right hand side of the state equation for the output. For a given instance of time t, the first part represents the direct contribution to the objective function of the client. The second part, however, characterizes the value of the changes in the rate of increase in output due to the decisions made at that specific time. Therefore, the second term represents an indirect contribution to the objective function. Consequently, the Hamiltonian needs to be maximized in each instance of time. In a given time instance, maximizing only the integrand would have been a myopic solution because of ignoring the indirect contribution of the increases in the output to the overall value. The maximum principle in dynamic optimization is a way to decouple the continuous problem into a set of static maximization problems in which the Hamiltonian acts as a surrogate objective to be maximized at each instance of time t (Sethi and Thompson 2000). The Hamiltonian for the vendor (i.e., $H_v(t)$) can be interpreted in a similar manner. Now, I present the equilibrium level of efforts for the client and vendor below in Lemma 1.

Lemma 1 *The equilibrium effort levels for the client and the vendor are:*

$$u(t) = \left(\frac{k(T-t)\alpha}{\gamma c_c} \left(\frac{p}{\delta c_v} \right)^{\frac{\beta}{\delta-1}} \right)^{\frac{1}{\gamma-\alpha}}; \quad v(t) = \left(\frac{p}{\delta c_v} \right)^{\frac{1}{\delta-1}}.$$

The equilibrium level of client's effort (i.e., $u(t)$) presented in Lemma 1 is strictly concave and decreasing with time. Furthermore, the client exerts no effort at the end of the planning horizon, because she does not have any utility from the collaboration after the project ends. Lemma 1 also implies that the equilibrium level of client's effort strictly increases as the payment per unit effort to the vendor (i.e., p) is increased. This result might seem counterintuitive, but can be explained as follows. If the client increases the payment per unit effort to the vendor, the vendor increases his effort level because he has a direct incentive to do so. In turn, the fact that the vendor is putting more effort into the project is an incentive for the client to increase her own effort level. Hence, if the client pays more per unit vendor's effort, it essentially creates an incentive to increase her own effort level. This is in line with the complementary nature of the business environment studied in this paper.

In contrast to the client's case, the equilibrium effort level of the vendor (i.e., $v(t)$) is constant throughout the planning horizon. This equilibrium level is essentially derived by equating the marginal gain of the vendor to the marginal cost of vendor's effort. This implies that vendor's equilibrium effort strictly increases in the per unit payment he receives from the client. Because neither of the marginal terms have time components, the vendor's effort stays constant throughout the planning horizon.

From a practical perspective, any increase in the transfer payment does not necessarily result in an increase in the value the client gets from the collaboration. Likewise, in the model I study in this section, the total net value the client gets from the project is concave in the payment parameter p. More specifically, it increases up to a specific level, and then starts to decrease. Hence, after substituting the equilibrium effort levels presented in Lemma 1 into the client's objective function, one can obtain the optimum level of payment using the first and second order conditions with respect to p. This optimal payment level is presented in Lemma 2.

Lemma 2 *The optimal payment level per unit vendor's effort is given by:*

$$p^* = \left(\frac{\beta(\gamma-\alpha)\gamma c_c}{\alpha(2\gamma-\alpha)\delta} \left(\frac{kT\alpha}{\gamma c_c} \right)^{\frac{\gamma}{\gamma-\alpha}} \right)^{\frac{(\gamma-\alpha)(\delta-1)}{(\gamma-\alpha)\delta-\beta\gamma}} (\delta c_v)^{\frac{\gamma(1-\beta)-\alpha}{\gamma(1-\beta)-\alpha\delta}} .$$

By substituting this payment term in Lemma 1, one can derive the equilibrium effort levels that maximize client's value.

2.2.2 Output Dependent Contract

In this setting, the client does not transfer payments based on the effort level of the vendor. Rather, the client offers a portion of the output, i.e., $\int_0^T lq(t)dt$, to the vendor. More specifically, I analyze a pure revenue sharing contract here and there is no notion of fixed payment between the parties in this setting (Gil and Lafontaine 2012). The costs for both the client and the vendor stay unchanged compared to the first model. Hence, the objective functions of client and vendor, and the constraints can be written as:

$$\max_{u(t)} \quad \left\{ \int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - \int_0^T lq(t)dt \right\},$$
$$\max_{v(t)} \quad \left\{ \int_0^T lq(t)dt - \int_0^T c_v v(t)^\delta dt \right\},$$

$$\text{subject to} \quad \dot{q}(t) = u(t)^\alpha v(t)^\beta;$$
$$\int_0^T lq(t)dt - \int_0^T c_v v(t)^\delta dt \geq 0;$$
$$u(t) \geq 0;\ v(t) \geq 0.$$

The equilibrium effort levels for the parties are now presented below.

Lemma 3 *The equilibrium effort levels for the client and the vendor are:*

$$u(t) = \left((T-t)^\delta \left(\tfrac{l\beta}{\delta c_v} \right)^\beta \left(\tfrac{(k-l)\alpha}{\gamma c_c} \right)^{\delta-\beta} \right)^{\frac{1}{(\gamma-\alpha)\delta-\beta\gamma}};$$
$$v(t) = \left((T-t)^\gamma \left(\tfrac{l\beta}{\delta c_v} \right)^{\gamma-\alpha} \left(\tfrac{(k-l)\alpha}{\gamma c_c} \right)^{\alpha} \right)^{\frac{1}{(\gamma-\alpha)\delta-\beta\gamma}}.$$

Here the equilibrium effort level of the vendor (i.e., $v(t)$) changes with time, in contrast to that in the effort dependent structure. The reason is that in this case the transfer payment depends on the output that changes with time. Hence, there is incentive for the vendor to adjust his effort level continuously. Moreover, effort levels of both parties are strictly concave and decreasing in time due to Remark 1. Besides, both effort levels converge to zero at the project completion time, because there is no terminal value at the end of the planning horizon. The equilibrium effort levels presented in Lemma 3 hold for any payment parameter l. After substituting the equilibrium effort levels given in Lemma 3 to the objective function of the client, it is easy to see that the value to the client is concave in l. Therefore, the first and second order conditions with respect to l reveal the optimal payment parameter that is presented below.

Lemma 4 *The optimal transfer payment per unit output to the vendor is given by:*

$$l^* = k\frac{\beta}{\delta}.$$

Hence, by setting the payment parameter l to the above level, the client maximizes the value for herself. It is an interesting finding that this term does not depend on the cost per unit effort terms, the sensitivity of the output with respect to client's effort, or the cost elasticity of the client's effort. Moreover, the optimal share that the client offers to the vendor $\frac{l^*}{k}$ does not depend on any characteristics of the client. Therefore, from the vendor's point of view, the share he gets from a project depends only on the sensitivity of the output to his effort (i.e., β) and the cost elasticity of his effort level (i.e., δ). More specifically, the vendor is awarded with a bigger share of the output when the output becomes more sensitive to his effort or if he can decrease the cost elasticity of his effort.

2.2.3 Hybrid Contract

In this contract, the objective functions of the client and the vendor, and the constraints can be written as:

$$\max_{u(t)} \left\{ \int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - \int_0^T lq(t)dt - \int_0^T pv(t)dt \right\},$$
$$\max_{v(t)} \left\{ \int_0^T lq(t)dt + \int_0^T pv(t)dt - \int_0^T c_v v(t)^\delta dt \right\},$$

$$\text{subject to} \quad \dot{q}(t) = u(t)^\alpha v(t)^\beta;$$
$$\int_0^T lq(t)dt + \int_0^T pv(t)dt - \int_0^T c_v v(t)^\delta dt \geq 0;$$
$$u(t) \geq 0;\ v(t) \geq 0.$$

Deriving the closed form solution of this problem is not possible, because this structure is similar to that in Abel (1881) that has been shown to be analytically intractable. However, it is possible to derive the closed form solution for special cases. One such case, for example, is $\delta = 1$. Nonetheless, I present some interesting experimental results for the hybrid case without any restrictions on the parameter values. I compare and contrast the performances of the three different settings numerically accompanied by sensitivity analysis on the model parameters. I present these results, along with many others, in the next section.

2.3 Discussion and Managerial Insights

In this section, I discuss the findings regarding different aspects of the problem and outline the managerial insights.

2.3.1 Effort Dependent Structure

I begin with discussing the results for the effort dependent structure. In this section, I first analyze the behaviors of client's and vendor's efforts in the equilibrium. Next, I study how the net values gained by both parties are impacted by different parameters.

2.3.1.1 Effort Levels and Payment Term

The equilibrium effort levels can be easily derived using the results of Lemmas 1 and 2. Now, in the next proposition, I analyze the impacts of c_c, c_v, and k on these effort trajectories and the optimal payment term per unit vendor's effort (i.e., p^*) derived in Lemma 2.

Proposition 1 *In the effort dependent structure:*

1. *As the client's valuation (i.e., k) increases, efforts of both parties as well as p^* increase.*
$$\frac{du}{dk} > 0, \quad \frac{dv}{dk} > 0, \text{ and } \frac{dp^*}{dk} > 0.$$

2. *As the cost multiplier term of client (i.e., c_c) decreases, efforts as well as p^* again increase.*
$$\frac{du}{dc_c} < 0, \quad \frac{dv}{dc_c} < 0, \text{ and } \frac{dp^*}{dc_c} < 0.$$

3. *As the cost multiplier term of vendor (i.e., c_v) increases, efforts decrease but p^* increases.*
$$\frac{du}{dc_v} < 0, \quad \frac{dv}{dc_v} < 0, \text{ and } \frac{dp^*}{dc_v} > 0.$$

Clearly, with an increase in the valuation of the project, the client increases her effort level in order to generate more output. Further, the client increases the payment term in order to entice the vendor to exert more effort. As a result, the vendor also increases his effort level. When the effort becomes less costly for the client, again the client increases her effort level (see part (b)). Moreover, because of the complementary nature of the project, the benefit of increased client effort is higher at the increased level of vendor's effort (see Equation (2.2)). Hence, interestingly, as c_c decreases, the client increases the payment term in order to ensure a higher effort level from the vendor as well.

Let us now analyze part (c) of the proposition that is interesting (and somewhat counterintuitive) in the sense that the impacts of change in c_v on effort levels and payment term are not the same. As the cost multiplier term of the vendor increases, the vendor tends to decrease his effort level. In order to mitigate this effect, the client increases the payment term. However this is not sufficient for inducing the vendor not to decrease his effort in the equilibrium solution. Therefore, because of the complementary nature of the project, the client also decreases her effort. Next, I analyze the impacts of other parameters on effort levels and the payment term.

Proposition 2 *In the effort dependent structure:*[2]

1. *With an increase in output elasticity to client's effort (i.e., α), client's effort increases iff $k > E_{\alpha c}$, and vendor's effort, as well as p^*, increases iff $k > E_{\alpha v}$.*

$$\frac{du}{d\alpha} > 0 \text{ iff } k > E_{\alpha c}.$$
$$\frac{dv}{d\alpha} > 0 \text{ and } \frac{dp^*}{d\alpha} > 0 \text{ iff } k > E_{\alpha v}.$$

2. *With an increase in output elasticity to vendor's effort (i.e., β), client's effort*

[2] The threshold values in this proposition and all the subsequent propositions are provided in Appendix A.

increases iff $k > E_{\beta c}$, and vendor's effort, as well as p^*, increases iff $k > E_{\beta v}$.

$$\frac{du}{d\beta} > 0 \text{ } iff \text{ } k > E_{\beta c}.$$
$$\frac{dv}{d\beta} > 0 \text{ } and \text{ } \frac{dp^*}{d\beta} > 0 \text{ } iff \text{ } k > E_{\beta v}.$$

3. *With an increase in cost elasticity to client's effort (i.e., γ), client's effort decreases iff $k > E_{\gamma c}$, and vendor's effort, as well as p^*, decreases iff $k > E_{\gamma v}$.*

$$\frac{du}{d\gamma} < 0 \text{ } iff \text{ } k > E_{\gamma c}.$$
$$\frac{dv}{d\gamma} < 0 \text{ } and \text{ } \frac{dp^*}{d\gamma} < 0 \text{ } iff \text{ } k > E_{\gamma v}.$$

4. *With an increase in cost elasticity to vendor's effort (i.e., δ), effort levels of parties decrease iff $k > E_{\delta c}$.*

$$\frac{du}{d\delta} < 0 \text{ } and \text{ } \frac{dv}{d\delta} < 0 \text{ } iff \text{ } k > E_{\delta c}.$$

When the output elasticity to client's effort (i.e., α) increases, the output improves at a faster rate with an increase in client's effort (see Equation (2.2)). As a result, when α increases, the client has an incentive to increase both her effort level and the payment term to the vendor (in order to incentivize vendor to increase her effort level as well). However, both of these also increase the cost of the client. Hence, the client needs to trade-off between the benefit and the cost of increased effort and payment term. Since the benefit of increased output is higher for high valuation projects (see Equation (2.1)), the payment term increases with α for such projects. Furthermore, as the time passes in the collaboration, the benefit of increased effort reduces for the client. Hence, towards the end of the project, it is less beneficial for the client to increase effort even for the high valuation projects. As a result, I find that $E_{\alpha c}$ increases in time that has passed in the collaboration and approaches infinity at the end of the project. This implies that, at the later stages of collaboration, the client decreases her effort level as α increases, irrespective of her valuation of the output.

The result in part (b) of the proposition can be explained in a similar manner. More specifically, when the output elasticity to vendor's effort (i.e., β) increases, the client has an incentive to increase vendor's payment in order to extract more effort from the vendor. However, as explained earlier, it is beneficial for the client to do so only when the valuation of the project is high enough. Let us now analyze part (c) of the proposition. Clearly, the client's effort level is comparatively high for the high valuation projects (see Proposition 1(a)). Hence, in such projects, an increase in γ increases the client's cost at a higher rate (as discussed in Section 2.1.1). Therefore, the client has an incentive to reduce her effort level with an increase in γ despite the fact that the reduced effort level decreases the output. This, in turn, also incentivizes the client to reduce the payment term to the vendor. Consequently, the vendor reduces his effort level as well.

Finally, as the cost elasticity to vendor's effort increases, the vendor has an incentive to reduce his effort. Hence, due to the collaborative nature of the project, the client also tends to decrease her effort level. However, the decrease in effort levels of both parties also decreases the output. Hence, the client needs to balance the tradeoff between the decrease in output and the increase in cost savings. In the high valuation projects, the effort levels are higher for both parties. Hence, for such projects, the reduction in cost because of reduced effort is higher than the decrease in output. As a result, the effort levels of both parties decrease with an increase in δ.

2.3.1.2 Net Values to the Parties

So far I have discussed the nature of the optimal payment term and the equilibrium effort levels. Now I present and analyze the results for net values gained by both parties in the project. These results are derived using Lemmas 1 and 2.

Lemma 5 *In the effort dependent structure, given the expression of p^* in Lemma 2:*

1. *The client's net value from the project is*

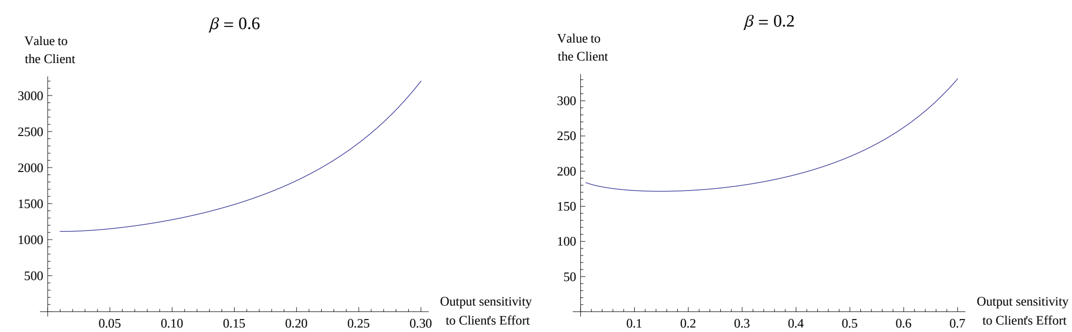
$$T\left(\frac{(\gamma-\alpha)^2 c_c}{\alpha(2\gamma-\alpha)}\left(\frac{kT\alpha}{\gamma c_c}\left(\frac{1}{\delta c_v}\right)^{\frac{\beta}{\delta-1}}\right)^{\frac{\gamma}{\gamma-\alpha}}(p^*)^{\frac{\beta\gamma}{(\delta-1)(\gamma-\alpha)}}-\left(\frac{1}{\delta c_v}\right)^{\frac{1}{\delta-1}}(p^*)^{\frac{\delta}{\delta-1}}\right).$$

2. *The vendor's net value from the project is*

$$T\left(\frac{\delta-1}{\delta}\right)\left(\frac{1}{\delta c_v}\right)^{\frac{1}{\delta-1}}(p^*)^{\frac{\delta}{\delta-1}}.$$

It is easy to observe in Lemma 5 that the value either party gets from the collaboration decreases with the cost per unit effort of the parties. Clearly, this result is intuitive and expected. Next, I use numerical experiments to derive some interesting (and not very intuitive) results regarding the relationship between net values to the parties and output elasticities to effort levels.

Figure 2.1 presents an example with the following parameter values: $T = 10$, $\gamma = 2.1$, $\delta = 1.1$, $k = 3.6$, $c_c = 2$, and $c_v = 1$. These values are reasonable in the sense that vendor's cost parameters are less than the client's, i.e., $\delta < \gamma$ and $c_v < c_c$. The value of β is 0.6 and 0.2 in Figures 2.1(a) and 2.1(b), respectively. Therefore, Figure 2.1(a) (resp., 2.1(b)) represents high (resp., low) output sensitivity to vendor's effort. In Figure 2.1(a), client's net value always increases as the output becomes more sensitive to her effort. Let us now elaborate this result. In this example, as α increases, the client increases her effort level in order to leverage the improved output elasticity. Hence, because of the complementary nature, the vendor also exerts more effort. Since the output is highly sensitive to vendor's effort (i.e., $\beta = 0.6$), the value of incremental increase in output (because of increased efforts) is more than the cost of increased effort for the client. Therefore, the net value of client always increases with an increase in α. On the other hand, in Figure 2.1(b), the client's net value decreases with α at the low values of α. In this case, since the output elasticity to vendor's effort is low (i.e., $\beta = 0.2$), the value of incremental increase in output (because of increased efforts) is less than the cost of increased effort for the client.



(a) High Output Elasticity to Vendor's Effort (b) Low Output Elasticity to Vendor's Effort

Figure 2.1: Impact of the Output Elasticity to Client's Effort

In order to check the robustness of these results, I conducted the experiments for several different parameter settings. The results show that the behaviors presented in Figure 2.1 are consistent across all scenarios. Furthermore, I find that the results are similar when vendor's net value is considered in the experiment instead of client's net value. I now summarize the findings of the experiments in the following observation.

Observation 1 *In the effort dependent structure:*

1. *When the output sensitivity to effort level is low for both parties, an increase in the output sensitivity to either parties' effort may decrease the net values of both parties.*

2. *When the output sensitivity to effort level is high for at least one party, an increase in the output sensitivity to either parties' effort always increases the net values of both parties.*

2.3.2 Output Dependent Structure

Again, I begin with analyzing the behaviors of client's and vendor's efforts in the equilibrium. After that, I study how the net values gained by both parties are impacted by different parameters.

2.3.2.1 Equilibrium Effort Levels

Using the results in Lemmas 3 and 4, one can easily characterize the effort levels of both parties in the output dependent structure. Now I analyze the impacts of different parameters on these effort levels. Some of these results and their explanations are similar to those in the effort dependent structure. However, since they are not exactly the same and some of the insights are unique in this setting, I briefly discuss these results. I begin with studying the impacts of the cost multipliers for effort levels (i.e., c_c and c_v) and the client's valuation (i.e, k).

Proposition 3 *In the output dependent structure, efforts of both parties increase with an increase in k and with a decrease in c_c or c_v.*

$$\frac{du}{dk} > 0, \quad \frac{dv}{dk} > 0, \; \frac{du}{dc_c} < 0, \; \frac{dv}{dc_c} < 0, \; \frac{du}{dc_v} < 0, \; \textit{and} \; \frac{dv}{dc_v} < 0.$$

Similar to that in the effort dependent structure, the client increases her effort level with an increase in valuation of the project (i.e., k). Hence, because of the complementary nature of the project, the vendor also increases his effort. On the other hand, the client reduces her effort level when her cost per unit effort (i.e., c_c) increases. Again, because of the complementarity, the vendor also decreases his effort. Finally, the impact of the vendors' cost per unit effort (i.e., c_v) can be explained in a similar manner. Now, in the next proposition, I present the impacts of α, β, γ, and δ on the effort levels of both parties.

Proposition 4 *In the output dependent structure:*

1. *With an increase in output elasticity to client's effort (i.e., α), client's effort increases iff $k > O_{\alpha c}$ and vendor's effort increases iff $k > O_{\alpha v}$.*

$$\frac{du}{d\alpha} > 0 \text{ } iff \text{ } k > O_{\alpha c}.$$
$$\frac{dv}{d\alpha} > 0 \text{ } iff \text{ } k > O_{\alpha v}.$$

2. *With an increase in output elasticity to vendor's effort (i.e., β), client's effort increases iff $k > O_{\beta c}$ and vendor's effort increases iff $k > O_{\beta v}$.*

$$\frac{du}{d\beta} > 0 \text{ } iff \text{ } k > O_{\beta c}.$$
$$\frac{dv}{d\beta} > 0 \text{ } iff \text{ } k > O_{\beta v}.$$

3. *With an increase in cost elasticity to client's effort (i.e., γ), effort levels of both parties decrease iff $k > O_{\gamma c}$.*

$$\frac{du}{d\gamma} > 0 \text{ } and \text{ } \frac{dv}{d\gamma} > 0 \text{ } iff \text{ } k > O_{\gamma c}.$$

4. *With an increase in cost elasticity to vendor's effort (i.e., δ), client's effort decreases iff $k > O_{\delta c}$ and vendor's effort decreases iff $k > O_{\delta v}$.*

$$\frac{du}{d\delta} < 0 \text{ iff } k > O_{\delta c}.$$
$$\frac{dv}{d\delta} < 0 \text{ iff } k > O_{\delta v}.$$

Similar to that in the effort dependent structure, when the output elasticity to client's effort (i.e., α) increases, the output improves at a faster rate with an increase in client's effort (see Equation (2.2)). As a result, the client has an incentive to increase her effort level when α increases. Moreover, since vendor's share of the output, i.e., $\frac{l^*}{k} = \frac{\beta}{\delta}$ (see Lemma 4), is unaffected by the increase in α, vendor's net value increases. Therefore, he has an incentive to increase his effort level as well. However, increasing their effort levels are costly for both parties. Hence, they need to consider the trade-off between the benefit and the cost of increased efforts. Since the benefit of increased output is higher for high valuation projects (see Equation (2.1)), both the client and the vendor increase their effort levels when α increases in such projects. Furthermore, as the time passes in the collaboration, the benefit of increased effort reduces. Hence, towards the end of the project, it is less beneficial for the parties to increase their efforts even for the high valuation projects. As a result, I find that the threshold levels in the proposition increase in time that has passed in the collaboration and approach infinity at the end of the project. This implies that, at the later stages of collaboration, both parties decrease their effort levels as α increases, irrespective of the valuation of the output.[3]

[3] All of the thresholds in Proposition 4 have similar properties with respect to time but I do not repeat the related discussion due to brevity.

Let us now analyze part (b) of the proposition. When the output elasticity to vendor's effort (i.e., β) increases, the vendor has an incentive to increase his effort because of two reasons. First, when β increases, the output improves at a faster rate with an increase in vendor's effort (see Equation (2.2)). Second, the payment received by the vendor increases in β (see Lemma 4). However, as explained earlier, it is beneficial for the parties to increase their effort levels only when the valuation of the project is high enough. Next, part (c) of the proposition states that both parties decrease their effort levels with γ in high valuation projects. In such projects, the client's effort level is comparatively high, and therefore her cost increases at a faster rate with γ. As a result, the client has an incentive to reduce her effort level with an increase in γ despite the fact that the reduced effort level decreases the output. Consequently, the vendor reduces his effort level as well.

Finally, as the cost elasticity to vendor's effort increases, the vendor has an incentive to reduce his effort because of two reasons. First, the payment received by the vendor decreases with δ (see Lemma 4). Second, with an increase in δ, it becomes costlier for the vendor to exert more effort. In turn, the client also has an incentive to decrease her effort level. However, the decrease in effort levels of both parties also decreases the output. Hence, as explained in the discussion of part (a), the effort levels of both parties decrease with an increase in δ only for the high valuation projects.

2.3.2.2 Net Values to the Parties

Using the results in Lemmas 3 and 4, it is easy to derive the net values of the parties. I present these expressions in the following lemma.

Lemma 6 *In the output dependent structure:*

1. *The client's net value from the project is:*

$$\frac{T^{\left(\frac{\beta\gamma+\alpha\delta-2\gamma\delta}{\beta\gamma+\alpha\delta-\gamma\delta}\right)\frac{(\gamma-\alpha)\delta-\beta\gamma}{(2\gamma-\alpha)\delta-\beta\gamma}}}{\left[\frac{k(\delta-\beta)}{\delta}\left(\left(\frac{k\alpha(\delta-\beta)}{\gamma\delta c_c}\right)^{\alpha\delta}\left(\frac{k\beta^2}{\delta^2 c_v}\right)^{\beta\gamma}\right)^{\frac{1}{\gamma\delta-\beta\gamma-\alpha\delta}}-c_c\left(\left(\frac{k\alpha(\delta-\beta)}{\gamma\delta c_c}\right)^{\delta-\beta}\left(\frac{k\beta^2}{\delta^2 c_v}\right)^{\beta}\right)^{\frac{\gamma}{\gamma\delta-\beta\gamma-\alpha\delta}}\right]^{-1}}.$$

2. *The vendor's net value from the project is:*

$$\frac{T^{\left(\frac{\beta\gamma+\alpha\delta-2\gamma\delta}{\beta\gamma+\alpha\delta-\gamma\delta}\right)\frac{(\gamma-\alpha)\delta-\beta\gamma}{(2\gamma-\alpha)\delta-\beta\gamma}}}{\left[k\beta\delta\left(\left(\frac{k\alpha(\delta-\beta)}{\gamma\delta c_c}\right)^{\alpha\delta}\left(\frac{k\beta^2}{\delta^2 c_v}\right)^{\beta\gamma}\right)^{\frac{1}{\gamma\delta-\beta\gamma-\alpha\delta}}-c_v\left(\left(\frac{k\alpha(\delta-\beta)}{\gamma\delta c_c}\right)^{\alpha}\left(\frac{k\beta^2}{\delta^2 c_v}\right)^{\gamma-\alpha}\right)^{\frac{\delta}{\gamma\delta-\beta\gamma-\alpha\delta}}\right]^{-1}}.$$

In order to derive relationships between the net values to the parties and the output elasticities to effort levels, I conduct similar experiments as those in Section 2.3.1.2. Based on the detailed experiments with several different parameter settings, I find that the results presented in Observation 1 are also valid for the output dependent structure. The explanations for these results are also similar to that in the effort dependent structure, and therefore not repeated here. I further explore this result by investigating the scenario when the output depends only on the vendor's effort. For this case, the impact of output elasticity to vendor's effort on the client's value is presented in the following proposition.

Proposition 5 *When the output depends only on the vendor's effort (i.e., $\alpha = 0$), the client's value decreases with β iff $\beta < e^{\frac{(\delta-\beta)^2}{2\delta(\beta-2\delta)}}\delta\left(\frac{c_v}{Tk}\right)^{\frac{1}{2}}$.*

This analysis further validates the experimental results presented in Observation 1. More specifically, $\alpha = 0$ in this proposition corresponds to part (a) of the observation. In this case, the proposition shows that the client's value decreases with β when β is low. Finally, the experiments show that the results presented in Observation 1 also hold for the hybrid contract.

2.3.3 Comparison of the Contracts

The purpose of this section is to compare the contracts analyzed in this section. This study would assist managers of the client firm in selecting the most beneficial contract while establishing a co-creation environment with a vendor. Before discussing the results, I should note that both effort dependent and output dependent structures are special cases of the hybrid structure. If one restricts the payment terms p or l to be zero, the hybrid structure becomes a pure output dependent or a pure effort dependent structure, respectively. Therefore, the hybrid structure cannot be strictly dominated by any of the other structures. However, the effort dependent and output dependent structures are more prominent in practice (Løwendahl 2005). Besides, managing a business collaboration with a hybrid contract is more challenging because of its more complicated nature. Hence, I study under what circumstances (i) the effort dependent structure or the output dependent structure dominates the other one, and (ii) the hybrid structure dominates both of the other structures.

Recall that the net values gained by the parties in the effort dependent structure and the output dependent structure are presented in Lemmas 5 and 6, respectively. Clearly, it is cumbersome to present the condition analytically when one structure is more beneficial than the other for the client. Hence, I present a representative numerical example in Figures 2.2 and 2.3 to illustrate the results. In this example, the parameter values are: $T = 10$, $\gamma = 2.5$, $\delta = 1.5$, $k = 3.6$, $c_c = 2$, $c_v = 3$ with different values of output sensitivity to client's and vendor's efforts (i.e., α and β). In the figures, $\{H, E, O\}$ stand for the optimal values for the client in hybrid, effort dependent, and output dependent structures, respectively. Figure 2.2 shows that the output dependent structure dominates the effort dependent structure only when the output sensitivity to vendor's effort is more than 0.7. This implies that if the output is highly sensitive to vendor's effort, then it is better for the client to offer a share of the output to the vendor. I verified that this result holds for other problem instances as well. Hence, I summarize the finding in the following observation.

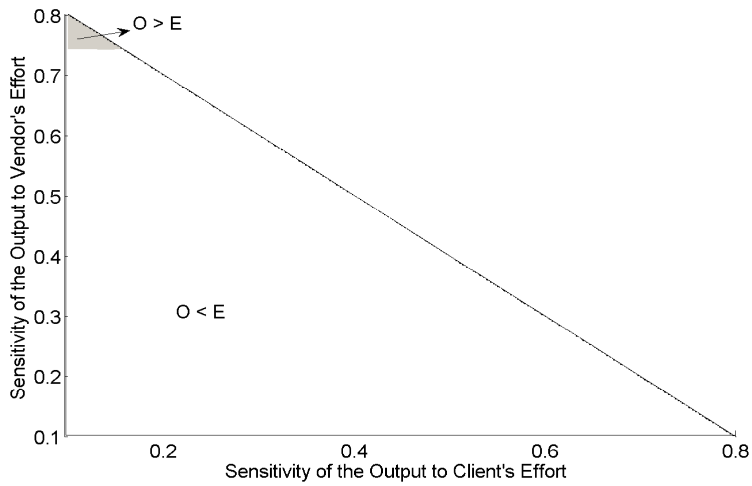


Figure 2.2: Comparison of Effort and Output Dependent Structures

Observation 2 *If the output sensitivity to vendor's effort is high, the output dependent structure is better than the effort dependent structure for the client; otherwise the effort dependent structure is better.*

Since the solution of the hybrid structure is not analytically tractable, I numerically determine the optimal transfer payment terms per unit vendor's effort and output (i.e., p^* and l^* respectively). For the majority of the parameter values, I observe that the hybrid structure converges to the effort dependent structure and that both are better than their output dependent counterpart. Interestingly, the hybrid structure never converged to the output dependent structure in the entire set of cases I analyzed. As shown in Figure 2.3, the hybrid structure converges to the effort dependent structure when the output sensitivity to vendor's effort is roughly less than 0.5. On the other hand, if it is more than 0.5, the hybrid structure dominates both of the other structures. This also implies that, if the output is more sensitive to the effort of the vendor, then the client is better off by offering a portion

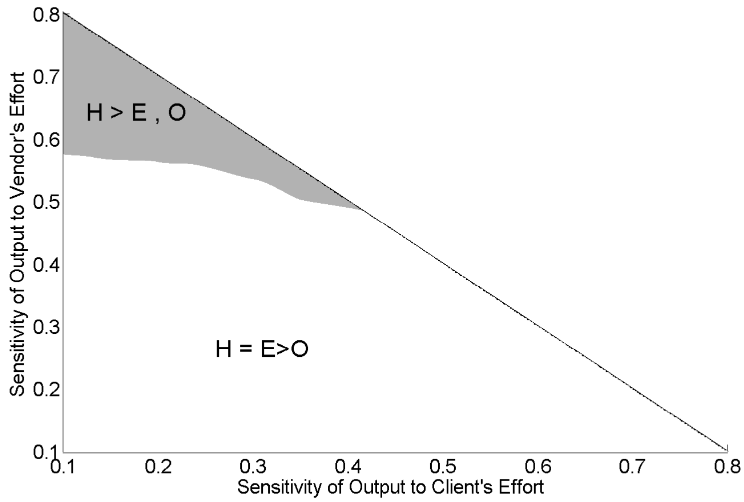


Figure 2.3: Comparison of the Hybrid Structure with the Other Structures

of the output to the vendor as well as paying for the vendor's efforts. I summarize the findings and discussion in the following observation.

Observation 3 *If the output is moderately or highly sensitive to vendor's effort, then it is better for the client to offer the hybrid structure. Otherwise, the effort dependent structure is as good as the hybrid structure and the client does not need to offer a share of the output to the vendor.*

I should also note that, if the effort levels of the parties are high, then the output sensitivity parameters can be thought of as reflecting the effectiveness levels of the parties. In other words, the parameters α and β would represent the effectiveness of the client and the vendor, respectively. Increasing one of these parameters, ceteris paribus, would result in higher output. Hence, Observation 3 implies that it is better for the client to offer a share as well as compensate for the effort of the vendor if he is very effective. This result can be applied to intra-organizational settings as well. For example, a firm might be considered as a client in that she is being served by all of its employees. All employees create value together for the firm and the executives can be thought of as the vendors that have high effectiveness versus the normal employees with relatively less effectiveness. Most publicly traded firms offer their executives incentive-based compensation plans, such as performance shares on the basis of corporate performance (Ferracone 2011). This compensation scheme is in parallel to my finding that if the effectiveness of the vendor is high (i.e., an executive), then the client is better off by offering a share of the output in addition to paying for the effort level of the vendor.

Finally, I analyze the scenario when $\alpha + \beta$ is a constant. Changing the sensitivity levels, while keeping their total constant, can be thought of as a shift in the assignment of responsibilities between the parties in the generation of the output. This result is presented in Figure 2.4 with $T = 10$, $\gamma = 2.5$, $\delta = 1.5$, $k = 3.6$, $c_c = 2$, $c_v = 3$, $\alpha + \beta = 0.9$ for the hybrid structure. This figure shows that the client's value is higher when the output is highly sensitive to either party's effort (i.e., high α or β) compared to the cases where it is equally sensitive to the efforts. In other words, when a single party assumes most of the responsibility in the collaboration, the client's value is higher. This result serves for comparative purposes, because depending on the nature of the project and characteristics of the parties, it may not be possible to shift the sensitivity of the output dramatically towards any of the parties in the design phase of the project. Therefore, the important insight here is the fact that client's net value would be higher in projects where the output depends more on the effort level of a single party. Moderate responsibility for both parties in the generation of the output might not be to the best interest of the client.

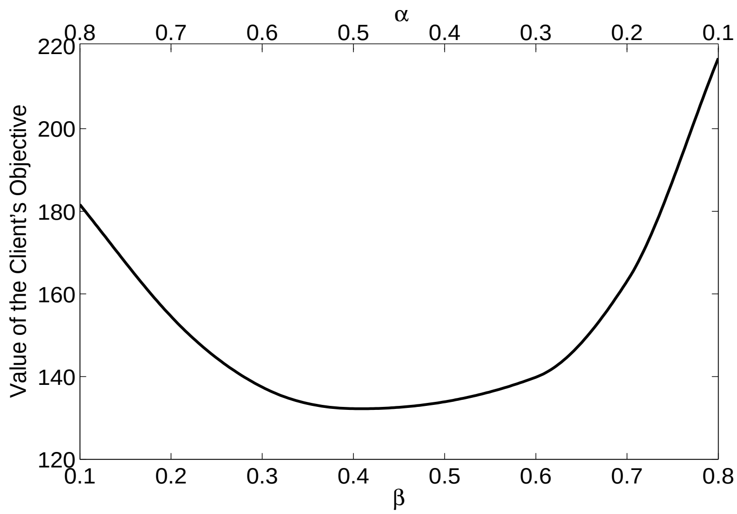


Figure 2.4: Client's Value with Respect to Assignment of Responsibilities

2.4 Extensions

In this section, I extend the analysis in different directions. First, I allow the client or the vendor to get utility from the output even after the project is finished in the following two subsections. Next, I consider a case where effort cost of the vendor decreases due to training or vendor's self-learning.

2.4.1 Salvage Value for the Client

Here, I consider that the client gets utility from the output both during the development time and after it is finished. For brevity, I present the results only for the effort dependent structure in this setting. Hence, given the salvage value parameter S_c, the objective function of the client is presented below. Vendor's objective function and other constraints are the same as those discussed in Section 2.2.1.

$$\max_{u(t)} \left\{ \int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - \int_0^T pv(t)dt + S_c q(T) \right\}$$

The Hamiltonians for the client and the vendor are the same as before. However, now $\lambda_1(T) = S_c$. This is the only difference from the earlier case. Next, in Lemma 7, I present the equilibrium effort levels.

Lemma 7 *In presence of the salvage value for the client, the equilibrium effort levels of the client and the vendor are:*

$$u(t) = \left(\frac{S_c + k(T-t)\alpha}{\gamma c_c} \left(\frac{p}{\delta c_v} \right)^{\frac{\beta}{\delta-1}} \right)^{\frac{1}{\gamma-\alpha}}; \; v(t) = \left(\frac{p}{\delta c_v} \right)^{\frac{1}{\delta-1}}.$$

Now, similar to the earlier structure, the client optimizes the level of p that maximizes her net value. This optimal payment level (i.e., p^*) is presented in the lemma below.

Lemma 8 *In presence of the salvage value for the client, the optimal payment per unit vendor's effort is given by:*

$$p^* = \left(\frac{ \frac{\left(S_c^{\frac{2\gamma-\alpha}{\gamma-\alpha}} - (S_c+kT)^{\frac{2\gamma-\alpha}{\gamma-\alpha}} \right) c_c}{\left(\frac{\alpha \left(\frac{1}{\delta c_v} \right)^{\frac{\beta}{\delta-1}}}{\gamma c_c} \right)^{\frac{-\gamma}{\gamma-\alpha}}} + \frac{(S_c+kT)^{\frac{\gamma}{\gamma-\alpha}} \left(S_c + kT - S_c \left(\frac{S_c}{S_c+kT} \right)^{\frac{\gamma}{\gamma-\alpha}} \right)}{\left(\left(\frac{\alpha}{\gamma c_c} \right)^{\alpha} \left(\frac{1}{\delta c_v} \right)^{\frac{\beta\gamma}{\delta-1}} \right)^{\frac{1}{\gamma-\alpha}}} }{ \frac{kT\left(2 - \frac{\alpha}{\gamma}\right)}{\frac{\beta}{\delta} \left(\frac{1}{\delta c_v} \right)^{\frac{-1}{\delta-1}}} } \right)^{\frac{(\gamma-\alpha)(\delta-1)}{(\gamma-\alpha)\delta-\beta\gamma}}.$$

As expected, if I set the salvage value to zero, the optimal payment term as well as the equilibrium effort levels converge to the case with no salvage value. I also find that if $S_c > 0$, the optimal payment parameter in this scenario is greater than that in the no salvage value case. In the presence of salvage value, the output is valuable to the client even after the project is completed. Hence, the client provides additional incentive to the vendor to spend more effort by increasing the payment term. As a result, due to the collaborative nature, the efforts from both parties are increased. In turn, the total output generated is more than that in the no salvage value case.

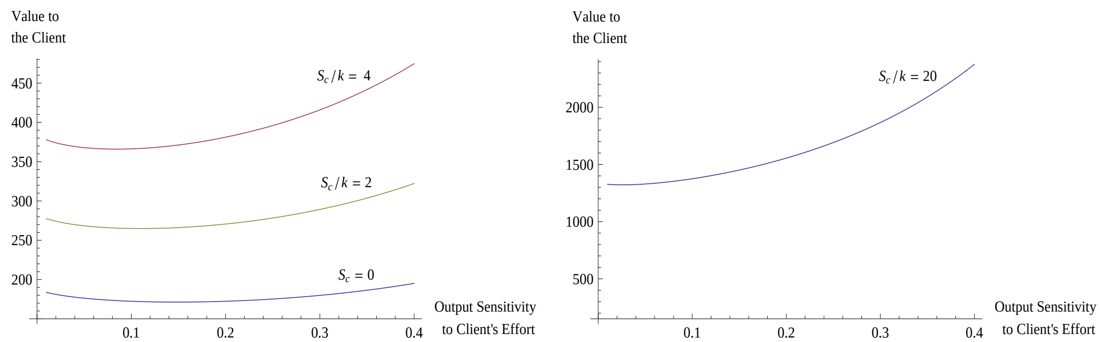
As can be seen in Lemma 7, the terminal effort level of the client is positive in this case (in contrast, it was zero in the no salvage value case). Next, although the vendor's effort in this case is more than that in the no salvage value case, it is again stationary throughout the collaboration period. Note that if I let $\frac{S_c}{k} \to \infty$, or set $k = 0$, which implies that the client does not get utility throughout the collaboration period but only at the project completion time, the model converges to a static model with no sense of continuous analysis except the accumulation of output that is valuable only at the end.

Interestingly, the relationship between value and output sensitivity terms that I discuss in Observation 1 changes its behavior in the salvage value settings. The mathematical conditions are cumbersome to present here, however, this behavior is observed in the entire set of problems I analyzed. I summarize this finding in the following observation.

Observation 4

1. *When the ratio of the values received by the client at the end of the project and during the collaboration (i.e.,* $\frac{S_c}{k}$*) is low, an increase in the output sensitivity to either party's effort (i.e.,* α *or* β*) may decrease the net values of both parties.*

2. *When the ratio of the values received by the client at the end of the project and during the collaboration (i.e.,* $\frac{S_c}{k}$*) is high, an increase in the output sensitivity to either party's effort (i.e.,* α *or* β*) always increases the net values of both parties.*



(a) Low Terminal Value

(b) High Terminal Value

Figure 2.5: Client's Value with Respect to Output Sensitivity to Client's Effort

This observation is illustrated in Figure 2.5 using the same parameter values as those in Figure 2.1(b). As shown in Figure 2.5(a), if one increases the importance of the terminal value in client's utility, the behavior of decreasing value with output sensitivity fades away. Figure 2.5(b) considers the same problem instance with $\frac{S_c}{k} = 20$ where I observe that this behavior is completely gone. Hence, as stated in the observation, if $\frac{S_c}{k}$ is large enough, the net values always increase with the output sensitivity to effort levels. Next, I study vendor's salvage value in the following subsection.

2.4.2 Salvage Value for the Vendor

In this scenario, the vendor is assumed to build reputation or history from what he has achieved in his previous business collaborations. His success, reflected in the output, might be considered by other potential clients as a signal of how successful the vendor would be in his future businesses. This can be considered as a salvage value of the project to the vendor. For brevity, I present the results only for the output dependent contract in this setting. Given the salvage value parameter S_v, the objective function of the vendor and the reservation utility constraint are presented below. Client's objective function and other constraints are the same as those in Section 2.2.2.

$$\max_{v(t)} \int_0^T lq(t)dt - \int_0^T c_v v(t)^\delta dt + S_v q(T)$$

$$\text{subject to} \quad \int_0^T lq(t)dt - \int_0^T c_v v(t)^\delta dt + S_v q(T) \geq 0.$$

Now, in the lemma below, I present the equilibrium effort levels.

Lemma 9 *In presence of the salvage value for the vendor, the equilibrium effort levels for the client and the vendor are:*

$$u(t) = \left(\left(\tfrac{(S_v + l(T-t))\beta}{\delta c_v}\right)^\beta \left(\tfrac{(k-l)(T-t)\alpha}{\gamma c_c}\right)^{\delta-\beta} \right)^{\frac{1}{(\gamma-\alpha)\delta-\beta\gamma}},$$
$$v(t) = \left(\left(\tfrac{(S_v + l(T-t))\beta}{\delta c_v}\right)^{\gamma-\alpha} \left(\tfrac{(k-l)(T-t)\alpha}{\gamma c_c}\right)^{\alpha} \right)^{\frac{1}{(\gamma-\alpha)\delta-\beta\gamma}}.$$

By comparing these equilibrium effort levels with those in Lemma 3, it is easy to see that the effort levels in the presence of salvage value for the vendor are greater than those without any salvage value given a specific level of the payment parameter l. Moreover, when $S_v = 0$, the equilibrium effort levels for both parties converge to the case with no salvage value. On the other hand, the determination of the optimal payment parameter l is not possible due to the necessity of integrating involved functions like Gauss hypergeometric function 2F1 (Gasper and Rahman 2004). However, I further analyzed this case via numerical approximation methods. The findings are summarized in the following observation.

Observation 5

1. *When $\frac{S_v}{k}$ is low, an increase in the output sensitivity to either party's effort (i.e., α or β) may decrease the net values of both parties.*

2. *When $\frac{S_v}{k}$ is high, an increase in the output sensitivity to either party's effort (i.e., α or β) always increases the net values of both parties.*

These results are similar to those in Observation 4 with salvage value for the client, and can be explained in a similar manner. Finally, in the next subsection, I analyze how the training or the self-learning impacts the decisions of the parties.

2.4.3 Training

In practice, the clients might engage in activities like supplier development programs or training their vendors in order to increase vendor productivity or lower their participation costs. In return, the client expects better value from the project. In this section, I consider that due to training or self-learning, the cost per unit effort for the vendor (i.e., c_v) decreases with time. I assume that the training effort is constant throughout the planning horizon. For example, the number of personnel who are responsible for vendor training might stay constant throughout the collaboration period.

Specifically, for a given training effort w, I model the change in the cost per unit effort for the vendor as $\dot{c}_v = -wc_v$. It implies that the instantaneous decrease in the cost is proportional to the training effort and the current level of the cost itself. This behavior is analogous to the loss of *goodwill* in advertising models (Nerlove and Arrow 1962). Hence, I have $c_v = c_0 e^{-wt}$. Here, c_0 denotes the starting level of the cost per unit effort. In this scenario, the cost for training is considered as $c_T w^\theta T$, where c_T is the cost multiplier for the training effort per unit time and θ is the power term. I consider $\theta > 1$, which implies that less costly training resources are utilized first. I can now write the objective functions of the client and the vendor and the constraints as:

$$
\begin{aligned}
\max_{u(t)} \quad & \int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - \int_0^T pv(t)dt - c_T w^\theta T, \\
\max_{v(t)} \quad & \int_0^T pv(t)dt - \int_0^T c_0 e^{-wt} v(t)^\delta dt, \\
\text{subject to} \quad & \dot{q}(t) = u(t)^\alpha v(t)^\beta; \\
& \int_0^T pv(t)dt - \int_0^T c_0 e^{-wt} v(t)^\delta dt \geq 0; \\
& u(t) \geq 0;\ v(t) \geq 0.
\end{aligned}
$$

The equilibrium effort levels for both parties are given in the following lemma.

Lemma 10 *With training, the equilibrium effort levels for the client and the vendor are:*

$$
u(t) = \left(\frac{k(T-t)\alpha}{\gamma c_c} \left(\frac{p}{\delta c_0} e^{wt} \right)^{\frac{\beta}{-1+\delta}} \right)^{\frac{1}{-\alpha+\gamma}};\ \ v(t) = \left(\frac{p}{\delta c_0} e^{wt} \right)^{\frac{1}{-1+\delta}}.
$$

In contrast to the no-learning case, the equilibrium level of the vendor's effort (i.e., $v(t)$) is not stationary. In fact, it strictly increases with time if the client provides training to the vendor, i.e., if $w > 0$. Moreover, client's effort is no longer strictly concave and decreasing in time. The first and second order conditions with respect to time reveal that client's effort has an increasing-decreasing behavior with the reflection point at $t = T - \frac{\delta-1}{w\beta}$. On the other hand, if $T < \frac{\delta-1}{w\beta}$, client's effort strictly decreases with time similar to that in the no-learning case. Besides, Lemma 10 also reveals that the equilibrium effort levels of both parties are always greater than their no training counterparts for a given level of payment per unit effort (i.e., p).

Analytical determination of the optimal payment parameter p and the training effort level w is not possible. However, I numerically optimize p and w that maximizes client's value in a variety of problem instances. In general, for a given payment term

p, as the training effort increases, the client's value first increases and then decreases. However, when the cost elasticity of vendor's effort (i.e., δ) is low, the client's value always decreases with an increase in the training effort. Hence, when δ is low, it is beneficial for the client to choose the lowest level of training effort, i.e., zero. I summarize this finding in the following observation.

Observation 6 *It is better for the client to provide training only if the cost elasticity of vendor's effort (i.e., δ) is high.*

I would like to note that, self-learning for the vendor is a special case of the training model I discuss in this section. In the self learning environment, $c_T = 0$, and w is not the effort level but the learning rate. This implies that the vendor becomes more efficient in time because of self-learning, not due to costly training efforts. I derive the closed form expressions in this case that are not reported due to brevity. However, my findings confirm that the benefit of vendor's self-learning is more beneficial to the client when the cost elasticity of vendor's effort (i.e., δ) is high.

2.5 Conclusions

In this paper, I analyze the value co-creation environments, considering the fact that the success of the project depends on both parties, not just the vendor. Therefore, I analyze a value co-creation environment where the output necessitates the efforts of both parties. I also consider the fact that both vendor and client may change their effort levels over time. Moreover, I allow the client to get utility from the output as it is being developed. Taking these facts into consideration, I examine three different settings from a differential game perspective. I derived the effort levels of both parties at the equilibrium and the optimal payment parameters in the contracts. I further present several interesting findings based on the sensitivity analyses of the equilibrium effort levels with respect to the parameters representing different characteristics of the parties.

I also study how different contracts compare under different settings in order to find the best contract for the client. I find that, as long as the sensitivity of output to vendor's effort is not very high, the effort dependent structure is better than the output dependent structure for the client. On the other hand, if the output is very sensitive to the effort of the vendor, then the client should offer payments based on output in order to give enough incentive to the vendor to spend more effort and generate more value together. If the selection includes the hybrid structure as well and if the output is moderately or highly sensitive to vendor's effort, it is better for the client to utilize a hybrid contract. This implies that the vendor should be offered a share of the output, as well as payments related to the effort he spends in the generation of the output. If the output is not much sensitive to vendor's effort, the client should prefer an effort dependent structure.

I also find that, if the output becomes more sensitive to the effort levels of either party, the client's value does not always increase. Similarly, the parties might not always increase their effort levels if the output becomes more sensitive to any of the effort levels. On the other hand, if the client receives substantial value from the project at the end of the collaboration, then the client's value always increases if the output becomes more sensitive to her effort. Finally, I consider that the client may offer training to their vendors in order to increase their productivity or lower their participation costs. I show that the client should provide training to only those vendors that have relatively high costliness.

3. IMPLICATIONS OF DOUBLE MORAL HAZARD PROBLEM IN VALUE CO-CREATION ENVIRONMENTS

3.1 Problem Definition and the First Best Scenario

Unlike Section 2, in this section, I assume that neither the client firm nor the vendor firm is aware of the other party's effort level throughout the collaboration. This leads the client with a double moral hazard problem, and I explore its implications for the collaboration in this section. I consider in this section a more general setting and hence the contracts are more flexible. Rather than just sharing revenues, or making a payment solely based on vendor's effort level without any other fixed payment term (which is the case in Section 2), I consider in this section that the contracts can also include fixed payment terms. The settings and the problems I am answering are different in Sections 2 and 3, yet some of the variables are defined very similarly to each other. In addition, some of the results and findings are very similar as well. However, for the sake of completeness, I provide all definitions in this section as well.

In this study, I consider a business setting where a client and a vendor enter into a collaborative agreement. The client's objective is to find the contract that will maximize her value, i.e., the difference between the value of the output and all costs related to the collaboration. On the other hand, the vendor maximizes his value in any setting, which is the difference between the payment he receives and the costs related to participating in the collaboration. As discussed in the introduction section, the client and the vendor receive on-going value from the project. Therefore, the settings I analyze incorporate both strategic decision making and dynamic changes. Hence, I model the problem as a differential game. As discussed earlier, I consider problem settings that differ based on whether the effort levels are monitored or not. Next, I introduce our model starting with the input parameters of the parties.

3.1.1 Input Parameters

I denote the effort levels of the client and the vendor at time t by $u(t)$ and $v(t)$, respectively. These levels represent the resources, like labor-hours, exerted by each party. The total time horizon of the collaboration is denoted by T. I model the cost of client's effort with a general power term structure as $c_c u(t)^\gamma$. Note that the commonly used quadratic cost function (e.g., Tsay and Agrawal 2000) is a special case of this structure. Here, c_c is the cost multiplier for client's effort. The parameter γ is considered to be more than 1 to reflect the fact that the marginal cost of effort increases with the level of effort. In other words, the client utilizes her least costly options first. Similarly, the cost for the vendor is $c_v v(t)^\delta$ with the corresponding cost multiplier c_v and the cost elasticity term $\delta > 1$. I also refer to γ and δ as costliness of client and vendor, respectively. Clearly, the cost parameters c_c and c_v can be easily estimated using the cost structures of the firms. The cost elasticities of client's and vendor's effort level (i.e., γ and δ, respectively) can also be estimated based on firms' overtime policies, hire-fire policies, etc.

In general, the effort levels are not observable or verifiable. However, in some environments, the effort levels can be monitored. Monitoring relieves the double moral hazard problem. However, if the effort levels are monitored, the costs related to monitoring need to be taken into account. In practice, IT enabled services and cross-located personnel mainly help in monitoring these efforts. Specifically, Internet has facilitated real-time access to information across supply chains that enables decision models and software to take actions for streamlining supply chain operations (Swaminathan and Tayur 2003). RosettaNet (www.rosettanet.org) and GS1 (www.gs1.org) are global standards organizations that develop common platforms for communication between the firms to enable collaboration and automation of transactions across industries and in global supply chains. Some of the benefits of these standards and other IT enabled services are real-time data transfer and automated communication that enable firms to reduce contracting costs (Erhun and Keskinocak 2011). IT, if successfully combined with processes and strategies, creates an environment that fosters timely reporting, interaction, and visibility that are required in a value co-creation environment. In addition, the parties in the collaboration generally have personnel located in the other party's sites, and they may hold regular meetings, conference calls, presentations, etc. Thus, the effort level of each party can be monitored with some cost.

In the problem settings I analyze, if one party's effort level is monitored, the other party's effort level can be inferred from the output that is observable. Therefore, monitoring both parties' effort levels is not optimal. Hence, in this section, only vendor's effort is monitored if the client chooses to do so. Monitoring cost comprises of fixed and variable costs. I denote the fixed component as F_m, e.g., investments in IT technologies. In addition, it is more costly to observe the effort if the effort level is increased. In other words, it is easier to observe the vendor's effort if it is low. As the vendor increases his effort level and utilizes other more costly options such as overtime, monitoring of these efforts should increase as well. Hence, similar to the cost of vendor's effort, the variable portion of monitoring cost at time t can be written as $c_m v(t)^\delta$. I would like to note that considering a power term different than δ in the monitoring cost does not change the key insights of this section. In this formulation, the parameter c_m is the cost multiplier for monitoring vendor's effort. Next, I discuss the output parameters of the model.

3.1.2 Output Parameters

In many business settings, the client gets utility from the output as the collaboration is in progress.[1] Therefore, I model the output, which I denote by $q(t)$, as continuous, doubly differentiable, strictly concave for positive effort levels, and non-decreasing in effort levels. Furthermore, the instantaneous increase in output is due to the collaborative work between the client and the vendor. Therefore, both parties need to exert effort in order to generate output. Similar to the other studies, the output is considered to be a Cobb-Douglas function (Cohen et al. 1996, Kim and Nettessine 2012). This implies that if one party exerts higher levels of effort, then the other party will have incentive to do so as well. More specifically, I model the instantaneous increase in the output as $\dot{q}(t) = u(t)^{\alpha} v(t)^{\beta}$ with $\alpha, \beta \geq 0$, and $\alpha + \beta < 1$.[2] The Cobb-Douglas functional form is flexible in the sense that it can represent a traditional setting where the vendor assumes the whole responsibility in the generation of the output. This is possible by setting α to 0.

Several past studies empirically justify using Cobb-Douglass functional form for the speed of improvement in output in automotive and food industries, e.g., see Cohen et al. (1996). Here, the parameters α and β represent the output elasticity (or sensitivity) to client's and vendor's efforts, respectively, that can be easily estimated using the data from the past projects. The relative weight of the output elasticity parameters should be different across different business settings. For example, in a traditional setting, where the vendor assumes almost all responsibility in the gen-

[1] As discussed in the introduction section, I do not rule out the possibility that the parties receive value also at the end of the collaboration. I study such a setting in Section 3.5.

[2] The constraint $\alpha + \beta < 1$ is actually related to the concavity of the output function. In effect, one does not need to assume $\alpha + \beta < 1$ if a set of more relaxed conditions, i.e., $\alpha < \gamma$ and $\beta < \delta$ holds. If these relaxed conditions are also not satisfied, then it would mean that there is increasing return to scale. Hence, in such a case, the parties could increase their effort levels arbitrarily to produce more output in a less costly manner. Clearly, it does not make practical sense, and therefore I impose this constraint.

eration of the output, α should be close to 0. On the other hand, if it is a value co-creation setting, then both of these parameters should be greater than 0.

Since the client gets value from the project while it is being developed, I can define it as

$$\int_0^T kq(t)dt. \tag{3.1}$$

I would like to note that I do not discount the net values of the parties in my analyses since the planning horizon is usually short to medium term. However, it is easy to modify the model for long term planning horizons by including the discount factor. I find that the consideration of the discount factor does not affect the key insights of this section. In the above equation, k is used to convert the output to a utility measure. One possible interpretation of k is dollar value per unit output. Hence, I also refer to k as the valuation of the project. From the discussion above, one can write $q(t)$ as

$$q(t) = \int_0^t u(s)^\alpha v(s)^\beta ds;\ q(0) = 0. \tag{3.2}$$

All the notations are summarized in Table 3.1.

3.1.3 Model Preliminaries

In the settings I analyze, the client is the principal and the vendor is the agent. Building the discussion in the reverse setting is also possible without any further complication. The client offers a business relationship to the vendor and the vendor accepts it if the gain is more than his reservation utility. These reservation utilities are denoted by R_c and R_v for client and vendor, respectively. If the offer is accepted by the vendor, both parties start working together and select and adjust their effort levels dynamically. I begin with analyzing the first best (FB) contract.

Symbol	Definition
c_c	Cost multiplier for client's effort
c_v	Cost multiplier for vendor's effort
c_m	Cost multiplier for monitoring vendor's effort
F_m	Fixed portion of the cost for monitoring vendor's effort
k	Valuation of the project (i.e., client's value per unit output)
R_c	Reservation utility of the client
R_v	Reservation utility of the vendor
S_c	Salvage value of the client
S_v	Salvage value of the vendor
α	Output elasticity (or sensitivity) to the client's effort level, $\alpha \geq 0$
β	Output elasticity (or sensitivity) to the vendor's effort level, $\beta > 0$
γ	Cost elasticity (costliness) of the client's effort level, $\gamma > 1$
δ	Cost elasticity (costliness) of the vendor's effort level, $\delta > 1$
T	Length of planning horizon
$q(t)$	Output at time t **(state variable)**
$u(t)$	Level of client's effort at time t **(control variable)**
$v(t)$	Level of vendor's effort at time t **(control variable)**
l	Transfer payment per unit output **(decision variable)**
F_d	Fixed transfer payment **(decision variable)**
$H(.)$	Transfer payment function **(decision variable)**

Table 3.1: List of Parameters and Variables in Section 3

3.1.4 First Best

The first best setting represents the ideal case. Here, the client and the vendor behave as if they are a single firm and the effort levels are observable and verifiable. As a result, total value that could be generated in the collaboration is maximized. Therefore, I compare the first best solution with the solutions of the output dependent and effort dependent contracts in Section 3.4 in order to study their relative performances. The single objective in the first best case is to maximize profits that is the difference between the total value of the output and the participation costs of the parties. Therefore, one can write the unified objective function and the constraints as the following:

$$\max_{u(t),v(t)} \left\{ \int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - \int_0^T c_v v(t)^\delta dt \right\} \tag{3.3}$$

subject to

$$\dot{q}(t) = u(t)^\alpha v(t)^\beta; \tag{3.4}$$

$$\int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - \int_0^T c_v v(t)^\delta dt \geq R_c + R_v; \tag{3.5}$$

$$u(t) \geq 0;\ v(t) \geq 0. \tag{3.6}$$

As discussed, the objective function in Equation (3.3) is simply the difference between the total value generated in the collaboration and the cost of effort for the client and the vendor. Equation (3.4) depicts how the output accumulates over time. In addition, the client and the vendor would collaborate only if they can generate value together more than the total of their reservation utilities R_c and R_v. This fact is reflected in Constraint (3.5). Lastly, Constraint (3.6) states the technical detail that the effort levels cannot be negative. The solution of this optimal control problem is provided in the following lemma. All of the proofs are available in Appendix B.

Lemma 11 *The optimal effort levels for the client and the vendor in the first best scenario are:*

$$u(t) = \left(\frac{k\alpha(T-t)}{\gamma c_c}\right)^{\frac{\delta}{\gamma(\delta-\beta)-\alpha\delta}} \left(\frac{\beta\gamma c_c}{\alpha\delta c_v}\right)^{\frac{\beta}{\gamma(\delta-\beta)-\alpha\delta}},$$
$$v(t) = \left(\frac{k\beta(T-t)}{\delta c_v}\right)^{\frac{\gamma}{\gamma(\delta-\beta)-\alpha\delta}} \left(\frac{\alpha\delta c_v}{\beta\gamma c_c}\right)^{\frac{\alpha}{\gamma(\delta-\beta)-\alpha\delta}}.$$

It is easy to see that the optimal effort levels presented in Lemma 11 decrease as the project progresses (i.e., as t increases). When the client and the vendor choose these effort levels, the total value generated in the collaboration is as provided in the lemma below.

Lemma 12 *The total value generated in the collaboration is:*

$$\frac{T(\gamma(\delta-\beta)-\alpha\delta)^2}{((2\gamma-\alpha)\delta-\beta\gamma)\gamma\delta}(kT)^{\frac{\gamma\delta}{\gamma(\delta-\beta)-\alpha\delta}}\left(\left(\frac{\alpha}{c_c\gamma}\right)^{\alpha\delta}\left(\frac{\beta}{c_v\delta}\right)^{\beta\gamma}\right)^{\frac{1}{\gamma(\delta-\beta)-\alpha\delta}}.$$

The client and the vendor collaborate if the value presented in Lemma 12 is more than the total of their reservation utilities, i.e., if Constraint (3.5) is satisfied. Later, I use this first best value to compare it with the values in other models.

3.1.4.1 Time-Variable Value

In some environments, unit value of the output in the collaboration might be more in the later stages of the project. Therefore, I analyze what happens if client's value per unit output k is not constant but increases in time, i.e., $\frac{dk(t)}{dt} > 0$. Should the effort levels increase as the project progresses, or should they again decrease similar to that in Lemma 11? I find that the effort levels still decrease through time even if the output becomes more valuable towards the end of the collaboration. In addition, I find that considering k as time-invariant or time-variant does not change the main insights of this section. Therefore, for simplicity, I treat k as a time-invariant constant hereafter in this section.

3.2 Unobservable Effort Levels

First best approach assumes that (i) effort levels are observable or verifiable, and (ii) the objective of both parties is to maximize the total value in the collaboration. However, in many real business settings, effort levels of parties are not observable and parties have different and often conflicting objectives. Therefore, a double moral hazard problem arises in many business settings. Accordingly, in this section, I assume that effort levels of both parties are not observable. Therefore, the client cannot transfer payments based on vendor's effort level. Clearly, the client can observe her own effort level and can infer the vendor's effort by observing the output.

However, she cannot write a contract based on vendor's effort. The reason is that in cases of breaches of the contract, the court should be able to verify the effort levels of both parties. However, neither party can verify or prove their effort levels to the court. Hence, such a contract is not practicable. Therefore, the client offers the vendor a portion of the output and a fixed fee, i.e., $\int_0^T lq(t)dt + F_d$. Here, l denotes the transfer payment per unit output, and F_d is the fixed fee. The structure of the costs for both client and vendor stays unchanged compared to the first best model. Hence, the objective functions of client and vendor, and the constraints can be written as:

$$\max_{u(t),l,F_d} \left\{ \int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - \int_0^T lq(t)dt - F_d \right\},$$
$$\max_{v(t)} \left\{ \int_0^T lq(t)dt + F_d - \int_0^T c_v v(t)^\delta dt \right\},$$

subject to

$$\dot{q}(t) = u(t)^\alpha v(t)^\beta;$$
$$\int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - \int_0^T lq(t)dt - F_d \geq R_c;$$
$$\int_0^T lq(t)dt + F_d - \int_0^T c_v v(t)^\delta dt \geq R_v;$$
$$u(t) \geq 0; \quad v(t) \geq 0.$$

In this setting, the client chooses her equilibrium effort level $u(t)$ and the vendor chooses his equilibrium effort level $v(t)$. The first constraint shows the evolution of output through time. The second and third constraints state that the client and vendor will participate in the collaboration only if they receive value more than or equal to their reservation utilities. The solution of this differential game for a given l reveals the equilibrium effort levels of the parties that is presented in the lemma below assuming that the participation constraints hold.

Lemma 13 *When the effort levels are not monitored, the equilibrium effort levels for the client and the vendor are:*

$$u(t) = \left((T-t)^{\delta} \left(\frac{l\beta}{\delta c_v} \right)^{\beta} \left(\frac{(k-l)\alpha}{\gamma c_c} \right)^{\delta-\beta} \right)^{\frac{1}{(\gamma-\alpha)\delta-\beta\gamma}},$$

$$v(t) = \left((T-t)^{\gamma} \left(\frac{l\beta}{\delta c_v} \right)^{\gamma-\alpha} \left(\frac{(k-l)\alpha}{\gamma c_c} \right)^{\alpha} \right)^{\frac{1}{(\gamma-\alpha)\delta-\beta\gamma}}.$$

Effort levels of both parties are strictly concave and decreasing in time. This is because $\frac{1}{(\gamma-\alpha)\delta-\beta\gamma} > 0$ is satisfied as shown in Remark 2 that is provided in Appendix B. In addition, both effort levels converge to zero at the project completion time, because there is no terminal value at the end of the planning horizon. These results are in line with the first best scenario. In the next subsection, I derive the optimal payment terms based on Lemma 13.

3.2.1 Optimal Payment Terms

In this contract, the client makes the vendor's participation constraint binding. The reason is that the client has no incentive to leave any value more than needed in order to make the vendor participate in the collaboration. Hence, $F_d = -\int_0^T lq(t)dt + \int_0^T c_v v(t)^{\delta} dt + R_v$. After substituting F_d and the equilibrium effort levels provided in Lemma 13 into the objective function of the client, it is easy to observe that the client's value is concave in l. Therefore, the first and second order conditions with respect to l reveal the optimal payment parameters that I present in Lemma 14. As discussed, the fixed fee that is provided in the same lemma is calculated in a way such that the vendor's net utility is equal to his reservation utility.

Lemma 14 *When effort levels of both parties are not observable, the optimal payment per unit output, and the fixed payment terms are given by:*

$$l^* = k\frac{\beta(\gamma-\alpha) - \sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}}{\beta\gamma - \alpha\delta},$$

$$F_d^* = T^{\frac{(2\gamma-\alpha)\delta-\beta\gamma}{(\gamma-\alpha)\delta-\beta\gamma}} \frac{(\gamma-\alpha)\delta-\beta\gamma}{(2\gamma-\alpha)\delta-\beta\gamma} \; l^* \left(\left(\frac{(k-l^*)\alpha}{\gamma c_c} \right)^{\alpha\delta} \left(\frac{l^*\beta}{\delta c_v} \right)^{\beta\gamma} \right)^{\frac{1}{(\gamma-\alpha)\delta-\beta\gamma}} \left(\frac{\beta-\delta}{\delta} \right).$$

By setting the payment parameters l and F_d to the levels presented in Lemma 14, the client maximizes the total value in the system and her utility. This is because the vendor receives only his reservation utility irrespective of the other parameter values. It is an interesting finding that the optimal payment term l^* does not depend on the cost per unit effort terms for both parties, i.e., c_c and c_v. I show in the proof of this lemma that l^* is between 0 and k, hence I can think of $\frac{l^*}{k}$ as the optimal share. The output elasticity terms α and β, and the cost elasticity terms γ and δ might be similar in different client-vendor dyads. Therefore, in a given industry or even across different industries, one would expect to see common share terms (e.g., 50%-50%, or 60%-40%) in revenue-sharing contracts irrespective of the participation costs of the parties, and the client's valuation of the project. This fact is actually evidenced in Bhattacharyya and Lafontaine (1995), and I complement their finding in the case of continuous value co-creation settings. On the other hand, the fixed payment term F_d^* depends on the cost per unit effort terms for both parties. Therefore, one would expect more variability in this fixed term in contracts. This finding is also observed in real business settings (Bhattacharyya and Lafontaine 1995).

As I discussed, the optimal payment term l^* is insensitive to the per unit cost terms c_c and c_v. Now, in the next proposition, I summarize the behavior of l^* with respect to output and cost elasticity terms α, β, γ, and δ.

Proposition 6 *The optimal payment parameter l^* strictly decreases with α and δ, and strictly increases with β and γ.*

Proposition 6 implies that if the output becomes more sensitive to client's effort (i.e., if α increases) or if vendor's marginal cost parameter increases (i.e., if δ increases), then the optimal payment term decreases. This result can be explained as follows. As α increases, the ratio of output sensitivity to cost sensitivity with respect to client's effort (i.e., α/γ) increases. This ratio represents client's productivity in a sense. Similarly, β/δ represents vendor's productivity. Therefore, as either α or δ increases, the relative productivity of the client (i.e., $\frac{\alpha/\gamma}{\beta/\delta}$) increases. Hence, the client tends to assume more responsibility in the collaboration. In other words, the client attempts to reduce the effort level of the vendor that is achieved by reducing the payment term. Using the similar argument, if the output becomes more sensitive to vendor's effort (i.e., if β increases) or if client's marginal cost parameter increases (i.e., if γ increases), the relative productivity of the vendor increases. Hence, in this case, it is beneficial for the client to increase the optimal payment term in order to entice vendor to work more in the collaboration.

3.2.2 Behavior of Effort Levels

Using the results in Lemmas 13 and 14, I can easily characterize the effort levels of both parties in the double moral hazard case. Clearly, efforts of both parties increase with k and decrease with c_c or c_v. Let us first explain this result. The client increases her effort level with an increase in valuation of the project (i.e., k) in order to increase the output. Hence, because of the complementary nature of the project, the vendor also increases his effort. On the other hand, the client reduces her effort level when her cost per unit effort (i.e., c_c) increases in order to reduce her effort cost. Again, because of the complementarity, the vendor also decreases his effort. Finally, the impact of the vendors' cost per unit effort (i.e., c_v) can be explained in a similar manner. Now, in the next proposition, I present the impacts of output sensitivity parameters and marginal cost terms on the effort levels of the parties.

Proposition 7 *In the output dependent structure:*[3]

1. *With an increase in output elasticity to client's effort (i.e., α), client's effort increases iff $k > O_{\alpha c}$ and vendor's effort increases iff $k > O_{\alpha v}$.*

2. *With an increase in output elasticity to vendor's effort (i.e., β), client's effort increases iff $k > O_{\beta c}$ and vendor's effort increases iff $k > O_{\beta v}$.*

3. *With an increase in cost elasticity to client's effort (i.e., γ), client's effort decreases iff $k > O_{\gamma c}$, and vendor's effort decreases iff $k > O_{\gamma v}$.*

4. *With an increase in cost elasticity to vendor's effort (i.e., δ), client's effort decreases iff $k > O_{\delta c}$ and vendor's effort decreases iff $k > O_{\delta v}$.*

When the output elasticity to client's effort (i.e., α) increases, the output improves at a faster rate with an increase in client's effort (see Equation (3.2)). As a result, the client has an incentive to increase her effort level when α increases. Moreover, because of the complementary effort levels, the vendor has an incentive to increase his effort level as well. However, increasing their effort levels are costly for both parties. Hence, they need to consider the trade-off between the benefit and the cost of increased efforts. Since the benefit of increased output is higher for high valuation projects (see Equation (3.1)), both the client and the vendor increase their effort levels when α increases in such projects. Furthermore, as the time passes in the collaboration, the benefit of increased effort reduces. Hence, towards the end of the

[3]The threshold values in this proposition and all the subsequent propositions are provided in Appendix B.

project, it is less beneficial for the parties to increase their efforts even for the high valuation projects. As a result, I find that the threshold levels in the proposition increase with the time elapsed in the collaboration and approach infinity at the end of the project. This implies that, at the later stages of collaboration, both parties decrease their effort levels as α increases, irrespective of the valuation of the output. All of the thresholds in Proposition 7 have similar properties with respect to time but I omit the details due to brevity.

Let us now analyze part (b) of the proposition. When the output elasticity to vendor's effort (i.e., β) increases, the output improves at a faster rate with an increase in vendor's effort (see Equation (3.2)). Hence, with an increase in β, the vendor has an incentive to increase his effort level. However, as explained earlier, it is beneficial for the parties to increase their effort levels only when the valuation of the project is high enough. Next, part (c) of the proposition states that both parties decrease their effort levels with an increase in cost elasticity to client's effort (i.e., γ) for high valuation projects. In such projects, the client's effort level is comparatively high, and therefore her cost increases at a faster rate with γ. As a result, the client has an incentive to reduce her effort level with an increase in γ despite the fact that the reduced effort level decreases the output. Consequently, the vendor reduces his effort level as well. Finally, as the cost elasticity to vendor's effort (i.e., δ) increases, it becomes costlier for the vendor to exert more effort. As a result, the vendor has an incentive to reduce his effort level with an increase in δ. In turn, the client also has an incentive to decrease her effort level. However, the decrease in effort levels of both parties also decreases the output. Hence, as explained in the discussion of part (a), the effort levels of both parties decrease with an increase in δ only for the high valuation projects.

3.2.3 Client's Net Value

Given the optimal payment terms in Lemma 14, I can calculate client's value that is presented below.

Lemma 15 *When effort levels of both parties are not observable, client's net value is* $T^{\frac{(2\gamma-\alpha)\delta-\beta\gamma}{(\gamma-\alpha)\delta-\beta\gamma}} \frac{(\gamma-\alpha)\delta-\beta\gamma}{(2\gamma-\alpha)\delta-\beta\gamma} \left(\left(\frac{(k-l^*)\alpha}{\gamma c_c} \right)^{\alpha\delta} \left(\frac{l^*\beta}{\delta c_v} \right)^{\beta\gamma} \right)^{\frac{1}{(\gamma-\alpha)\delta-\beta\gamma}} \left(k - \frac{(k-l^*)\alpha}{\gamma} - \frac{l^*\beta}{\delta} \right) - R_v,$ *where* $l^* = k\frac{\beta(\gamma-\alpha)-\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}}{\beta\gamma-\alpha\delta}$.

In Section 3.4, I analyze the behavior of the client's net value in the output dependent contract, and compare it with that in the first best solution and in the contract discussed in the next section.

3.3 Monitoring Vendor's Effort

If the effort levels of both parties are not observable, the client has a double moral hazard problem as I analyze in the previous section. However, as discussed in Section 3.1.1, in some settings, the vendor's effort can be made observable or verifiable through IT systems, regular meetings, site visits, submitting progress reports, etc. The cost for making vendor's effort observable has a fixed portion F_m as well as the variable part $c_m v(t)^\delta$, as discussed in Section 3.1.1. I consider that the vendor assumes the monitoring costs. However, I show later that the allocation of the monitoring costs between the client and the vendor does not affect the performance of the contract. In this environment, the payment function $H(.)$ can be written in terms of what is observable, i.e., output and vendor's effort. The client maximizes her value by optimizing the effort trajectory $u(t)$ throughout the planning horizon. Similarly, the vendor optimizes his effort trajectory $v(t)$ for all t. Hence, the objective functions of the client and the vendor, and the constraints can be written as:

$$\begin{aligned}
\max_{u(t)} \quad & \left\{ \int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - H(.) \right\}, \\
\max_{v(t)} \quad & \left\{ H(.) - \int_0^T (c_v + c_m) v(t)^\delta dt - F_m \right\}
\end{aligned}$$

subject to

$$\begin{aligned}
& \dot{q}(t) = u(t)^\alpha v(t)^\beta; \\
& \int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - H(.) \geq R_c; \\
& H(.) - \int_0^T (c_v + c_m) v(t)^\delta dt - F_m \geq R_v; \\
& u(t) \geq 0; \; v(t) \geq 0.
\end{aligned}$$

Because the vendor's effort is observable and the client's effort can be inferred from the output, the double moral hazard problem turns into a standard principal agent model with perfect information. In the next subsection, I introduce a contract that coordinates this setting.

3.3.1 Optimal Contract when the Vendor's Effort is Monitored

Now I present an optimal contract below where the client covers vendor's total cost and reservation utility if the vendor exerts effort more than a threshold.

Proposition 8 *When the vendor's effort is monitored, the optimal payment term is as follows:*

$$H(.) = \begin{cases} \int_0^T (c_v + c_m) \bar{v}(t)^\delta dt + F_m + R_v, & \text{if } v(t) \geq \bar{v}(t). \\ 0, & \text{if } v(t) < \bar{v}(t). \end{cases}$$

Here, $\bar{v}(t)$ is the value of $v(t)$ in the solution of **Problem P** *discussed below.*

Problem P:

$$\begin{aligned}
\max_{u(t),v(t)} \quad & \int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - \int_0^T (c_v + c_m) v(t)^\delta dt - F_m, \\
\text{subject to} \quad & \dot{q}(t) = u(t)^\alpha v(t)^\beta; \\
& \int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - \int_0^T (c_v + c_m) v(t)^\delta dt - F_m \geq R_c + R_v; \\
& u(t) \geq 0;\ v(t) \geq 0.
\end{aligned}$$

Because the client can write the contract based on vendor's effort, he requests the vendor to exert the effort level that would maximize the value in the system. Therefore, $\bar{v}(t)$ is the solution of the following system that considers both parties as a unified firm. The client chooses her effort level as a best response to vendor's effort $v(t)$, and covers only the effort cost of the vendor and the costs related to monitoring in addition to his reservation utility. Hence, the client claims all of the net value in the collaboration. This is the reason why the allocation of the monitoring cost (between the client and the vendor) does not affect the performance of the contract as discussed earlier.

Given the above system, it is easy to derive the equilibrium effort levels of the parties that are presented in Lemma 16. Obviously, if the value generated is less than the total of the reservation utility terms, i.e., $R_c + R_v$, the parties do not collaborate.

Lemma 16 *When the vendor's effort is monitored, the equilibrium effort levels for the client and the vendor are:*

$$\begin{aligned}
u(t) &= \left(\frac{k(T-t)\alpha}{\gamma c_c}\right)^{\frac{\delta}{\gamma(\delta-\beta)-\alpha\delta}} \left(\frac{\beta\gamma c_c}{\alpha\delta(c_v+c_m)}\right)^{\frac{\beta}{\gamma(\delta-\beta)-\alpha\delta}}, \\
v(t) &= \left(\frac{k(T-t)\beta}{\delta(c_v+c_m)}\right)^{\frac{\gamma}{\gamma(\delta-\beta)-\alpha\delta}} \left(\frac{\alpha\delta(c_v+c_m)}{\beta\gamma c_c}\right)^{\frac{\alpha}{\gamma(\delta-\beta)-\alpha\delta}}.
\end{aligned}$$

The equilibrium level of client's effort (i.e., $u(t)$) presented in Lemma 16 is strictly concave and decreasing with time. Furthermore, the client exerts no effort at the end of the planning horizon, because she does not have any utility from the collaboration after the project ends. This creates an incentive for the vendor to exert no effort at the end of the project as well.

3.3.2 Behavior of Effort Levels

In this section, I analyze the impacts of different characteristics of the client and the vendor on the effort trajectories presented in Lemma 16. I begin with the impacts of c_c, c_v, and k on the effort levels in the next proposition.

Proposition 9 *When the vendor's effort level is monitored:*

1. *As the client's valuation (i.e., k) increases, effort levels of both parties increase.*
2. *As the cost multiplier term of the parties (i.e., c_c and c_v) or cost of monitoring vendor's effort (i.e., c_m) decrease, effort levels of both parties increase.*

Clearly, with an increase in the valuation of the project, the client increases her effort level in order to generate more output. This entices the vendor to increase his effort level as well. Next, when exerting effort becomes less costly for the client (i.e., c_c decreases), the client increases her effort level (see part (b)), and therefore the vendor also increases his effort. Similarly, for the vendor, when either the unit effort cost (i.e., c_v) or the cost of monitoring effort (i.e., c_m) decreases, he increases his effort level. As a result, the client also increases her effort level. Next, I analyze the impacts of other parameters on effort levels and the payment term.

Proposition 10 *When the vendor's effort level is monitored:*

1. *With an increase in output elasticity to client's effort (i.e., α), client's effort increases iff $k > E_{\alpha c}$, and vendor's effort increases iff $k > E_{\alpha v}$.*

2. *With an increase in output elasticity to vendor's effort (i.e., β), client's effort increases iff $k > E_{\beta c}$, and vendor's effort increases iff $k > E_{\beta v}$.*

3. *With an increase in cost elasticity to client's effort (i.e., γ), client's effort decreases iff $k > E_{\gamma c}$, and vendor's effort decreases iff $k > E_{\gamma v}$.*

4. *With an increase in cost elasticity to vendor's effort (i.e., δ), client's effort decreases iff $k > E_{\delta c}$, and vendor's effort decreases iff $k > E_{\delta v}$.*

The output improves at a faster rate with an increase in client's effort when the output elasticity to client's effort (i.e., α) increases (see Equation (3.2)). As a result, the client has an incentive to increase her effort level for the purpose of improving output when α increases. Subsequently, the client tends to request more effort from the vendor by utilizing the contract presented in Proposition 8. However, increased efforts increase the participation costs of the client and the vendor in the collaboration. Therefore, the client should consider the trade-off between the benefits and the costs of increased efforts. Since the benefit of increased output is higher for high valuation projects (see Equation (3.1)), both parties increase their effort levels with α for such projects. In addition, the benefit of increased effort reduces in the later stages of the collaboration. Thus, the client has less incentive to increase her effort even for the high valuation projects towards the end of the project. In effect, I find that $E_{\alpha c}$ increases in time and approaches infinity at the end of the project. Therefore, irrespective of her valuation of the output, the client decreases her effort level as α increases at the later stages of collaboration. I would like to note that all of the thresholds in Proposition 10 have similar properties with respect to time. In addition, other parts of the proposition can be explained similarly. Hence, I do not repeat the discussion due to brevity.

3.3.3 Client's Net Value

Given the equilibrium effort levels presented in Lemma 16 and the contract presented in Proposition 8, I can derive the net value of the client that is presented below. By design of the contract, the vendor receives only his reservation utility.

Lemma 17 *When the vendor's effort is monitored, client's net value is:*

$$\frac{T(\gamma(\delta-\beta)-\alpha\delta)^2}{((2\gamma-\alpha)\delta-\beta\gamma)\gamma\delta}(kT)^{\frac{\gamma\delta}{\gamma(\delta-\beta)-\alpha\delta}}\left(\left(\frac{\alpha}{c_c\gamma}\right)^{\alpha\delta}\left(\frac{\beta}{(c_v+c_m)\,\delta}\right)^{\beta\gamma}\right)^{\frac{1}{\gamma(\delta-\beta)-\alpha\delta}}-F_m-R_v.$$

Now I analyze the behavior of client's value and several other results in the following section.

3.4 Discussion and Managerial Insights

In this section, I discuss our findings and outline managerial insights regarding how the value generated in the collaboration is affected by different characteristics of the parties and which contract is better under different circumstances.

3.4.1 Client's Net Value

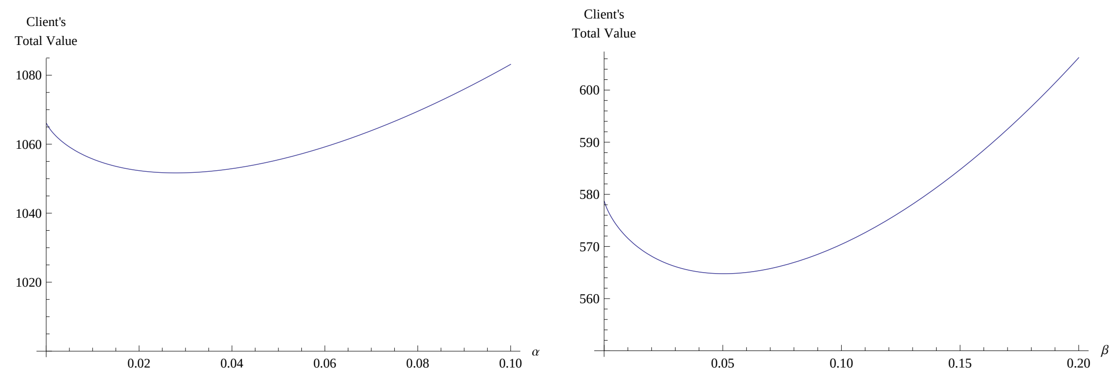
As discussed earlier, when the effort levels are not observed, an output dependent contract is utilized. On the other hand, when the effort level of the vendor is monitored, an effort dependent contract is used. In this subsection, I present the results and insights that are similar for both of these contracts. First, in both contracts, with the help of the fixed fee, the client extracts all value from the collaboration leaving the vendor only his reservation utility. Hence, I use client's value and total value interchangeably in this section. Next, it is easy to observe from Lemmas 15 and 17 that the value client gets from the collaboration increases with client's valuation of the project, i.e., k, and decreases with the cost per unit effort of the parties, i.e., c_c and c_v. In addition, client's value decreases with the parameters related to monitoring vendor's effort, i.e., c_m and F_m, if his effort is monitored. Clearly, these results are expected. Now I present some other results that are not very intuitive (and maybe somewhat counter-intuitive). In this direction, Proposition 11 discusses how the client's value is affected by the output sensitivity to effort levels.

Proposition 11 *In both output dependent and effort dependent contracts, when the output sensitivity to client's or vendor's effort (i.e., α or β) is low, an increase in the same output sensitivity decreases the net value of the client.*

$$\frac{d(\text{Client's Value})}{d\alpha} < 0 \text{ when } \alpha \text{ is low, and } \frac{d(\text{Client's Value})}{d\beta} < 0 \text{ when } \beta \text{ is low.}$$

Intuitively, if the output becomes more sensitive to the effort level of either party, client's net value should increase. However, as shown in Proposition 11, when the relative responsibility in the generation of the output is low for either the client (i.e., α is low) or the vendor (i.e., β is low), then increasing the responsibility of the same party does not necessarily increase client's value (or total value). However, after the sensitivity increases substantially, the value starts to increase. This result is illustrated using two examples in Figure 3.1. In this figure, I consider a scenario where the effort level of the vendor is monitored with the following parameter values: $T = 10$, $\gamma = 3$, $\delta = 2.5$, $k = 10$, $c_c = 2$, $c_v = 1.2$, and $c_m = 0.3$. These values are reasonable in the sense that the vendor's cost parameters are less than those of the client, i.e., $\delta < \gamma$ and $c_v < c_c$. Further, the value of β is 0.8 in Figure 3.1(a), and the value of α is 0.5 in Figure 3.1(b).

In both Figures 3.1(a) and 3.1(b), the client's net value increases with the sensitivity parameters only if these sensitivity parameters are not low. Let us now elaborate this result. As the sensitivity to the client's effort (i.e., α) increases, the



(a) Net Value with Output Elasticity to Client's Effort (b) Net Value with Output Elasticity to Vendor's Effort

Figure 3.1: Impact of the Output Elasticity Parameters on Client's Value

client increases or decreases her effort level (see Propositions 7 and 10). Let us first consider the scenario when the client's effort increases with α. In this case, when α is low, with an increase in α, the value of incremental increase in output (because of increased efforts) is less than the cost of increased effort for the client. On the other hand, when α is not low, since the output is relatively more sensitive to the effort levels, the value of incremental increase in output (because of increased efforts) is more than the cost of increased effort for the client. The similar argument exists for the case when the client's effort decreases with α. Hence, as shown in Figure 3.1(a), when α is low, the client's net value decreases with α, but the reverse is true at the higher values of α. The behavior with respect to β (in Figure 3.1(b)) can be explained similarly.

3.4.2 Comparison with the First Best

The first best approach assumes that the client and the vendor act as a single firm. In addition, it is assumed that the effort levels are observable and verifiable. Hence, first best presents the maximum value that can be generated in the collaboration. In this subsection, I examine how the contracts compare with the first best case under different circumstances. I start with the output dependent contract that is utilized in the double moral hazard case.

3.4.2.1 Performance of the Output Dependent Contract

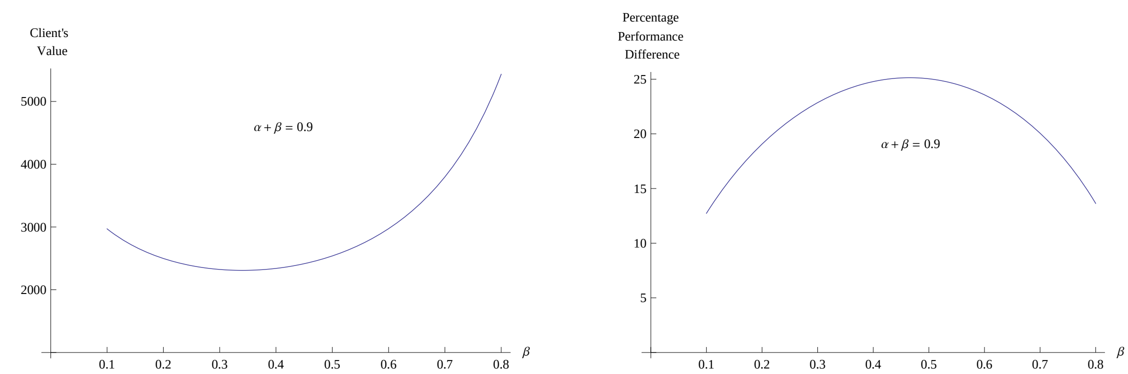
First, I conduct a numerical study to analyze how the percentage difference in client's value between the first best scenario and the output dependent contract changes with the output sensitivity parameters α and β. In this study, I find that as the output becomes more sensitive to the effort of either party, the percentage difference increases. In other words, the disadvantage of the unobservable effort levels amplifies when the output becomes more sensitive to the effort of either party. This is attributed to the fact that the client faces a more severe double moral hazard problem when the output becomes more sensitive to the effort of either party. The details of this numerical study are omitted for brevity, but I summarize the finding in the following observation.

Observation 7 *The percentage difference in client's value between the first best scenario and the output dependent contract increases as the output becomes more sensitive to either the client's effort or the vendor's effort (i.e., either α or β increases).*

I next analyze a scenario where the total of the sensitivity parameters (i.e., $\alpha+\beta$), is kept constant, instead of keeping one sensitivity parameter constant while varying the other. Changing the sensitivity parameters while keeping their total constant can be thought of as a shift in the assignment of responsibilities between the parties in the generation of the output. This analysis is presented in Figure 3.2 with $k = 12$, $T = 10$, $\gamma = 1.9$, $\delta = 1.7$, $c_c = 2$, $c_v = 1.6$, and $\alpha + \beta = 0.9$. Figure 3.2(a) shows that the client's value is higher when the output sensitivity is comparatively higher to one party's effort (i.e., higher α or β) compared to the case where the output is almost equally sensitive to the efforts of both parties (i.e., $\alpha \approx \beta$). In other words, when a single party assumes most of the responsibility in the collaboration, the client's value is higher. This result serves for comparative purposes, because depending on the nature of the project and the characteristics of the parties, it may not be possible to shift the sensitivity of the output dramatically towards any of the parties in the design phase of the project. Therefore, the important insight here is the fact that the client's net value would be higher in projects where the output depends more on the effort level of a single party. Moderate responsibilities for both parties in the generation of the output might not be to the best interest of the client in the output dependent contract. Now I summarize this finding below.

Observation 8 *In the output dependent contract, the client's value is higher when the output sensitivity is comparatively higher to one party's effort (i.e., higher α or β) compared to the case when the output is almost equally sensitive to the efforts of both parties (i.e., $\alpha \approx \beta$).*

Figure 3.2(a) does not provide any intuition regarding how the output dependent contract compares with the first best and whether the decreasing-increasing behavior is due to the game that is played between the parties. Therefore, in Figure 3.2(b), I present the percentage difference in client's net value between the first best case and the output dependent contract (while keeping $\alpha + \beta$ constant). Based on the finding in Observation 7, I may expect that as the output becomes more sensitive to the vendor's effort (i.e., as β increases), the client will face a more severe double moral hazard problem. Therefore, one may conclude that the performance difference between the first best case and the output dependent contract increases with β. However, as Figure 3.2(b) suggests, this intuition is not correct. When the output depends more on either party's effort, the performance difference between the



(a) Net Value in the Output Dependent Contract

(b) Percentage Difference between the First Best Case and the Output Dependent Contract

Figure 3.2: Output Dependent Contract: Impact of the Assignment of Responsibilities between the Parties

contracts is less. One possible explanation is that when the output depends more on the vendor's effort, the vendor assumes most of the responsibility in the generation of the output. This corresponds to high β values on the right hand side of the graph. On the other hand, when the output depends more on the client's effort, the client assumes most of the responsibility in the generation of the output. This corresponds to low β values on the left hand side of the graph. Therefore, because of the low responsibility of either party, the client has a "one-sided" moral hazard problem on both ends of the graph. However, when the output depends on both parties' efforts, the client faces a "double" moral hazard problem. Therefore, the output dependent contract performs worst around the middle of the horizontal axis in the figure. These findings are summarized in the following observation.

Observation 9 *The percentage difference in client's net value between the first best case and the output dependent contract is lower when the output sensitivity is comparatively higher to one party's effort (i.e., higher α or β) compared to the case when* *the output is almost equally sensitive to the efforts of both parties (i.e., $\alpha \approx \beta$).*

3.4.2.2 Performance of the Effort Dependent Contract

The first two results of the output dependent contract (i.e., Observations 7 and 8) are also valid in the effort dependent contract. Hence, for brevity, I do not discuss these results in detail. However, the third result (i.e., Observations 9) is not valid here. More specifically, when the total of the output sensitivity parameters (i.e., $\alpha + \beta$) is kept constant, the percentage difference in client's net value between the first best case and the effort dependent contract always increases with β (compared to the increasing-decreasing behavior in the output dependent contract). I illustrate this result in Figure 3.3 using $k = 12$, $T = 10$, $\gamma = 1.9$, $\delta = 1.7$, $c_c = 2$, $c_v = 1.2$, $c_m = 0.3$, and $\alpha + \beta = 0.9$. Since the vendor's effort is monitored in the effort dependent contract, the client does not face a moral hazard problem as it is the case in Figure 3.2(b). This explains why the behavior in Figure 3.3 is not inverse U-shaped. I summarize this result in the following observation.

Observation 10 *When the total of the output sensitivity parameters (i.e., $\alpha + \beta$) remains constant, the percentage difference in client's net value between the first best case and the effort dependent contract increases with the output sensitivity to vendor's effort (i.e., β).*

Now, in the next subsection, I compare output dependent and effort dependent contracts, and derive the conditions when one of them is preferred over the other.

3.4.3 Comparison of the Contracts

In this section, I compare the output dependent contract and the effort dependent contract with respect to different characteristics of the parties. This analysis would assist managers of the client firm in selecting the most beneficial contract while estab-

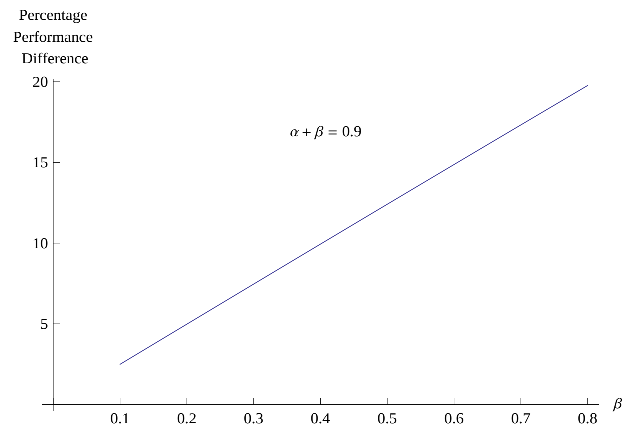


Figure 3.3: Effort Dependent Contract: Impact of the Assignment of Responsibilities between the Parties

lishing a value co-creation environment. Compared to the first best case, the output dependent contract has inefficiency due to unobservable effort levels. On the other hand, the effort dependent contract has inefficiency due to costs pertaining to the monitoring of vendor's efforts. Recall that the client's net value in output dependent and effort dependent contracts are presented in Lemmas 15 and 17, respectively.

If the cost multiplier for monitoring vendor's effort (i.e., c_m) or the fixed portion of the cost for monitoring (i.e., F_m) increases, the client tends to utilize the output dependent contract as it does not require the monitoring of vendor's effort. These results are monotonic as expected. However, as I study in this section, other characteristics of the parties affect client's selection of the contract in non-monotonic ways. In the next proposition, I begin with studying how vendor's cost multiplier term c_v affects which contract is better.

Proposition 12 *When the fixed cost for monitoring vendor's effort (i.e., F_m) is negligible, output dependent contract dominates the effort dependent contract iff the* *vendor's cost multiplier term (i.e., c_v) is sufficiently low. More specifically, the output dependent contract dominates iff*

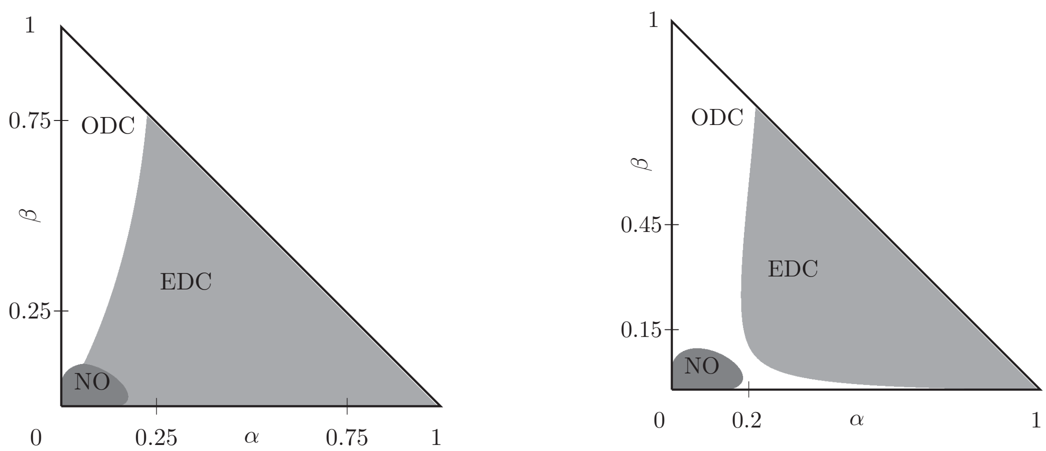
$$c_v < \left(\frac{\left(\frac{(\gamma-\alpha)\delta-\beta\gamma}{(\gamma-\alpha)(\delta-\beta)+\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}}\right)^{\frac{(\gamma-\alpha)\delta-\beta\gamma}{\beta\gamma}} \left(\frac{\beta\gamma-\alpha\delta}{\alpha(\beta-\delta)+\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}}\right)^{\frac{\alpha\delta}{\beta\gamma}}}{\frac{\beta\gamma-\alpha\delta}{c_m\left(\beta(\gamma-\alpha)-\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}\right)}} - \frac{1}{c_m} \right)^{-1}.$$

When the cost multiplier term c_v is low, the vendor's equilibrium effort is relatively high, and therefore the total monitoring cost is high. Hence, in such a case, the severity of the double moral hazard problem is less than the costs related to monitoring vendor's effort. Therefore, as shown in Proposition 12, when it is less costly for the vendor to exert effort, the client should choose not to monitor vendor's effort and utilize the output dependent contract. If the fixed cost for monitoring vendor's effort (i.e., F_m) is not negligible, then the threshold value for c_v is higher than the value presented in Proposition 12. I do not present the expression due to brevity, but the key insight remains the same.

Proposition 12 reveals another interesting finding. When F_m is not significant, client's valuation of the project (i.e., k) and the cost multiplier for client's effort (i.e., c_c) do not play a role in determining which contract is better, as they do not appear in the threshold expression in Proposition 12. Recall that the client's valuation of the project and her participation cost affect the equilibrium effort levels of both parties, and more importantly, the client's net value in the project. The reason why they do not affect which contract is better is that the percentage changes in client's net value in both of these contracts are the same as k or c_c varies. However, I find that if the fixed cost for monitoring vendor's effort (i.e., F_m) is significant, k and c_c indeed affect the contract choice. In such a case, if vendor's cost multiplier term c_v is more than the threshold presented in Proposition 12, then an increase in k favors the effort dependent contract while an increase in c_c favors the output dependent contract. On the other hand, if c_v is below that threshold, then an increase in k favors the output dependent contract, while an increase in c_c favors the effort dependent contract.

Let us now analyze how output sensitivity parameters α and β affect which contract is better. It is not possible to derive the thresholds with respect to α and β analytically, because the difference in client's value between the contracts is highly nonlinear with respect to them. Hence, I present two representative numerical examples in Figure 3.4. In these graphs, "EDC" and "ODC" represent the regions where the effort dependent contract and the output dependent contract, respectively, are preferred. The phrase "NO" denotes the region where no contract is feasible, and therefore there is no collaboration. In this example, the parameter values are: $T = 10$, $\gamma = 2.1$, $\delta = 1.7$, $k = 12$, $c_c = 2.0$, $c_v = 1.0$, $c_m = 0.3$, $R_c = 400$, and $R_v = 200$ with different values of output sensitivity to client's and vendor's efforts. In Figure 3.4(a), the fixed cost for monitoring vendor's effort (i.e., F_m) is zero. On the other hand, in Figure 3.4(b), $F_m = 14$. In other words, F_m is not negligible in Figure 3.4(b).

Figure 3.4(a) reveals that the output dependent contract dominates the effort dependent contract only when the output sensitivity to vendor's effort (i.e., β) is considerably higher than the output sensitivity to client's effort (i.e., α). However, if the output is equally sensitive to both parties' efforts or relatively more sensitive to client's effort, the effort dependent contract yields more value to the client. In addition, I observe from both figures that when the output sensitivity to the effort levels of both parties are very low, they do not collaborate (see the left-bottom area labeled with "NO"). In such a case, neither contract can generate more value than



(a) Without Fixed Cost of Monitoring (b) With Fixed Cost of Monitoring

Figure 3.4: Client's Contract Choice

the total reservation utilities of the parties.

As stated earlier, the difference between the settings depicted in Figures 3.4(a) and 3.4(b) is that the fixed cost for monitoring vendor's effort (i.e., F_m) is not negligible in Figure 3.4(b). In this figure, the set of parameter values where the output dependent contract dominates the effort dependent contract is not convex. This might not be intuitive at first, but can be explained as follows. For example, in Figure 3.4(b), when $\alpha = 0.20$ and β is lower than 0.15, the output dependent contract dominates. The reason is that, in such a case, the output does not depend highly on neither of the parties's effort levels, and the net value of the client is low. Hence, the client avoids paying the fixed payment term for monitoring and chooses to use the output dependent contract. On the other hand, when $\alpha = 0.20$ and β is between 0.15 and 0.45, the effort dependent contract dominates. This is because when the output depends on the effort levels of both parties, the severity of the double moral hazard problem increases as discussed in Section 3.4.2.1. However, when

$\alpha = 0.20$ and β is more than 0.45, again the output dependent contract dominates. In this case, the vendor exerts relatively more effort and the cost of monitoring is increased. In addition, because the client spends relatively less effort compared to the vendor, severity of the double moral hazard problem is decreased as explained in Section 3.4.2.1. I verified that these results hold for other problem instances as well. Hence, I summarize these findings in the following observation.

Observation 11

1. *When the fixed cost for monitoring vendor's effort (i.e., F_m) is negligible, the output dependent contract is better than the effort dependent contract only when the output sensitivity to vendor's effort (i.e., β) is considerably higher than the output sensitivity to client's effort (i.e., α). However, if the output is equally sensitive to both parties' efforts or relatively more sensitive to client's effort, the effort dependent contract yields more value to the client.*

2. *When the fixed cost for monitoring vendor's effort (i.e., F_m) is not negligible, the set of parameters where the output dependent contract dominates the effort dependent contract is non-convex and is similar to that in Figure 3.4(b).*

Now, in the next section, I present some interesting extensions of the base model and provide useful insights.

3.5 Extensions

In this section, I study two additional settings. First, I study pure revenue-sharing settings[4] where the parties share the output without the transfer of a fixed fee. After that, I allow the client or the vendor to get utility from the output even

[4]A through discussion of the pure revenue sharing is discussed in Section 2 as the output dependent contract. For the sake of completeness, I am providing the discussion and details here as well.

after the project is finished. In the next subsection, I start with analyzing the pure revenue-sharing contracts.

3.5.1 Pure Revenue-Sharing

Pure revenue-sharing contracts that have no fixed payment components are used in some collaboration environments (Gil and Lafontaine 2012). In these settings, the client and the vendor collaborate, and they share the output. Because the contract is written only in terms of the output, there is no need to monitor the effort level of either party in this setting. For simplicity, I consider that the reservation utilities of both parties are zero in this setting. However, the key findings and insights do not change in the presence of positive reservation utilities.

Recall that, in the output dependent contract, the client extracts all the value from the collaboration by adjusting the fixed payment term (see Section 3.2). As a result, the vendor's value in the output dependent contract is same as his reservation utility. However, there is no fixed payment term in the pure revenue-sharing case. Hence, in this case, the vendor's value is more than his reservation utility in the equilibrium. Nonetheless, the equilibrium effort levels in the pure revenue-sharing case for any payment parameter l are the same as those in the output dependent contract (see Lemma 13). After substituting these effort levels into the client's objective function, I observe that the client's value is concave in l. Hence, the first and second order conditions with respect to l reveal the optimal payment parameter that is presented in the following lemma.

Lemma 18 *The optimal transfer payment per unit output to the vendor in the pure revenue-sharing setting is given by:* $l^* = k\frac{\beta}{\delta}$.

Using the results in Lemmas 13 and 18, it is easy to derive the net values of both parties. I present these expressions in the next lemma.

Lemma 19 *In the pure revenue-sharing setting:*

1. *The client's net value from the project is:*
$$\frac{T^{\left(\frac{\beta\gamma+\alpha\delta-2\gamma\delta}{\beta\gamma+\alpha\delta-\gamma\delta}\right)\frac{(\gamma-\alpha)\delta-\beta\gamma}{(2\gamma-\alpha)\delta-\beta\gamma}}}{\left[\frac{k(\delta-\beta)}{\delta}\left(\left(\frac{k\alpha(\delta-\beta)}{\gamma\delta c_c}\right)^{\alpha\delta}\left(\frac{k\beta^2}{\delta^2 c_v}\right)^{\beta\gamma}\right)^{\frac{1}{\gamma\delta-\beta\gamma-\alpha\delta}}-c_c\left(\left(\frac{k\alpha(\delta-\beta)}{\gamma\delta c_c}\right)^{\delta-\beta}\left(\frac{k\beta^2}{\delta^2 c_v}\right)^{\beta}\right)^{\frac{\gamma}{\gamma\delta-\beta\gamma-\alpha\delta}}\right]^{-1}}.$$

2. *The vendor's net value from the project is:*
$$\frac{T^{\left(\frac{\beta\gamma+\alpha\delta-2\gamma\delta}{\beta\gamma+\alpha\delta-\gamma\delta}\right)\frac{(\gamma-\alpha)\delta-\beta\gamma}{(2\gamma-\alpha)\delta-\beta\gamma}}}{\left[k\beta\delta\left(\left(\frac{k\alpha(\delta-\beta)}{\gamma\delta c_c}\right)^{\alpha\delta}\left(\frac{k\beta^2}{\delta^2 c_v}\right)^{\beta\gamma}\right)^{\frac{1}{\gamma\delta-\beta\gamma-\alpha\delta}}-c_v\left(\left(\frac{k\alpha(\delta-\beta)}{\gamma\delta c_c}\right)^{\alpha}\left(\frac{k\beta^2}{\delta^2 c_v}\right)^{\gamma-\alpha}\right)^{\frac{\delta}{\gamma\delta-\beta\gamma-\alpha\delta}}\right]^{-1}}.$$

As discussed, the client cannot extract all the value in the pure revenue-sharing setting. Therefore, the client picks the payment term that maximizes only her value. On the other hand, the optimal payment term in Section 3.2 (i.e., $l^* = k\frac{\alpha\beta-\beta\gamma+\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}}{\alpha\delta-\beta\gamma}$) maximizes the total value generated in the system. Therefore, setting the payment term to $k\frac{\beta}{\delta}$ generates lower total value than that in the output dependent contract. It is easy to show that $k\frac{\beta}{\delta}$ is less than $k\frac{\alpha\beta-\beta\gamma+\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}}{\alpha\delta-\beta\gamma}$ when β is not small. Hence, I obtain the following result.

Proposition 13 *When the output sensitivity to vendor's effort (i.e., β) is not low, the payment term offered in the pure revenue-sharing contract is less than the required amount to maximize the total value in the collaboration.*

Although the client's value and the total value generated in the pure revenue-sharing contract is different from those in the output dependent contract, most of the findings in the output dependent contract holds in the pure revenue-sharing setting. For example, Proposition 7 (with different threshold values), Proposition 11, and the behavior depicted in Figure 3.2 are all valid in the pure-revenue sharing case. Finally, in the next section, I analyze the effects of salvage value on the dynamics of the collaboration.

3.5.2 Salvage Value

Here I consider that the client or the vendor gets utility from the output both during the collaboration period and after the collaboration. For example, the client can use the output even after the collaboration ends. In this scenario, the value obtained after the collaboration period can be considered as a salvage value. On the other hand, the vendor can build reputation or history from his business collaboration. His success, reflected in the output, might be considered by other potential clients as a signal of how successful the vendor would be in his future businesses. This can be considered as a salvage value of the project to the vendor.

For brevity, I present the results only for the effort dependent contract in this section. Hence, given the salvage value parameters S_c and S_v for the client and the vendor, the objective functions of the parties and the constraints are presented below.

$$\max_{u(t),H(.)} \left\{ \int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - H(.) + S_c q(T) \right\},$$
$$\max_{v(t)} \left\{ H(.) - \int_0^T (c_v + c_m) v(t)^\delta dt - F_m + S_v q(T) \right\}.$$

subject to

$$\dot{q}(t) = u(t)^\alpha v(t)^\beta;$$
$$\int_0^T kq(t)dt - \int_0^T c_c u(t)^\gamma dt - H(.) + S_c q(T) \geq R_c;$$
$$H(.) - \int_0^T (c_v + c_m) v(t)^\delta dt - F_m + S_v q(T) \geq R_v;$$
$$u(t) \geq 0; \quad v(t) \geq 0.$$

The contract that coordinates this setting is very similar to the contract presented in Proposition 8. Hence, I omit the details. However, I present the equilibrium effort levels in the following lemma.

Lemma 20 *In presence of the salvage values for both client and vendor, their equilibrium effort levels in the effort dependent contract are:*

$$u(t) = \left(\left(\frac{(S_c+S_v+k(T-t))\alpha}{\gamma c_c} \right)^{\delta} \left(\frac{\beta\gamma c_c}{\alpha\delta(c_v+c_m)} \right)^{\beta} \right)^{\frac{1}{\gamma(\delta-\beta)-\alpha\delta}},$$
$$v(t) = \left(\left(\frac{(S_c+S_v+k(T-t))\beta}{\delta(c_v+c_m)} \right)^{\gamma} \left(\frac{\alpha\delta(c_v+c_m)}{\beta\gamma c_c} \right)^{\alpha} \right)^{\frac{1}{\gamma(\delta-\beta)-\alpha\delta}}.$$

This lemma shows that the optimal effort levels decrease in time similar to that in the no salvage value scenario. However, the terminal effort levels here are positive due to the salvage value. Next, similar to the earlier case, the client covers only the effort cost of the vendor and the costs related to monitoring in addition to the vendor's reservation utility. Let us denote the total of the salvage value terms with S, i.e, $S = S_c + S_v$. As expected, if I set S to zero, the equilibrium effort levels converge to the case with no salvage value. In addition, I would like to note that if I let $\frac{S}{k} \to \infty$ or set $k = 0$ (that implies that the client does not get utility throughout the collaboration period but only at the project completion time), our model converges to a static model with no sense of continuous analysis except the accumulation of output that is valuable only at the end.

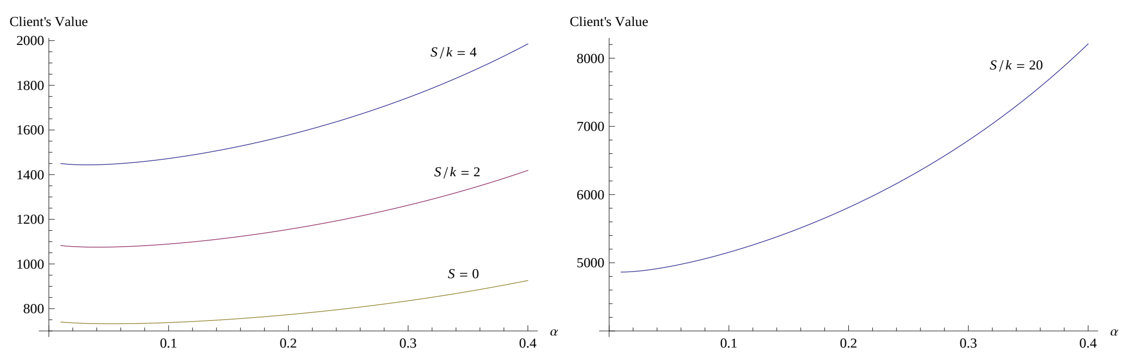
Most results remain valid in this setting compared to those in the effort dependent contract studied in Section 3.3. However, the relationship between value and output sensitivity terms discussed in Proposition 11 and Figure 3.1 changes its behavior when the parties have salvage value in the project. The analytical conditions are cumbersome to present, however, the following behavior is observed in the entire set of problem instances I analyzed.

Observation 12

1. *When the ratio of the total salvage value to the ongoing value during the collaboration (i.e., $\frac{S}{k}$) is low, an increase in the output sensitivity to either party's effort (i.e., α or β) may decrease the client's net value.*

2. *When the ratio of the total salvage value to the ongoing value during the collaboration (i.e., $\frac{S}{k}$) is high, an increase in the output sensitivity to either party's effort (i.e., α or β) always increases the client's net value.*

Observation 12 is illustrated in Figure 3.5 using $T = 10$, $\beta = 0.4$, $\gamma = 3$, $\delta = 2.5$, $k = 12$, $c_c = 2$, $c_v = 1.2$, and $c_m = 0.3$. As shown in Figure 3.5(a), if I increase the importance of the terminal value in the collaboration, the behavior of decreasing client's value with output sensitivity fades away. Figure 3.5(b) considers the same problem instance with $\frac{S}{k} = 20$ where I observe that this behavior is completely gone. Hence, as stated in the observation, if $\frac{S}{k}$ is large enough, the client's net value always increases as the output becomes more sensitive to the effort of either party.



(a) Low Terminal Value (b) High Terminal Value

Figure 3.5: Value to the Client with Respect to the Output Sensitivity to Client's Effort

3.6 Conclusions

In the business settings I analyze, such as consulting, IT services, and auditing, the traditional view of analyzing supplier-client or consultant-client relationship needs to be re-visited in light of the fact that the output of these relationships depend on the effort levels of both parties, not just the vendor's efforts. Therefore, I analyze a value co-creation environment where the output necessitates the efforts of both parties.

In this section, I also consider the fact that the effort levels of the parties may not be monitored. This leads to a double moral hazard problem, however, the client may choose to observe the effort level of the vendor with a cost. In addition, both parties may change their effort levels dynamically. Moreover, I allow the client to get utility from the output as it is developed. Considering all these facts, I examine two different settings from a differential game perspective. I derive the equilibrium effort levels of the client and the vendor as well as the optimal payment parameters or contracts. Next, I present several interesting findings based on the sensitivity analyses of the equilibrium effort levels with respect to the parameters representing different characteristics of the parties. The client's valuation of the project and the progress in the collaboration play an important role in the behavior of the effort levels. Depending on these factors, the equilibrium effort levels might increase or decrease with the changes in the output sensitivities to effort levels and the cost elasticities of efforts.

I also compare the performances of different contracts in order to find the best one for the client under different circumstances. This analysis also reveals whether the vendor's effort should be monitored or not. I find that, as long as the participation cost of the vendor is not very low, his effort should be monitored and the client should use an effort dependent contract. Otherwise, the client should not monitor vendor's effort, and should operate under double moral hazard with an output dependent contract. Another interesting finding is that, if the sensitivity of output to vendor's effort is relatively higher than the sensitivity to client's effort, the client should use the output dependent contract and should not monitor vendor's effort.

I also find that, under certain circumstances, an increase in the sensitivity of the output to either parties' effort level does not necessarily increase the net values of the parties. However, if the relative value received at the end of the collaboration is high (compared to the on-going utility received), any increase in the output sensitivity to either party's effort level always increases the net value. Finally, I consider the pure revenue-sharing contracts and find that the total value generated in such an environment is lower than that in the output dependent contract.

4. IMPACT OF INVENTORY ON RECOMMENDATION SYSTEMS FOR DVD RENTAL FIRMS

4.1 Problem Description

Most of the major DVD-rental firms, such as Netflix and Blockbuster, offer subscription based plans to their customers. The operations of Netflix, for example, are divided into two main categories: online streaming and DVD rentals by mail. These two services are complementary to each other rather than separate, because not all movie titles are available for streaming. As of 2012, Netflix offered around 60,000 titles for streaming versus more than 140,000 titles in DVD format (Liedtke 2012). The customers are in need of DVDs for completeness as well as simplicity, and also if they do not have high speed broadband (Netflix 2012). Besides, the contribution of domestic DVD service to the total profit of Netflix in the second quarter of 2012 was \$134 million, whereas the domestic online streaming contributed only \$83 million. On the other hand, international streaming resulted in a loss of \$89 million (Netflix 2012). Netflix expects to achieve profitability in the international markets over the long-term. These numbers suggest that the DVD service keeps the company afloat. Accordingly, Netflix CEO Hastings noted that their goal is to keep the DVD service as healthy as possible for many years to come (Grover and Edwards 2011). Hence, in this section, I restrict the analysis to DVD rentals by mail.

Customers arrive at the websites of DVD-rental firms and request movies that are sent by mail based on their availability. Although they differ in the sophistication level, most of these firms utilize recommender systems. Figure 4.1 shows a snapshot of the recommendation pages presented to two different customers of Netflix at the same time. This figure illustrates that the number of recommendations and the

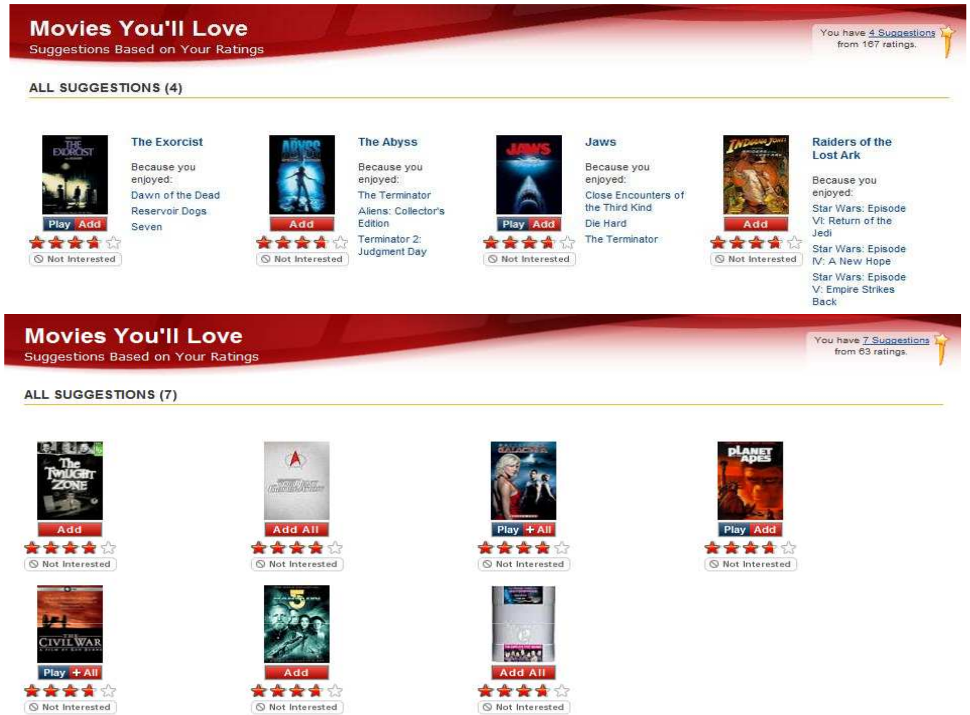


Figure 4.1: Recommendations to Two Different Customers by Netflix

list of recommended movies differ across customers. Note that, in the short run, recommendations just shift demand from one product to another in subscription based settings, rather than creating new demands as in the case of e-commerce websites (such as Amazon.com).

In the context of DVD-rental firms, the satisfaction of a customer depends on the availability of the movie she requests (Hastings et al. 2008, Shih et al. 2009). Therefore, these firms need to satisfy the demands of as many customers as possible. Given the number of customer arrivals in a given time period, the problem of maximizing the expected customer satisfaction is equivalent to minimizing the portion of the expected demand that cannot be satisfied due to the unavailability of movies. Hence, the objective is to find the best combination of recommendations to influence customer demand such that the highly available movies are shipped more and the unsatisfied demand is minimized. Another factor that influence customer satisfaction is whether they are recommended the movies they value higher. I control this aspect of satisfaction by allowing the model to pick only those movies that are expected to be liked by the customer.

An important input for the model is the rating estimates of movies by the users. Since several past studies have presented effective algorithms for estimating ratings, I consider that these rating values are exogenous to the model. However, I examine and analyze the impact of the prediction quality of recommender system on the solution. Besides, since the DVD-rental firms may have millions of users and thousands of movie titles, sometimes clustering has been utilized for both movies (Milkman et al. 2009, Jedidi et al. 1998) and users (Bassamboo et al. 2009, Bassamboo and Randhawa 2007) in order to make operational decisions. In the similar fashion, I also define the model at the cluster level. Note that such modeling is very general in the sense that the firms can always use the movies and/or customers at the individual level (e.g., by setting the number of movie clusters as the total number of movies) when the computational resources are sufficient. This is an important issue to clarify because computational resources are becoming less costly and more readily available with the advent of newer technologies, such as cloud computing (Thibodeau 2012). For example, Netflix wants to move as much as 95% of their IT services to the cloud (Velte et al. 2009).

If the computational resources are not sufficient, then one needs to utilize clustering. The clustering may result in the loss of individual level information, but it helps us in utilizing inventory information in a more effective manner. I show later that clustering in the setup does not deteriorate the quality of the solution significantly. Moreover, the solution obtained using clustering can be easily tailored to implement at the individual level (I discuss it in detail in Section 4.1.4). I also show that the solution obtained by utilizing inventory information (using customer clusters) outperforms the one obtained without utilizing inventory information (but using the individual customer information). For a detailed review on clustering, the readers can refer to Jain et al. (1999). Next, I discuss the model in detail.

4.1.1 Problem Formulation

I begin with presenting the stochastic version of the problem where demand and inventory are random variables. However, as mentioned in the introduction section, my analyses are based on a deterministic approximation. Hence, I present the deterministic model and a detailed justification of this approximation in the next subsection. Since the firms use a threshold level to maintain trust, I consider only those movies as candidates for recommendation for which the estimated rating is above the threshold. The difference between the proposed policy and the all-inclusive policy is that the latter recommends all movies above the threshold whereas the former chooses an optimal subset among those movies. Clearly, the number of recommendations in such policies is not constant, as shown in Figure 4.1.

Let n_{it} be the number of customers arriving from user cluster i in time period t. Several empirical studies have shown that n_{it} can be estimated effectively using past data, and therefore it can be treated as deterministic (Chung 2010, Lehmann and Weinberg 2000). Next, let us denote the set of user clusters by U, the set of movie clusters by J, and the time horizon by T. In the model, a customer requests exactly one movie in each arrival. Without any loss of generality, I can treat multiple requests as multiple arrivals. The decision variable x_{ijt} denotes the proportion of customers in cluster i that are recommended movie cluster j in period t. For the implementation purpose, x_{ijt} can also be interpreted as the proportion of movies in cluster j that are recommended to customers in cluster i in period t. The stochastic model is now presented below. I summarize the parameters and variables in Table 4.1.

$$\min \mathbb{E}\left[\sum_{t=1}^{T}\sum_{j\in J}\left(\tilde{d}_{jt} - \tilde{K}_{jt}\right)^{+}\right] \tag{4.1}$$

subject to

$$\tilde{d}_{jt} = \sum_{i\in U} \tilde{e}_{ijt};\ i \in U,\ j \in J,\ t = 1, 2, \ldots, T \tag{4.2}$$

$$\tilde{e}_{ijt} \sim B(n_{it}, A_{ijt});\ i \in U,\ j \in J,\ t = 1, 2, \ldots, T \tag{4.3}$$

$$\tilde{K}_{jt} = \tilde{K}_{j(t-1)} - \tilde{d}_{j(t-1)} + \left(\tilde{d}_{j(t-1)} - \tilde{K}_{j(t-1)}\right)^{+} + \tilde{W}_{j(t-1)}; j \in J; \forall t \tag{4.4}$$

$$\tilde{W}_{j(t-1)} = \sum_{s=1}^{t-1}\left[\tilde{d}_{js} - \left(\tilde{d}_{js} - \tilde{K}_{js}\right)^{+}\right] Q_{s(t-1)};\ j \in J,\ , t = 1, 2, \ldots, T \tag{4.5}$$

$$R'_{ijt} = R_{ij} + (R_{max} - R_{ij})\, x_{ijt} E;\ i \in U,\ j \in J,\ t = 1, 2, \ldots, T \tag{4.6}$$

$$A_{ijt} = \frac{R'_{ijt}}{\sum_{s\in J} R'_{ist}};\ i \in U,\ j \in J,\ t = 1, 2, \ldots, T \tag{4.7}$$

$$0 \leq x_{ijt} \leq 1;\ i \in U,\ j \in J,\ t = 1, 2, \ldots, T \tag{4.8}$$

$$x_{ijt} = 0;\ (i, j) \notin V,\ i \in U,\ j \in J,\ t = 1, 2, \ldots, T \tag{4.9}$$

The objective is to minimize the expected value of the total dissatisfaction of customers. As discussed before, I can state it as the total number of movie requests that are not fulfilled throughout the planning horizon. In a given time period t, demand for a specific movie cluster j is stochastic and denoted by $\tilde{d}_{jt}$. The available inventory of movie cluster j in period t is another stochastic variable $\tilde{K}_{jt}$ except when $t = 1$, i.e., the starting inventory levels are known. I would like to note that the DVD-rental firms usually can not replenish their inventories in real-time. The main reason is that, in the DVD rental context, most DVDs are acquired with revenue-sharing contracts, and these contracts do not allow the DVD rental firms to purchase new DVDs. More specifically, DVD purchase quantities for a movie are determined one to four weeks before the movie is available for renting to consumers (Chung 2010). Therefore, I do not allow the model to replenish inventories. However, the model can be easily modified to consider a scenario where the inventory can be replenished. Overall, the dissatisfaction for movie cluster j in period t can be written as $\left(\tilde{d}_{jt} - \tilde{K}_{jt}\right)^{+}$. Consequently, the expectation of the total dissatisfaction throughout the planning horizon for all movies can be written as in Equation (4.1).

Constraint set (4.2) decomposes demand $\tilde{d}_{jt}$ based on the user cluster it stems from. Here, $\tilde{e}_{ijt}$ is the demand for movie cluster j that is requested by users in cluster i in period t. I can state that $\tilde{e}_{ijt} \sim B(n_{it}, A_{ijt})$, i.e., $\tilde{e}_{ijt}$ is binomially distributed with mean $A_{ijt}n_{it}$ and variance $A_{ijt}(1 - A_{ijt})n_{it}$ as constraint set (4.3) presents. Here, A_{ijt} is the probability that a user in user cluster i requests a movie in movie cluster j in period t. Next, the constraint set (4.4) denotes the evolution of available inventory $\tilde{K}_{jt}$ over time. The first term on the right hand side is the random variable that denotes the available inventory of movie cluster j in the previous time period. The second term and the terms in the parentheses together denote the negative of the amount of movies shipped in the previous period. More specifically, the second term is the total demand in the previous period, and the parentheses denotes the demand that could not be satisfied. The last term $\tilde{W}_{j(t-1)}$ denotes the number of returns of movies from cluster j in period $(t-1)$ that are available to ship in period t. Now the constraint set (4.5) states that the returns are a function of the number of shipped DVDs in the earlier periods. Here, Q_{st} indicates the proportion of all the movies rented in period s that are returned in period t. Based on the rental and return data, it is easy to estimate Q_{st}, e.g., Chung (2010) estimates Q_{st} using data collected from Blockbuster. Moreover, Chung (2010) shows that it is reasonable to consider Q_{st} as deterministic.

Symbol	Definition	Remarks
Parameters		
U	Set of customer clusters	Index i: $i \in U$
J	Set of movie clusters	Index j: $j \in J$
T	Planning horizon	Index $t : t = 1, 2, \ldots, T$
θ	Threshold level for recommendation	Higher value implies that less movies are available for recommendation
E	Effectiveness of the recommender system	
n_{it}	Number of arrivals	
Q_{st}	Proportion of all the movies rented in period s that are returned in period t	$1 \leq s \leq t$
R_{ij}	Estimated rating for the movies	
R_{max}	Maximum possible rating	Usually 5 or 10
V	Set of pairs for which estimated ratings are above threshold	$(i, j) \in V \ \ \forall R_{ij} > \theta$
Variables		
R'_{ijt}	Perceived utility of movies after recommendations are shown	
A_{ijt}	Probability of requesting a movie	
$\tilde{K}_{jt}$	Inventory of movies available for shipping	Random Variable
$\tilde{e}_{ijt}$	Demand of movie cluster j in period t from user cluster i	Random Variable
$\tilde{d}_{jt}$	Total demand of movie cluster j in period t	Random Variable
$\tilde{W}_{jt}$	Number of returns	Random Variable
Z_{jt}	Expected demand that cannot be satisfied	
x_{ijt}	Proportion of customers in cluster i that are recommended movie cluster j in period t	**Decision Variable**

Table 4.1: List of Parameters and Variables in Section 4

Constraint set (4.6) is central to the discussion that the recommendations affect the rental decisions of customers. This argument is similar to that in Fleder and Hosanagar (2009) where the perception about a movie increases due to recommendations, as if the recommendations were advertisements. This constraint set states that if the movie cluster j is not recommended to a user in cluster i in period t (i.e., if $x_{ijt} = 0$), then there is no increase in the perception of the utility of the movie for that customer (i.e., $R'_{ijt} = R_{ij}$). Otherwise, if the recommendation is made (i.e., $x_{ijt} > 0$), the perception increases by the amount $(R_{max} - R_{ij})\, x_{ijt}E$. The term in the parenthesis is the difference between maximum possible rating and actual rating. Parameter E denotes the percentage increase in the difference term because of the recommendation. Since the firms recommend only those movies that have ratings above a threshold, the change in customer perception level is always positive. Therefore, I consider $E > 0$. It is easy for the firms to estimate the value of E by observing the increase in perceived utility after recommendations are shown. Although trust is affected by the recommendations conformity with the taste of the users, its dynamics should be observed in longer time periods during which the customers adapt the use of recommendation system based on their past experiences (Kim et al. 2009). Since I solve the problem for short time horizons, I do not consider E as a variable. Nevertheless, I present some interesting insights regarding the effect of changing the threshold level on customer satisfaction in short-run.

For each customer cluster and each time period, the constraint set (4.7) scales the perceived values of the movies and assigns them to A_{ijt}, which is the probability that a specific customer in cluster i will request a movie from cluster j in time period t. Thus, A_{ijt} is a function of both the raw rating estimates and all of the recommendations shown to the user. This is actually a variant of the constant-utility attraction model (Tang 2010), which stems from Luce's theory of strict utility (Luce 1959). Next, the constraint set (4.8) defines the bounds for the parameter values. As explained earlier, the all-inclusive policy, i.e., the prevalent industry practice, is to recommend a movie cluster j to a customer from cluster i only if the corresponding rating estimate is above a threshold level. This is what the constraint set (4.9) depicts by defining V such that $(i, j) \in V \;\; \forall R_{ij} > \theta$. One can easily set $\theta = 1$ to contain all of the movies in the set of recommendable movies. However, in order to ensure that the model does not recommend movies to users that they will not like, I incorporate the constraint set (4.9) into the formulation.

4.1.2 Justification of a Deterministic Approximation

Since the stochastic model is analytically intractable, I need to simplify it in order to derive useful managerial insights. The analysis of the distribution of the demand for movie cluster j in period t, i.e., $\tilde{d}_{jt}$, reveals that it can be approximated reasonably well and treated as a deterministic variable. Since $\tilde{d}_{jt} = \sum_{i \in U} \tilde{e}_{ijt}$ (see Equation (4.2)), I begin with the analysis of $\tilde{e}_{ijt}$. As discussed earlier, $\tilde{e}_{ijt}$ is binomially distributed with mean $A_{ijt}n_{it}$ and variance $A_{ijt}(1 - A_{ijt})n_{it}$. However, based on the de Moivre-Laplace theorem, which is essentially a special case of the central limit theorem, one can approximate the binomial distribution with the normal distribution with mean $A_{ijt}n_{it}$ and variance $A_{ijt}(1 - A_{ijt})n_{it}$ (Box et al. 1978). This is a good approximation when the following condition is satisfied: $\left|(1/\sqrt{n_{it}})\left(\sqrt{(1-A_{ijt})/A_{ijt}} - \sqrt{A_{ijt}/(1-A_{ijt})}\right)\right| < 0.3$. Netflix ships around 2 million DVDs per day (Bond 2011). If the number of user clusters is taken as 10, then the average value of n_{it} is $200,000$. On the other hand, A_{ijt} would be in the range of 0.02 to 0.20 depending on the number of movie clusters that are around 8 to 15. For $n_{it} = 200,000$ and $A_{ijt} = 0.05$, the left hand side of above condition is 0.0051. Clearly this is much less than 0.3, and therefore the normal distribution is a reasonably good approximation.

Since the sum of normal random variables is also normally distributed, the demand for the movie cluster j in period t, i.e, $\tilde{d}_{jt} = \sum_{i \in U} \tilde{e}_{ijt}$, is a normal random variable with mean $\sum_{i \in U} A_{ijt}n_{it}$ and variance $\sum_{i \in U} A_{ijt}(1 - A_{ijt})n_{it}$ (Ross 2009). Therefore, 99.7% of the values lie around 3 standard deviations of the mean.

It is easy to see in the problem that the variance of $\tilde{d}_{jt}$ is quite small compared to its mean value. This implies that most of the values of demand for a specific movie cluster are grouped around the mean. Consequently, for the analysis, one can approximate $\tilde{d}_{jt}$ by its mean value $\sum_{i\in U} A_{ijt}n_{it}$. For example, when the numbers of user cluster and movie cluster are 10 and 8 respectively, with the average values of n_{it} and A_{ijt} (i.e., $n_{it} = 200,000$ and $A_{ijt} = 0.125$), the standard deviation is $\sqrt{\sum_{i\in U} A_{ijt}(1-A_{ijt})n_{it}} = 467.70$ and mean is $\sum_{i\in U} A_{ijt}n_{it} = 250,000$. Hence, the coefficient of variation is only 0.187%. This implies that the demand would be around ± 0.561% of the mean with probability 0.9973. More specifically, $\Pr\left[248,597 < \tilde{d}_{jt} < 251,403\right] \approx 0.9973$.

Thus, in the analysis henceforth, I consider the demand for movie cluster j in period t, i.e., $\tilde{d}_{jt}$, as deterministic and denote it by d_{jt}. Treating demand as deterministic also results in deterministic returns and inventory. This follows by substituting $\tilde{W}_{j(t-1)}$ (from Equation 4.5) and the starting inventory level K_{j1} into Equation (4.4) and by rewriting Equation (4.4) recursively for each $\tilde{K}_{jt}$. Therefore, now I denote the available inventory and the number of returns by K_{jt} and W_{jt}, respectively. Although I approximate the stochastic model as a deterministic variant, I later analyze the impacts of errors in return probability and inventory levels. Now, I introduce the deterministic variant of the problem below.

$$\min_{x_{ijt}} \sum_{t=1}^{T} \sum_{j\in J} Z_{jt} \tag{4.10}$$

subject to

$$Z_{jt} \geq \sum_{i\in U} A_{ijt}n_{it} - K_{jt};\ j \in J,\ t = 1, 2, \ldots, T \tag{4.11}$$

$$K_{jt} = K_{j(t-1)} - \sum_{i\in U} A_{ij(t-1)}n_{i(t-1)} + Z_{j(t-1)} + W_{j(t-1)};\ j \in J,\ \forall t \tag{4.12}$$

$$
\begin{aligned}
W_{j(t-1)} &= \sum_{s=1}^{t-1}\left[\sum_{i\in U} A_{ijs}n_{is} - Z_{js}\right] Q_{s(t-1)};\ j\in J,\ t=2,3,\ldots,T && (4.13)\\
R'_{ijt} &= R_{ij} + (R_{max} - R_{ij})\, x_{ijt}E;\ i\in U,\ j\in J,\ t=1,2,\ldots,T && (4.14)\\
A_{ijt} &= \frac{R'_{ijt}}{\sum_{s\in J} R'_{ist}};\ i\in U,\ j\in J,\ t=1,2,\ldots,T && (4.15)\\
0 &\le x_{ijt}\le 1,\quad Z_{jt}\ge 0;\ i\in U,\ j\in J,\ t=1,2,\ldots,T && (4.16)\\
x_{ijt} &= 0;\ (i,j)\notin V,\ i\in U,\ j\in J,\ t=1,2,\ldots,T && (4.17)
\end{aligned}
$$

In the deterministic model, I introduce the variable Z_{jt} that denotes the unsatisfied demand for each movie cluster j in period t. Hence, $Z_{jt} = (d_{jt} - K_{jt})^{+}$. One can further write $Z_{jt} \ge \sum_{i\in U} A_{ijt}n_{it} - K_{jt}$, because $d_{jt} = \sum_{i\in U} A_{ijt}n_{it}$. Since Z_{jt} explicitly appears in the objective function as a minimization term, the inequality $Z_{jt} \ge \sum_{i\in U} A_{ijt}n_{it} - K_{jt}$ becomes an equality when the available inventory is less than demand. On the other hand, if the available inventory is more than demand, then Z_{jt} would become zero because it is defined as a nonnegative term. This is reflected in the constraint set (4.11). Other constraints are the deterministic analogs of the constraint sets in the stochastic version. Clearly, the presented model is nonlinear because of the constraint set (4.15). However, in order to use a linear solver, I can easily linearize it using standard linearization techniques. The linearized version is presented in Appendix C.

4.1.3 Comparison of the Stochastic Version and the Deterministic Approximation

In order to strengthen the validity of the deterministic approach, I conduct a simulation study to examine the effect of using mean values for the random variables rather than treating them as stochastic. The analysis presented earlier suggests that the percentage coefficient of variation of demand for a movie cluster in a given time period is around 0.187%. Here, I examine how this measure is affecting the quality of

Problem Class	Problem Size			Objective Value Comparison	
	$\|U\|$	$\|J\|$	T	*MAPD*	*PCV*
1	15	15	12	0.064%	0.07%
2	10	15	12	0.057%	0.07%
3	15	8	12	0.052%	0.07%
4	10	8	12	0.002%	0.00%
5	15	15	7	0.032%	0.04%
6	10	15	7	0.042%	0.05%
7	15	8	7	0.034%	0.04%
8	10	8	7	0.003%	0.00%

Table 4.2: Deterministic vs Stochastic

the solution in the deterministic approximation. To this end, I compare the objective function value of the deterministic problem (DO) with that of a simulation study (SO) where the demand process is treated as a random event. The proximity of these values supports the validity of the deterministic demand assumption. Table 4.2 summarizes the findings.

I carried out simulation for 8 different scenarios in a full factorial design where I set $|U|$ to 10 and 15; $|J|$ to 8 and 15; and T to 7 and 12. For each problem class, I run the simulation 100 times. In the table, I report the mean absolute percentage difference (MAPD) between the objective values, i.e., $\left|\frac{SO-DO}{DO}\right| \times 100$. The average value of MAPD across all scenarios is less than 0.04%. Moreover, the percentage coefficient of variation (PCV), i.e., $\frac{\text{Standard Deviation}}{\text{Mean}} \times 100$, is also reported in Table 4.2. Clearly, PCV is also very small indicating that the deterministic model is a good approximation.

4.1.4 Implementation at the Individual Customer Level

As discussed earlier, the decision variable x_{ijt} can be interpreted in the following two ways: (i) the proportion of users in cluster i to whom each movie in cluster j

should be recommended in period t, and (ii) the proportion of movies in cluster j that should be recommended to each user in cluster i in period t. The DVD-rental firms can implement the results of the model using either of these interpretations (depending on the strategies of the firm). For example, if $x_{ijt} = 0.60$ for a given i, j, and t, and the firm adopts the second interpretation, then 60% of the movies in cluster j should be recommended to each user in cluster i in period t. Moreover, within cluster j, the firm can strategically choose those 60% of the movies for recommendation that have higher availability. Further, this selection of movies may also take into account the customers' individual rating values in order not to recommend them movies below the rating threshold. Finally, the firm should not recommend those movies to a customer that have already been watched by her earlier. Using all these implementation strategies, the firm can use the cluster level results of the model in an effective manner.

The individual constraints discussed above do not alter the overall effects of the clustered recommendations, because the effects of random terms cancel out when the overall effect of individual recommendations is calculated. This can be proved in the similar manner as the justification of deterministic approximation in Section 4.1.2. Further, I would also like to re-emphasize that the firms can always use the movies and/or customers at the individual level (e.g., by setting the number of movie clusters as the total number of movies) when the computational resources are sufficient.

4.2 Solution Approaches

The solution of the deterministic problem defined in the previous section provides managers with recommendation decisions throughout the planning horizon. However, in practice, once the solution is implemented for a period, the managers can re-solve the problem in the next period with the consideration of realized demand

and returns. Hence, depending on whether the problem is solved in each period or only once at the start of the planning horizon, there are two solution approaches. I call them the dynamic and the static approaches, respectively.

4.2.1 Dynamic Approach

I begin with presenting the dynamic solution approach where the arrival and return events are observed after each time period and the available inventory is updated accordingly. The decision variable, as discussed earlier, is the proportion of users in each cluster to whom each movie cluster should be recommended in each period. In the dynamic approach, the problem is solved for the remaining planning horizon in any given period but the solution is implemented only for that period. The reason is that the firm re-solves the problem in each period with the updated inventory information. I find that the realistic instances of this problem cannot be solved to optimality in a reasonable amount of time using the state-of-the-art integer programming solvers. The problem becomes harder to solve as the problem size increases in user clusters, movie clusters, or the time horizon. Therefore, I develop an efficient and effective heuristic called the dynamic heuristic.

The dynamic heuristic is similar to the rolling horizon problem that is used for solving stochastic dynamic programs (SDP). Due to the curse of dimensionality, it is difficult to solve many SDPs to optimality in a reasonable amount of time. The rolling horizon approach is used as a proxy under such circumstances (Alden and Smith 1992). Similar to the rolling horizon problems, in the dynamic heuristic, a horizon length of H periods is chosen such that it is possible to solve the problem in a reasonable amount of time. Then, in every period t, the deterministic problem is solved where the decisions are the recommendations for periods $t, t+1, \ldots, t+H$ and the objective is to minimize the expected customer dissatisfaction over the same horizon. In this step, only the recommendations for period t are implemented. Then, the horizon rolls, i.e., the demand and return realization in period t is observed, the inventory position is updated for period $t+1$, a new problem corresponding to the new planning horizon $[t+1, t+H+1]$ is solved, and the recommendations are utilized for period $t+1$ only. This procedure is repeated until the end of the planning horizon is reached. The dynamic heuristic is practicable, easy to implement, runs quickly, and provides very good solutions for a wide variety of realistic problem instances. Moreover, this heuristic can be easily extended for the static approach that I discuss in the next subsection.

4.2.2 Static Approach

In the static approach, the firm solves the problem based on the expected values of parameters once at the beginning of the planning horizon, and utilize this solution for all periods. Since the problem is solved only once at the beginning of the planning horizon, inventory is not updated to their realized values in this model. This is the fundamental difference between the static and dynamic approaches. However, the static approach is useful for various reasons: (i) The solution of the static approach presents the expected value of the dissatisfaction over the planning horizon, and therefore it is useful for performance comparison purposes. (ii) As will be discussed in the next subsection, the dynamic approach outperforms the static solution only under special circumstances. (iii) The static approach is more amenable to analytical analyses as compared to the dynamic approach, and therefore it helps us in deriving useful managerial insights. (iv) Not all DVD-rental firms have capability to acquire timely information for each step of the dynamic approach. (v) If the costs related to obtaining/updating real time inventory information and solving the problem repetitively in each period are high, then the firm may adopt the static approach.

Similar to that in the dynamic approach, it is not possible to solve problems of realistic sizes to optimality in the static approach. Therefore, I modify the dynamic heuristic in order to solve the static approach of the problem and call it the static heuristic. Its difference from the dynamic heuristic is that the rentals and returns are not observed, but expectations of these events are used to update the parameters in each period. The dynamic approach is more informative than the static in the sense that the realized data is used to update inventory in each period. However, as I discuss next, the dynamic approach is not always better than its static counterpart.

4.2.3 Comparison between Dynamic and Static Approaches

I compare the performances of dynamic and static approaches in an experimental setting. In this experiment, I create two types of variations in the probability values. First, in order to introduce randomness, I draw individual request probabilities from their respective distributions around their mean values. Such randomness captures the fact that I consider clusters in the problem formulation and that the individual values may be different from the cluster means. Second, I incorporate bias in the estimation of the request probabilities by changing the mean value itself. Specifically, in the experiment, the errors in mean values are varied from -10% to +10%. In other words, I consider that the demand for a movie cluster might be underestimated or overestimated by the firm during optimization.

Figure 4.2 depicts the percentage gain that can be achieved by switching from the static approach to the dynamic approach with varying amount of error in the estimation of request probabilities. I present the results for two problem instances. Problem instance 1 represents a critically balanced system where the expected demand for some movie clusters are close to their available inventories. In such a case, the information about the realization of demand is important to the rental firm.

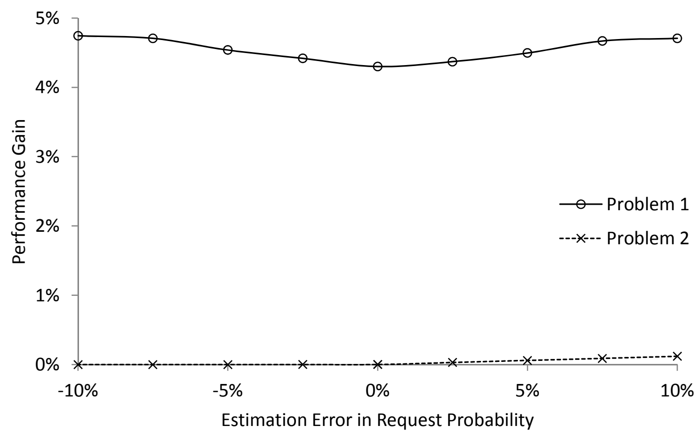


Figure 4.2: Percentage Gain by Using Dynamic Heuristic

Therefore, dynamic heuristic is better in shifting demand towards more available titles in this problem instance. As a result, the total dissatisfaction for the dynamic approach is comparatively lower. Moreover, the performance gain slightly increases with an increase in the error in request probabilities (for both underestimation and overestimation).

On the other hand, problem instance 2 depicts a case where the demand and requests for the movie clusters are not too close. In this case, the difference between static and dynamic solutions is very small, i.e., at most 0.1%, and almost insensitive to the estimation error. I ran the experiment for other problem instances, and the results are consistent with those presented in Figure 4.2. Hence, I summarize this finding in the following observation.

Observation 13 *When the demand and inventory are close to each other for some movies, it is better for the firm to adopt dynamically as long as the benefits outweigh the costs involved in implementing the dynamic solution. However, when no movie cluster is expected to have demand and requests to be close, the firm may consider using the static approach. It would help them in saving the costs related to collecting* *and updating parameter values in real-time and solving the problem in each period.*

4.2.4 Performance of the Static Heuristic

In order to examine the performance of the static heuristic, I compare its solution with the optimal solution for a variety of problem instances in realistic parameter ranges. I design a full factorial experiment where I set $|U|$ to 8, 10, and 12; $|J|$ to 10 and 12; and T to 7 and 10. For each of the twelve problem classes, there are twelve problem instances in which I set E to three different values (0.15, 0.20, and 0.25), vary the total demand randomly around its mean by 20%, and set Q_{st} parameters to represent two different return patterns.

The results of the experiment are presented in Table 4.3 where each row represents a problem class. In each problem instance, I calculate the reported comparison value as the difference between the heuristic value and the optimal value divided by the optimal value. Whenever the optimal solution could not be obtained, I use the best lower bound on the optimal solution. Hence, the performance of the heuristic is

Problem Class	**Problem Size**			**Objective Value Comparison**			
	$\|U\|$	$\|J\|$	T	*Min*	*Max*	*Median*	***Average***
1	8	10	7	0%	2.8%	0.3%	**0.6%**
2	8	10	10	0%	9.2%	0.0%	**2.2%**
3	8	12	7	0%	4.4%	0.2%	**0.7%**
4	8	12	10	0%	3.3%	0.0%	**0.8%**
5	10	10	7	0%	13.2%	0.0%	**1.5%**
6	10	10	10	0%	5.4%	0.2%	**1.1%**
7	10	12	7	0%	9.7%	0.2%	**1.6%**
8	10	12	10	0%	6.4%	0.4%	**1.3%**
9	12	10	7	0%	8.0%	0.2%	**1.2%**
10	12	10	10	0%	6.3%	0.1%	**0.6%**
11	12	12	7	0%	10.8%	0.2%	**1.7%**
12	12	12	10	0%	1.9%	0.1%	**0.4%**

Table 4.3: Performance of the Heuristic

sometimes better than that reported in Table 4.3. The sub-problems of the heuristic are generated in C++ and solved using CPLEX 12.1.0.

As can be seen in Table 4.3, the heuristic is able to obtain the optimum solution for at least one problem instance in each problem class (because the minimum difference in each row is 0%). The median difference is 0.1% and the average difference is only 1.1%. The solutions differ more than 10% in only 1.4% of the cases. As mentioned earlier, I use the best lower bound for the comparison whenever the optimal solution could not be obtained. For some of the large problem instances, the bound gets worse. Therefore, the large gap for some of the problem instances may be because of the worse bounds rather than the performance of the heuristic. Nonetheless, with the heuristic, I am able to find good quality solutions in comparatively much less time. Since the optimal value could not be obtained for many instances, I do not report the comparison of the solution times. However, it takes the heuristic an average of 16.3 seconds to find the solution on a 2.40 GHz Intel dual-core processor with 4.00 GB of RAM.

4.3 Results and Managerial Insights

Since the solution of the dynamic approach depends on the realizations of demand and returns, it is difficult to study it in an analytical setting. On the other hand, the static approach considers expected values for optimization. Hence, I use static approach to derive analytical results and present useful managerial insights. In the following subsections, I first derive the conditions under which the all-inclusive policy provides the optimal solution. Next, I numerically compare the performances of the all-inclusive policy and the heuristic. Then, I examine the impact of error in inventory information on the quality of the solution. After that, I analyze the trade-offs in choosing different threshold levels for recommendations. Finally, I compare the performance of my solution methodology with an individualized recommendation policy and then show the non-optimality of simple greedy heuristics that the DVD firms can utilize in their recommendation decisions. The conditions presented in all of the analytical results in this section, except one, utilize the original parameters of the problem. In other words, the managers can directly check the conditions using the parameter values, without solving any optimization problem.

4.3.1 Sub-optimality of the Prevalent Industry Practice

In this section, I analyze the conditions that are necessary and sufficient for the prevalent industry practice (i.e., the all-inclusive policy) to be optimal. I also argue that these conditions are unrealistic and hence the all-inclusive policy will perform sub-optimal in real business settings. I first define and interpret some basic expressions that will be useful in presenting the results. In the all-inclusive policy, if the rating is more than the threshold, i.e., $(i, j) \in V$, the perceived utility of a movie cluster j is $R_{ij} + (R_{max} - R_{ij})\, E$ due to the constraint set (4.14) and the fact that the movie cluster is recommended to all users in cluster i, i.e., $x_{ij} = 1$. Similarly, if the rating is less than the threshold, i.e., $(i, j) \notin V$, then $x_{ij} = 0$ and the perceived utility of the movie is unchanged at level R_{ij}. Therefore, after the all-inclusive policy recommendations, the sum of perceived utilities for user cluster i over all the movie clusters can be written as

$$S_i = \sum_{s \in J:(i,s) \in V} \left[R_{is} + (R_{max} - R_{is})\, E\right] + \sum_{s \in J:(i,s) \notin V} R_{is}. \tag{4.18}$$

Hence, as shown in constraint set (4.15), the probability that a user in cluster i requests movie cluster j is: (a) $\frac{R_{ij} + (R_{max} - R_{ij})E}{S_i}$ if $(i, j) \in V$, or (b) $\frac{R_{ij}}{S_i}$ if $(i, j) \notin V$. As a result, the total demand for movie cluster j in period t after the recommendations of all-inclusive policy is

$$F_{jt} = \sum_{i \in U:(i,j) \in V} \frac{\left[R_{ij} + \left(R_{max} - R_{ij}\right) E\right] n_{it}}{S_i} + \sum_{i \in U:(i,j) \notin V} \frac{R_{ij} n_{it}}{S_i}, \qquad (4.19)$$

where n_{it} denotes the number of requests from users in cluster i in period t. Using these expressions, I present Proposition 14 below. Proofs are available in Appendix C.

Proposition 14 *The prevalent industry practice is optimal* <u>*iff*</u> *at least one of the following three conditions is satisfied:*

(a) The parameter values are such that the expected demand for all movie clusters are greater than their inventories when everything above threshold is recommended (i.e., all-inclusive policy).

(b) The expected demands for the movie clusters that have at least one (i, j) pair in set V (i.e., the movie clusters that are in the set of recommendable movie clusters for at least one user cluster) are less than their inventories after the recommendations are shown using all-inclusive policy. The movie clusters in which no (i, j) pair is in set V do not need to satisfy this condition.

(c) There is no movie cluster that is recommendable to any user cluster.

Using Equations (4.18) and (4.19), one can restate these conditions mathematically as:

1. $F_{jt} \geq K_{jt}; \quad \forall j \in J,\ t = 1, 2, \ldots, T.$

2. $F_{jt} \leq K_{jt}; \quad \forall j \in J : \exists \gamma \in U : (\gamma, j) \in V,\ t = 1,\ 2, \ldots,\ T.$

3. $V = \emptyset.$

It is important to note that the conditions presented in the proposition represent extreme cases and are unlikely to happen in realistic settings. Further, since these conditions are if-and-only-if, a better recommendation policy can be derived when neither of these conditions satisfy. Therefore, I can conclude that the prevalent industry practice is not optimal for realistic problem instances, and a better solution methodology may be able to improve customer satisfaction.

Let us now analyze the conditions presented in the proposition. In part (a), the firm utilizes its entire inventory in the all-inclusive policy, and cannot satisfy more demand using any other methodology. Therefore, the industry practice is optimal. On the other hand, in part (b), demand is satisfied as much as possible for the movies that are recommended to at least one customer cluster. In this case, since all of the movies with ratings above the threshold are offered, the demands for the other movies are minimized. Hence, although the demands for these movies may not be all satisfied, they are at their minimum levels. Therefore, the all-inclusive policy is again optimal. Finally, in part (c), neither the all-inclusive policy nor the optimal policy can recommend anything. Hence, the all-inclusive policy is optimal.

It is also important to note that the conditions presented in Proposition 14 need to be satisfied throughout the planning horizon. All-inclusive policy is not clairvoyant in the sense that it does not consider future demand and returns. Hence, even if the conditions in the proposition are satisfied in a given period, the all-inclusive policy might not be optimal over the planning horizon. For example, consider the following case: In a time period before the demand for a movie cluster is expected to increase (e.g., Christmas and Christmas themed movies), the condition in part (b) is satisfied. However, since the demand for Christmas related movies is expected to increase in the next period, the optimal policy that generates more surplus inventory in the current period for Christmas movies will perform better than those policies that are myopic such as the all-inclusive policy. In summary, the current industry practice is not optimal in almost all non-trivial cases not only because it does not optimize decisions for a given period, but also because it is not forward-looking.

4.3.2 Comparison of the All-Inclusive Policy with the Proposed Heuristic

The preceding subsection argues and shows that the all-inclusive policy performs sub-optimal in realistic business settings. In this subsection, I complement this finding and examine numerically how much the all-inclusive policy is inferior to my proposed solution approach. To this end, I compare the all-inclusive policy with the proposed heuristic for different problem classes with realistic parameter ranges. The experiment has a full factorial design where I set $|U|$ to 8 and 10, $|J|$ to 10 and 12, and T to 10 and 12, in order to generate eight problem classes. For each of these problem classes, I generate eight problem instances by setting E to 0.25 and 0.30, varying total demand randomly around its mean by 20%, and setting Q_{st} parameters to represent two different return patterns. In each of these problem instances, I simulate the all-inclusive policy and compare it to the heuristic solution. The reported performance values are calculated as the difference between the solutions of all-inclusive policy and heuristic divided by the solution of the all-inclusive policy. As can be seen in Table 4.4, the solutions differ by as much as 59%. On the other hand, there are some instances where the difference is around 2% but never the same. Overall, the average difference between these two solutions is 11%. These results further strengthen my analytical finding that the prevalent industry practice is sub-optimal.

4.3.3 Impact of Error in Inventory Information on the Quality of the Solution

The findings I have discussed so far clearly indicate that, because of the sub-optimality of the all-inclusive policy, it is important to consider inventory explicitly in recommendation decisions. However, the level of inventory may not always be

Problem Class	Problem Size			Objective Value Comparison			
	$\lvert U \rvert$	$\lvert J \rvert$	T	*Min*	*Max*	*Median*	***Average***
1	8	10	10	3.4%	17.5%	10.9%	**10.4%**
2	8	10	12	2.2%	14.9%	7.1%	**7.9%**
3	8	12	10	4.1%	59.3%	6.8%	**13.7%**
4	8	12	12	3.1%	15.3%	5.9%	**6.7%**
5	10	10	10	4.1%	55.9%	19.9%	**23.0%**
6	10	10	12	2.4%	25.5%	5.9%	**8.8%**
7	10	12	10	4.6%	13.7%	5.5%	**7.3%**
8	10	12	12	2.9%	21.1%	9.9%	**10.4%**

Table 4.4: Performance of the Prevalent Industry Practice

estimated with perfect accuracy. Hence, I examine the impact of error in inventory information on the quality of the solution. First, a given instance of the problem is solved with incorrect inventory levels, and the recommendation decisions are recorded. Next, this set of decisions is utilized to simulate what would happen with the correct inventory levels. In each time period, the inventory for each movie cluster is varied between 70% to 130% of the corresponding actual levels, with 10% increments across different experiments. Hence, a specific problem instance is solved seven times. In these experiments, $\{|U|, |J|, T\}$ are set to $\{4, 5, 6\}$, and E to 0.25. I solve different problem instances by varying the number of requests (n_{it}), the starting inventory levels (K_{jt}), and the rating values (R_{ij}). The findings, however, do not change across these instances and can be categorized in two groups. Hence, I present the solution of only two representative problem instances for each group in Figure 4.3.

In Figure 4.3, I present the performance gap between the all-inclusive policy and the static heuristic. Here, problem instance 1 represents cases where the inventory has either ample shortages or surpluses for most of the movie clusters. On the other hand, problem instance 2 is one of the scenarios where most of the movie clusters have expected demand and inventory close to each other. Hence, the second instance represents a critically balanced system. As expected, when both methods utilize the accurate inventory levels, i.e., when the horizontal axis equals to 0, the performance gap between the all-inclusive policy and the static heuristic is at the maximum level (for both problem instances).

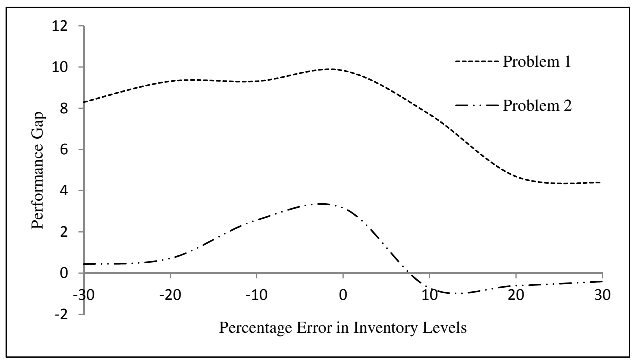


Figure 4.3: Gap between the All-inclusive Policy and the Static Solution

Figure 4.3 also shows that when the inventory is underestimated (i.e., negative values on the x-axis), the percentage gap remains almost unchanged with very little deterioration in problem instance 1. On the other hand, problem instance 2 exhibits a different behavior. Here, the quality of the solution diminishes rather quickly, and there is almost no benefit of using the solution procedure I present in this section when the underestimation error is more than 20%. Now I analyze the overestimation case (i.e., positive values on the x-axis). In problem instance 1, the performance gap diminishes rapidly if the inventory level is overestimated. However, the static heuristic still performs better than the all-inclusive policy. On the other hand, in problem instance 2, the deterioration in the quality of the solution is more dramatic. When the overestimation error is more than 10%, the static approach performs even worse than the all-inclusive policy.

In general, the static approach tries to shift demand from movies that have expected shortages to movies that have expected excess supplies. Hence, even if the inventory is underestimated, the static approach still tries to increase the demand of movies that have excess supplies. However, because of the underestimation, some portion of the inventory of such movies are not utilized that could otherwise attract those customers that are looking for popular movies (with expected shortages). Therefore, I observe that underestimation errors do not deteriorate the quality of the solution for problem instance 1. However, such errors impact the quality of the solution in problem instance 2, because it is a critically balanced system.

On the other hand, when there is overestimation error, the static approach assumes that the firm should be able to satisfy demand with the ample inventory. As a result, it does not shift demand among movies as much as it actually needs to. Hence, the quality of the solution deteriorates quickly. I observe that, in both problem instances, the quality of the solution worsens quickly with overestimation errors. However, because the performance gap even with no error is low in the second problem instance, the static solution becomes worse than the all-inclusive policy. I would also like to note that the problem instance 1 is more likely to be encountered in practice, because most movies either have excess demand (i.e., newly released hot movies) or excess supply (i.e., old movies).

Therefore, an important question arises. What does a firm need to do if there is high uncertainty regarding the inventory levels? Clearly, the answer depends on the firm's expected demand and inventory level. I summarize the answer to this question in the following observation based on the findings discussed above.

Observation 14 *If the firm does not have a critically balanced system, it could in-*

tentionally underestimate the inventory. The reason is that, in these environments, underestimation does not deteriorate the quality of the solution substantially. On the other hand, if the firm has a critically balanced inventory, underestimation as well as overestimation might hurt the quality of the solution. Therefore, intentional underestimation should not be very high. However, the dynamic solution approach performs better than the static approach in critically balanced systems. Hence, especially if there is also high uncertainty regarding the inventory levels, the firm should utilize the dynamic approach and slightly underestimate its inventory.

In the next subsection, I examine the issue that the DVD-rental firms encounter regarding the trade-off between short term profitability and long term customer trust in the recommendations.

4.3.4 Effect of Trust in the Recommender System

As discussed earlier, there are two different effects of the recommendation threshold θ. In the short-term, when the recommendation threshold is lowered, there is more flexibility in offering movie clusters. Hence, in the short-run, the firm is able to shift the demand better among the movies. On the other hand, if the firm chooses a lower threshold level, the customers would perceive that they are offered movies which they do not value highly. Hence, they may accept less and less recommendations in the long-run. This concept is referred to as trust in the recommendation literature (Andersen et al. 2001, Tintarev and Masthoff 2007) and implicitly captured by the parameter E in my model.

As discussed by Kim et al. (2009), the dynamics of trust should be analyzed in the long-run. Hence, while keeping E constant, I first derive an upper bound on the value of reducing threshold level in the short-run. Next, I experimentally study the effect of lowering the threshold level. For simplicity, I present the analytical results for one period, and therefore ignore the subscript for time period in the expressions. Furthermore, before presenting the generalized result, I present the result for the scenario when the threshold level is reduced from its maximum value.

First, I define and interpret some basic expressions that will be useful in presenting the results. If the recommendation threshold is set to its maximum value, then clearly there are no recommendations. Hence, in this case, the perceived utility terms are unchanged. Therefore, as discussed in Section 4.3.1, the demand for a movie cluster j is $F_j = \sum_{i \in U} \frac{R_{ij}}{\sum_{s \in J} R_{is}} n_i$. Now, the dissatisfaction for this movie cluster can be written as $G_j = (F_j - K_j)^+$. Finally, the total dissatisfaction for all movies becomes $H = \sum_{j \in J} G_j$. In the similar manner, the amount of unused inventory for the movie cluster j is $I_j = (K_j - F_j)^+$.

By decreasing the recommendation threshold, the firm has the option of recommending some movies that may increase their demands. However, the maximum increase in demand for a movie cluster is possible when only that movie cluster is recommended to the users. This bound on increase in demand for movie cluster j can be expressed as

$$Q_j = \sum_{p \in U:(p,j) \in V} \left(\frac{R_{pj} + (R_{max} - R_{pj})E}{\sum_{s \in J} R_{ps} + (R_{max} - R_{pj})E} - \frac{R_{pj}}{\sum_{s \in J} R_{ps}} \right) n_p.$$

Now, using these expressions, I present the following proposition.

Proposition 15 *If the recommendation threshold is reduced from its maximum value (i.e., R_{max}) to θ_1, then an upper bound on the increase in customer satisfaction is the minimum of the following two values:*

(a) The total dissatisfaction when θ is set to R_{max}.

(b) The summation of the minimum of following two values over each movie cluster

that became recommendable with the new threshold θ_1: (b1) the amount of unused inventory when θ is set to R_{max}, and (b2) the maximum increase in demand with the new threshold.

Mathematically, I can restate the condition as:

$$\min\left[H, \sum_{j\in J:\exists(p,j)\in V} \min(I_j, Q_j)\right].$$

Note that the condition presented in Proposition 15 is based on exogenous parameter values, similar to that in Proposition 14. Hence, managers can use this condition to analyze the benefit of reducing the threshold level without solving the model. Based on the discussion in Proposition 15, I now generalize the result for the impact of switching threshold level between any two values.

Proposition 16 *An upper bound on the increase in customer satisfaction when decreasing the threshold value from θ_1 to θ_2 is the minimum of the following two values:*

- *(a) The total dissatisfaction when the threshold is set to θ_1.*
- *(b) The summation of the minimum of following two values over each movie cluster that became recommendable with the new threshold θ_2: (b1) the amount of unused inventory when the threshold is set to θ_1, and (b2) the maximum increase in demand with the new threshold θ_2.*

Mathematically, this condition can be restates as:

$$\min\left[\sum_{j\in J} \max\left(0, D_j^{\theta_1} - K_j\right), \sum_{j\in J; V_j\neq\emptyset} \min\left[\max(0, K_j - D_j^{\theta_1}), G\right]\right],$$

where $i \in V_j \quad \forall\theta_1 < R_{ij} < \theta_2$. $D_j^{\theta_1}$ = demand of movie cluster j with optimal recommendations at $\theta = \theta_1$, and $G = \sum_{i\in U:R_{ij}>\theta_1}\left(\frac{R_{ij}+(R_{max}-R_{ij})E}{\sum_{s\in J} R_{is}+(R_{max}-R_{ij})E} - D_j^{\theta_1}\right)$.

Unlike other propositions, in Proposition 16, the condition requires us to know the demand with the optimal set of recommendations for movie cluster j at the

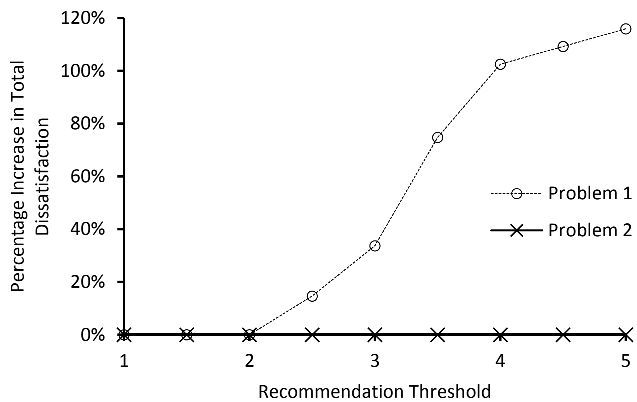


Figure 4.4: Impact of Recommendation Threshold on Customer Dissatisfaction

beginning threshold value θ_1. Actually, a more loose bound (not reported because of brevity) can be calculated without the need of solving the problem. However, once I derive this demand, I can easily obtain this upper bound on the increase in customer satisfaction when decreasing the threshold value to any other level. This is an important result for managers because it helps them in estimating the gain that may potentially be achieved by decreasing the threshold value from the current level to another level.

Propositions 15 and 16 show that the bound on the amount of gain obtained by reducing the threshold level is problem specific. I further study this phenomenon using a variety of problem instances. The findings can be broadly grouped into two scenarios. In Figure 4.4, I plot the results for two problem instances representing each of these scenarios. Clearly, the total dissatisfaction is lowest when the recommendation threshold is set to its minimum value (i.e., 1). In Figure 4.4, the y-axis is the percentage increase in total dissatisfaction if the recommendation threshold is increased from the minimum level. For problem instance 1, there is no effect of switching the threshold from 1 to 2. However, after this value, the effect of switching the threshold level becomes significant. The total dissatisfaction more than doubles when the threshold is increased to its maximum value 5. On the other hand, problem instance 2, which is actually an extreme case, shows no sensitivity to the threshold level at all. Therefore, managers should decide on the threshold level by simulating their own problems and deciding on the level that balances short-term gains and long-term customer trust. Next, in the following subsection, I study how my proposed approach compares with an individualized recommendation policy.

4.3.5 Comparison with an Individualized Recommendation Policy

As discussed earlier, the purpose of clustering is to reduce the number of variables in the optimization model in order to make it tractable. However, although the proposed methodology provides a near-optimal solution, it loses the information at the individual customer level. Hence, it is meaningful to compare the proposed methodology with an individualized policy where the solution is obtained without clustering the customers. Hence, I design an individualized policy and compare it with the proposed optimization model. This study would help us in analyzing the benefit of optimization at the cost of losing individual information. In this section, I also discuss how the clustering impacts the quality of the solution.

The individualized policy is based on the prevalent industry practice, i.e., all-inclusive policy. However, in order to make a more conservative comparison, I consider that the individualized policy is more sophisticated than the all inclusive policy. Let us now describe the details of this individualized policy. The individualized policy is actually a heuristic that begins by checking whether the estimated rating of a specific movie cluster for a specific user is above the recommendation threshold. If this condition is satisfied, then the requests and shipments of this specific movie cluster are checked for the previous period. If all of the requests were satisfied in the prior period, then this movie cluster is recommended to this specific customer;

Percentage Variation	Average Percentage Improvement
0	6.0
5	6.2
10	6.1
20	6.2
30	6.9

Table 4.5: Comparison between the Individualized Policy and the Dynamic Solution

else this movie cluster is not recommended to this customer. Next, I compare the dynamic heuristic with the individualized policy explained above. The results are presented in Table 4.5.

In this experiment, I draw individual rating values within a cluster around the cluster mean with some variations. This captures the fact that the individual rating values within a cluster may be different from the cluster mean. I consider five different levels of variations within clusters (i.e., 0%, 5%, 10%, 20%, and 30%). Here, 0% variation indicates that all individual rating values within a cluster are equal to the cluster mean. In the remaining four levels, the individual rating values are drawn around their cluster means with the corresponding variations. Each of these five levels of variations is presented in a row in Table 4.5. For each level of variation, I consider 40 problem instances by varying the parameter values in realistic ranges. In the second column of Table 4.5, I compare the performances of dynamic heuristic and individualized policy. More specifically, I present the average percentage improvement (across 40 problem instances) in using the dynamic heuristic over the individualized policy.

First and foremost, the results show that the solutions are not sensitive to the variations within clusters at any level of variation. This implies that the clustering does not worsen the solution in this context, and therefore the clustering approach considered in this section is reasonable. Finally, the results also show that the improvement in using dynamic heuristic over individualized policy is more than 6.0% at any level of variation. This clearly indicates that it is beneficial to utilize the proposed dynamic heuristic even though it optimizes the recommendation decision at the cluster level.

Finally, I present two examples in the following proposition where seemingly logical or greedy heuristics are shown to be non-optimal.

Proposition 17

(a) Even when the demand is less than the available inventory for a movie cluster, it may not be optimal to offer this highly available movie to everybody.

(b) Even when the demand for a movie cluster is more than its available inventory, it may be optimal to recommend it to some customers.

Proposition 17 suggests that it may be difficult to obtain a good solution using simple greedy heuristics. Hence, there is a need to design a heuristic that can exploit the structure of the problem effectively, such as the one proposed in Section 4.2. In the next section, I present the optimal recommendation policies for the DVD-rental firms in special cases and provide interesting managerial insights.

4.4 Optimal Recommendation Policies for Special Cases

The goal of this section is to characterize the optimal solution under special cases. These results would provide some guidance for the managers about the solution of the general case. Also, I discuss that, under certain conditions, the managers may directly use this optimal solution without solving the optimization model.

4.4.1 Two Movie Clusters, One User Cluster, and Arbitrary Number of Periods

In this subsection, for ease of discussion, I begin with a restricted case where the number of movie clusters and user clusters are two and one, respectively. For example, the two movie clusters may represent high and low rated movies. In addition, for brevity, I present the results for the single period case in Proposition 18. However, the multiple period case is trivial to develop from this proposition. In this proposition, I ignore the subscripts for customer cluster and time period, because of their singularity. Before presenting the proposition, I define $\delta = \frac{(R_1+R_2)K_1-R_1 n}{(R_{max}-R_1)(n-K_1)}$. As shown in Appendix C, this is the value of the effectiveness parameter that would make demand for movie cluster 1 equal to its inventory when it is recommended to all users. A less restricted case with arbitrary number of user clusters is analyzed in the next subsection.

Proposition 18 *For one user cluster, two movie clusters, one period, and $R_1 > \theta$:*

1. *If the expected demands for both movie clusters are less than the respective inventories, then an optimal policy is to recommend nothing, i.e., $x_1 = x_2 = 0$.*

2. *If the expected demands for both movie clusters are more than the respective inventories, then again an optimal policy is to recommend nothing, i.e., $x_1 = x_2 = 0$.*

3. *If only one movie cluster has surplus inventory (say cluster 1) and this available inventory is greater than the total demand (i.e., $n \leq K_1$), then an optimal policy is: $x_1 = 1$, $x_2 = 0$.*

4. *If only one movie cluster has surplus inventory (say cluster 1) and this available inventory is less than the total demand (i.e., $n > K_1$), then an optimal policy is: $x_1 = \min\left[\frac{\delta}{E}, 1\right], \quad x_2 = 0$.*

In part (a) of the proposition, the firm is able to meet all the demand before any recommendations. Therefore, there is no need to recommend anything. In part (b), both movie clusters are requested more than the firm can supply. Hence, again there is no need to recommend anything, because the firm is already utilizing its entire inventory and it is not possible to meet more demand. In part (c), the movie cluster 1 has more inventory than the total demand. Increasing the demand for movie cluster 1 is offset by the equal decrease in the demand for movie cluster 2. Here, it is optimal to increase the demand for movie cluster 1 as much as possible, because its demand can never be more than its inventory.

Finally, in part (d) of the proposition, the idea is to increase the demand for movie cluster 1 as much as possible but not more than its inventory. The surplus inventory cannot be utilized fully if the effectiveness parameter E is less than δ. Therefore, in this case, it is optimal to recommend movie cluster 1 to everybody with expected unsatisfied demand equal to $\left(\frac{R_2 n}{R_1+(R_{max}-R_1)E+R_2} - K_2\right)^+$. If E is more than δ, then recommending movie cluster 1 to everybody will result in expected shortage in movie cluster 1. Hence, setting $x_1 = \frac{\delta}{E}$ is optimal because it makes the demand and inventory for movie cluster 1 equal. In parts (c) and (d), it is optimal not to recommend movie cluster 2 to anybody because its demand increases in x_2.

4.4.2 Two Movie Clusters, Arbitrary Number of User Clusters, and Arbitrary Number of Periods

Next, I consider a more general case with two movie clusters and any number of user clusters. Before presenting the results, I denote

$$\gamma = \frac{(R'_{s1} + R_{s2})\left(K_1 - \sum_{i \in U \backslash s}\left(\frac{R'_{i1}}{R'_{i1}+R_{i2}} n_i\right)\right) - R'_{s1} n_s}{(R_{max} - R'_{11})\left(n_1 - K_1 + \sum_{i \in U \backslash s}\left(\frac{R'_{i1}}{R'_{i1}+R_{i2}} n_i\right)\right)}.$$

I explain later that this is a critical value of the effectiveness parameter when the demand of movie cluster 1 is set equal to its inventory. For brevity of discussion, I present the results for only one period in the proposition below where I ignore the subscript for time.

Proposition 19 *For two movie clusters, one period, and any number of user clusters:*

(a) If the expected demands for both movie clusters are less than the respective inventories, then an optimal policy is to recommend nothing, i.e., $x_{ij} = 0$*;* $\forall i, j$.

(b) If the expected demands for both movie clusters are more than the respective inventories, then again an optimal policy is to recommend nothing, i.e., $x_{ij} = 0$*;* $\forall i, j$.

(c) If only one movie cluster has surplus inventory (say cluster 1), then an optimal set of recommendations can be obtained by using the following steps:

(i) Sort and re-label the customer clusters according to R_{i1} *values, and find* k *such that* $R_{k1} > \theta$ *and* $R_{(k+1)1} \leq \theta$.

(ii) Initialization: $s = 0$, $x_{i2} = 0\ \forall i$.

(iii) Increase s *by 1. If* $s > k$*, then the optimal solution is:* $x_{i1} = 1$ *for* $i \leq k$, $x_{i1} = 0$ *for* $i > k$*, and* $x_{i2} = 0 \quad \forall i$*, and terminate. Otherwise (i.e., when* $s \leq k$*), set* $x_{p1} = 0$ *for* $p > s$*, and denote* $R'_{i1} = \begin{cases} R_{i1} + (R_{max} - R_{i1})\, E & i < s \\ R_{i1} & i \geq s \end{cases}$

(iv) If $n_s - K_1 + \sum_{i \in U \backslash s} \left(\frac{R'_{i1}}{R'_{i1} + R_{i2}} n_i \right) < 0$*, then set* $x_{s1} = 1$*, and go to Step (iii). Otherwise, go to Step (v).*

(v) If $E < \gamma$, then set $x_{s1} = 1$, and go to Step (iii). Otherwise, set the recommendations as: $x_{i1} = 1$ for $i < s$, $x_{s1} = \frac{\gamma}{E}$, $x_{i1} = 0$ for $i > s$, and $x_{i2} = 0\ \forall i$, and terminate.

The explanations for parts (a) and (b) are similar to those for Proposition 18. Also, similar to Proposition 18, in part (c), one should aim to increase the demand for movie cluster 1 as much as possible without making it more than its inventory. In this case, there is one equation, namely the demand-inventory balance equation for the movie cluster 1, but there are k unknowns. I propose a solution methodology which follows from the observation that the recommendation of movie cluster 1 to a specific user cluster does not affect the demand of movies for other user clusters. Hence, an optimal solution is to check (starting from the first user cluster) whether recommending movie cluster 1 to user cluster s can increase the demand of movie cluster 1 to the starting inventory level K_1. Step (iv) checks whether it is possible to increase the demand up to the inventory level by recommending only up to the user cluster s. If it is not enough, then the proposed policy is to recommend movie cluster 1 to everybody in the user cluster and go back to step (iii) to restart the procedure for the next user cluster. Otherwise, in step (v), I check whether the efficiency of the recommendations is below γ. In such a case, recommending movie cluster 1 to user cluster s does not increase its demand more than the inventory level, therefore I set $x_{s1} = 1$. Otherwise, x_{s1} is set to $\frac{\gamma}{E}$ so that the demand for movie cluster 1 equals to its inventory, and the optimal policy is stated at the end of step (v). Else, the procedure revisits step (iii) for the next user cluster.

4.4.3 Managerial Insights

There are two key advantages of the results presented in this section. First, Proposition 19 can be applied to settings for any number of user clusters. Hence, when users are very diverse in their preferences, utilizing the procedure described in the proposition can provide an individualized solution to managers. Second, when the diversity of the rating estimates are low, the firm may be able to group movies into two clusters without sacrificing much information about movies. If the firm is able to do so, Proposition 19 would provide them the explicit structure of the optimal solution. The managers may use this information effectively in making different strategic and tactical decisions. However, when the diversity of the rating estimates are high, the firm may not be able to group movies into two clusters effectively. In other words, the negative impact of grouping movies into two clusters may be more than the gain from the characterization of optimal solution. In such cases, the firm should rather utilize the methodology discussed in Section 4.1.1 with higher number of movie clusters. However, even for these cases, the managers may utilize the solution presented in Proposition 19 to derive some insights regarding the structure of the solution.

I would also like to mention that the optimal procedure presented in Proposition 19 can be extended to settings where the number of movie clusters is more than two (but not very high). Clearly, the procedure starts getting cumbersome as the number of movie clusters increases. However, since the computational efficiency is increasing substantially with time, one may be able to solve these procedures for higher number of movie clusters. The details are omitted for brevity.

4.5 Conclusions

This study builds on the argument that the recommendations should be tailored according to both the preferences of users and the inventory information. DVD-rental firms usually do not consider the inventory of the movies explicitly while making recommendations. The analytical and experimental results show that this prevalent industry practice is sub-optimal, and the customer satisfaction can be improved significantly by considering inventory in recommendation decisions.

I propose a practicable dynamic solution approach where the future decisions are based on the observation of actual rental and return events in current and past periods. The related analysis reveals that it is beneficial for the firm to adopt dynamically when the system is critically balanced, i.e., demand and inventory are close to each other for some movies. However, if the system is not critically balanced, then the firm may consider the solution of the static approach. It may help them in saving the costs related to updating parameter values and solving the problem in each period. On the other hand, if there is high uncertainty regarding the inventories and if the system is critically balanced, I show that the static approach might be slightly inferior to the prevalent industry practice. In this case, I propose utilizing the dynamic approach and intentionally underestimating inventories as quick recipes to the DVD-rental firms.

Next, I analyze the impact of the recommendation threshold level on short-term profitability and long-term customer trust trade-off. The results indicate that the gain from changing the threshold level is problem-specific. Hence, the managers need to carefully analyze their business settings in order to choose the best threshold level that will balance profitability in the short-term and user trust in the long-term.

In addition, I discuss how the proposed methodology can be utilized by the DVD rental firms at the individual customer level. Finally, I present the optimal solution in the case of two movie clusters, any number of user clusters, and any number of time periods. When the diversity of the rating estimates are low, the firm may be able to group movies into two clusters without sacrificing much information about movies. In such a case, the managers can characterize the optimal procedure that would help them in making different strategic and tactical decisions.

5. SUMMARY AND CONCLUSION

In this dissertation, I present three essays where I study coordination and collaboration settings within and across companies. In the first essay, I study the relationship between a client and a vendor in value co-creation environments. I consider that the client gets utility from the project throughout the development period. The output is contingent on the effort levels of each party and I allow these effort levels to be dynamic. Hence, the client needs to optimally decide the terms of the payment so as to maximize the project output and minimize its cost.

Taking these characteristics of the collaboration into consideration, I examine three differential game settings. I derive the effort levels of the client and the vendor at the equilibrium and the optimal payment parameters in the contracts. I also study how different contracts compare under different settings in order to find the best one for the client. The result is that, if the output is not very sensitive to vendor's effort, the effort dependent structure is better than the output dependent structure for the client. On the other hand, if the output is very sensitive to the effort of the vendor, then the client should offer payments based on output in order to give enough incentive to the vendor to spend more effort and generate more value together.

Moreover, if the client's contract selection also includes the hybrid structure, and if the output is moderately or highly sensitive to vendor's effort, the hybrid contract is better for the client. This implies that the vendor should be offered a share of the output, as well as payments related to the effort he spends in the generation of the output. If the output is not much sensitive to vendor's effort, the client should prefer an effort dependent structure. I further present several interesting findings based on the sensitivity analyses of the equilibrium effort levels and value of the client with respect to the parameters representing different characteristics of the parties. Finally, I consider that the client may offer training to their vendors in order to increase their productivity or lower their participation costs. I show that the client should provide training to only those vendors that have relatively high costliness.

In the second essay, I study another value co-creation environment. In this case, unlike the first essay, I assume that the effort levels are not observable but might be monitored and examine two different settings from a differential game perspective. My analysis reveals that the client's valuation of the project and the progress in the collaboration play an important role in the behavior of the effort levels. Depending on these factors, the equilibrium effort levels might increase or decrease with the changes in the output sensitivities to effort levels and the cost elasticities of efforts.

Moreover, I compare the performances of different contracts in order to find the best one for the client under different circumstances. This analysis also reveals whether the vendor's effort should be monitored or not. I find that, as long as the participation cost of the vendor is not very low, his effort should be monitored and the client should use an effort dependent contract. Otherwise, the client should not monitor vendor's effort, and should operate under double moral hazard with an output dependent contract. Another finding is that, if the sensitivity of output to vendor's effort is relatively higher than the sensitivity to client's effort, the client should use the output dependent contract and should not monitor vendor's effort. I also consider pure revenue-sharing contracts and find that the total value generated in such an environment is lower than that in the output dependent contract.

In the third essay, I consider a subscription based rental organization. In these environments, the satisfaction of customers depends on the availability of requested products. Hence, it is important for these firms to satisfy as much demand as possible. Recommender systems, in a DVD-rental context, are typically used to help customers in finding the right movies for them. However, DVD-rental firms usually do not consider the inventory of the movies explicitly while making recommendations. My results show that this prevalent industry practice is sub-optimal, and the customer satisfaction can be improved significantly by considering inventory in recommendation decisions. I propose two recommendation approaches to the firms depending on the nature of their business, i.e., whether or not they have a critically balanced system. I also analyze the impact of the recommendation threshold level on short-term profitability and long-term customer trust trade-off. The results indicate that the gain from changing the threshold level is problem-specific. Hence, the managers need to carefully analyze their business settings in order to choose the best threshold level that will balance profitability in the short-term and user trust in the long-term. Finally, I present the optimal solution for some special cases of the problem.

REFERENCES

Abel, N. H. 1881. *Oeuvres Completes, Two Volumes.* Grondahl and Son, Christiana, Norway.

Adomavicius, G., A. Tuzhilin. 2005. Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions. *IEEE Transactions on Knowledge and Data Engineering* **17**(6) 734–749.

Alden, J. M., R. L. Smith. 1992. Rolling horizon procedures in nonhomogeneous Markov decision processes. *Operations Research* **40**(2) S183–S194.

Amazon.com. 2011. *Amazon Annual Meeting of Shareholders Presentation.* Available at http://phx.corporate-ir.net/phoenix.zhtml?c=97664&p=irol-presentations (last accessed: February 26, 2013).

Andersen, V., C. Hansen, H. Andersen. 2001. Evaluation of agents and study of end-user needs and behavior for e-commerce. *COGITO Focus Group Experiment Report CHMI*-01-01, Denmark, 1–42.

Arrow, K. J., M. Kurz. 1970. *Public Investment, the Rate of Return, and Optimal Fiscal Policy*. The John Hopkins Press, Baltimore, MD.

Bajari, P., S. Tadelis. 2001. Incentives versus transaction costs: A theory of procurement contracts. *Rand Journal of Economics* **32**(3) 387–407.

Bassamboo, A., R. S. Randhawa. 2007. Optimal control in a Netflix-like closed rental system. Working paper, Kellogg School of Business, Chicago, IL.

Bassamboo, A., S. Kumar, R. S. Randhawa. 2009. Dynamics of new product introduction in closed rental systems. *Operations Research* **57**(6) 1347–1359.

Bennett, J., S. Lanning. 2007. The Netflix prize. *Proceedings of the KDD Cup and Workshop.* San Jose, CA, 3–6.

Bettencourt, L. A., A. L. Ostrom, S. W. Brown, R. I. Roundtree. 2002. Client co-production in knowledge-intensive business services. *California Management Review* **44**(4) 100–128.

Bhattacharyya, S., F. Lafontaine. 1995. Double-sided moral hazard and the nature of share contracts. *The RAND Journal of Economics* **26**(4) 761–781.

Bhuiyan, N., A. Baghel. 2005. An overview of continuous improvement: From the past to the present. *Management Decision* **43**(5) 761 – 771.

Bodapati, A. V. 2008. Recommendation systems with purchase data. *Journal of Marketing Research* **45**(1) 77–93.

Bond, P. 2011. Is Netflix good for Hollywood? *Hollywood Reporter* **417**(2) 54–57.

Box, G. E. P., W. G. Hunter, J. S. Hunter. 1978. *Statistics for Experimenters.* John Wiley and Sons, New York.

Brynjolfsson, E., Y. Hu, M. Smith. 2003. Consumer surplus in the digital economy: Estimating the value of increased product variety. *Management Science* **49**(11) 1580–1596.

Cachon, G., S. Netessine. 2004. Game theory in supply chain analysis. D. Simchi-Levi, S. Wu, Z. Shen, eds., *Handbook of Quantitative Supply Chain Analysis: Modeling in the E-Business Era*. Kluwer Academic Publishers, Norwell, MA.

Chen, L., P. Pu. 2005. Trust building in recommender agents. *Proceedings of the Workshop on Web Personalization, Recommender Systems and Intelligent User*. Reading, UK. 1–10.

Chung, C. 2010. Demand-forecasting, product-allocation and production-scheduling applications at Blockbuster. PhD Thesis. University of Texas at Dallas. Available at http://gradworks.umi.com/34/14/3414896.html (last accessed: February 26, 2013).

Cohen, M. A., J. Eliashberg, T.H. Ho. 1996. New product development: The performance and time-to-market tradeoff. *Management Science* **42**(2) 173–186.

Currie, W. L., B. Galliers, eds. 1999. *Rethinking Management Information Systems: An Interdisciplinary Perspective*. Oxford University Press, New York.

De, P., Y. Hu, M. S. Rahman. 2010. Technology usage and online sales: An empirical Study. *Management Science* **56**(11) 1930–1945.

Dockner, E., S. Jorgensen, N. V. Long, G. Sorger. 2000. *Differential Games in Economics and Management Science*. Cambridge University Press, Cambridge, UK.

Erhun, F., P. Keskinocak. 2011. Collaborative supply chain management. K. Kempf, P. Keskinocak, R. Uzsoy, eds., *Planning Production and Inventories in the Extended Enterprise*. Springer, New York.

Ferracone, R. 2011. Performance share plans - the best of both worlds. *Forbes* **March**(3). Available at http:/Forbes.com/sites/robinferracone/2011/03/03/performance-share-plans-the -best-of-both-worlds (last accessed: July 22, 2012).

Fleder, D., K. Hosanagar. 2009. Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity. *Management Science* **55**(5) 697–712.

Garfinkel R. , R. D. Gopal, B. Pathak, F. Yin. 2008. Shopbot 2.0: Integrating recommendations and promotions with comparison shopping. *Decision Support Systems* **46**(1) 61–69.

Gasper, G., M. Rahman. 2004. *Basic Hypergeometric Series*. Cambridge University Press, Cambridge, UK.

Gil, R., F. Lafontaine. 2012. Using revenue sharing to implement flexible prices: Evidence from movie exhibition contracts. *The Journal of Industrial Economics* **60**(2) 187–219.

Grover, R., C. Edwards. 2011. Can Netflix find its future by abandoning the past? *Bloomberg Businessweek* **September**(22) 29–30.

Hasija, S., E. J. Pinker, S. A. Robert. 2008. Call center outsourcing contracts under information asymmetry. *Management Science* **54**(4) 793–807.

Hastings, W. R, N. D. Hunt, M. B. Randolph, J. R. Ciancutti, S. M. Lanning. 2008. Approach for estimating user ratings of items. *United States Patent* Patent No.: US 7,403,910 B1.

Herlocker J., J. Konstan, L. Terveen, J. Riedl. 2004. Evaluating collaborative filtering recommender systems. *ACM Transactions on Information Systems* **22**(1) 5–53.

Hosanagar, K., R. Krishnan, L. Ma. 2008. Recommended for you: The impact of profit incentives on the relevance of online recommendations. *International Conference on Information Systems,* Paris, France, 1–17.

Hopp, W. J., S. M. R. Iravani, F. Liu. 2009. Managing white-collar work: An operations-oriented survey. *Production & Operations Management* **18**(1) 1–32.

Jain, A. K., M. N. Murty, P. J. Flynn. 1999. Data clustering: A review. *ACM Computing Surveys* **31**(3) 264–323.

Jedidi K., R. E. Krider, C. B. Weinberg. 1998. Clustering at the movies. *Marketing Letters* **9**(4) 393–105.

Karmarkar, U. S., R. Pitbladdo. 1995. Service markets and competition. *Journal of Operations Management* **12**(3-4) 397–411.

Kim, D. J., D. L. Ferrin, H. R. Rao. 2009. Trust and satisfaction, two stepping stones for successful e-commerce relationships: A longitudinal exploration. *Information Systems Research* **20**(2) 237–257.

Kim, S.-H., S. Nettessine. 2012. Collaborative cost reduction and component procurement under information asymmetry. *Management Science* **59**(1) 189–206.

Kouvelis, P., M. A. Lariviere. 2000. Decentralizing cross-functional decisions: Coordination through internal markets. *Management Science* **46**(8) 1049–1058.

Lehmann, D. R., C. B. Weinberg. 2000. Sales through sequential distribution channels: An application to movies and videos. *Journal of Marketing* **64**(3) 18-33.

Liedtke, M. 2012. Gaps in Netflix's online library likely to persist. *Yahoo Finance* **April**(9).

Luce, R. D. 1959. *Individual Choice Behavior: A Theoretical Analysis.* Wiley, New York.

Løwendahl, B. R. 2005. *Strategic Management of Professional Service Firms*. Copenhagen Business School Press, Copenhagen, Denmark.

Mentzer, J. T., J. H. Foggin, S. L. Golicic. 2000. Collaboration: The enablers, impediments, and benefits. *Supply Chain Management Review* **4**(4) 52.

Milkman, K. L., T. Rogers, M. H. Bazerman. 2009. Highbrow films gather dust: Time-inconsistent preferences and online DVD rentals. *Management Science* **55**(6) 1047–1059.

Murthi, B. P. S., S. Sarkar. 2003. The role of the management sciences in research on personalization. *Management Science* **49**(10) 1344–1362.

Nerlove, M., K. J. Arrow. 1962. Optimal advertising policy under dynamic conditions. *Economica* **29**(114) 129–142.

Netflix. 2012. *Netflix Q2 12 Letter to Shareholders.* Available at http://ir.netflix.com/results.cfm (last accessed: Feb 26, 2013).

Pathak, B., R. Garfinkel, R. D. Gopal, R. Venkatesan, F. Yin. 2010. Empirical analysis of the impact of recommender systems on sales. *Journal of Management Information Systems* **27**(2) 159–188.

Patrick, M. F. 2009. *Advanced Calculus*. American Mathematical Society, Providence, RI.

Pigoski, T. M. 1996. *Practical Software Maintenance: Best Practices for Managing Your Software Investment*. Wiley, New York.

Plambeck, E. L., T. A. Taylor. 2006. Partnership in a dynamic production system with unobservable actions and noncontractible output. *Management Science* **52**(10) 1509–1527.

Roels, G., U. S. Karmarkar, S. Carr. 2010. Contracting for collaborative services. *Management Science* **56**(5) 849–863.

Ross, S. 2009. *A First Course in Probability.* Prentice Hall, Upper Saddle River, NJ.

Rubinfield, D. L., S. Scotchmer. 1993. Contingent fees for attorneys: An economic analysis. *RAND Journal of Economics* **24**(3) 343–356.

Schafer, B. J., J. A. Konstan, J. Riedl. 2001. E-commerce recommendation applications. *Data Mining and Knowledge Discovery* **5**(1-2) 115–153.

Senecal, S., J. Nantel. 2004. The influence of online product recommendations on consumer's online choices. *Journal of Retailing* **80**(2) 159–169.

Sethi, S.P., G.L. Thompson. 2000. *Optimal Control Theory: Applications to Management Science and Economics*. Kluwer Academic Publishers, Boston, MA.

Shapiro, C., H. R. Varian. 1999. *Information Rules: A Strategic Guide to the Network Economy.* Harvard Business School Press, Cambridge, MA.

Shih, W., S. Kaufman, D. Spinola. 2009. *Netflix.* Case Study, Harvard Business School, Cambridge, MA.

Swaminathan, J. M., S. R. Tayur. 2003. Models for supply chains in e-business. *Management Science* **49**(10) 1387–1406.

Tang, C. S. 2010. A review of marketing-operations interface models: from co-existence to coordination and collaboration. *International Journal of Production Economics* **125**(1) 22–40.

Thibodeau, P. 2012. Netflix guts data center in shift to cloud. *Computer World* **September**(11).

Tintarev, N., J. Masthoff. 2007. A survey of explanations in recommender systems. *Workshop on Web Personalization, Recommender Systems and Intelligent User Interfaces*, Istanbul, Turkey, 1–10.

Toppin, G., F. Czerniawska. 2005. *Business Consulting: A Guide to How It Works and How to Make It Work*. Profile Books Ltd., London, UK.

Tsay, A. A., N. Agrawal. 2000. Channel dynamics under price and service competition. *Manufacturing & Service Operations Management* **2**(4) 372–391.

Velte, T., A. Velte, R. Elsenpeter. 2009. *Cloud Computing, A Practical Approach.* McGraw-Hill, New York.

Xiao, B., I. Benbasat. 2007. E-commerce product recommendation agents: Use, characteristics, and impact. *MIS Quarterly* **31**(1) 137-209.

APPENDIX A

PROOFS OF LEMMAS AND PROPOSITIONS IN SECTION 2

Proof of Remark 1

Because of the bounds on the parameter values, one can easily write:

$$(\gamma-\alpha)\delta-\beta\gamma=\gamma\delta-\alpha\delta-\beta\gamma>\gamma\delta-(1-\beta)\delta-\beta\gamma\geq\gamma\delta-\max(\gamma,\delta)>0.$$

This remark is utilized in many of the following proofs in Section 2 with or without mention. ∎

Proof of Lemma 1

The Hamiltonians for the client and the vendor (i.e., $H_c(t)$ and $H_v(t)$, respectively) can be written as below (Arrow and Kurz 1970):

$$H_c(t)=kq(t)-c_c u(t)^{\gamma}-pv(t)+\lambda_1(t)u(t)^{\alpha}v(t)^{\beta},\quad \text{and}$$
$$H_v(t)=pv(t)-c_v v(t)^{\delta}+\lambda_2(t)u(t)^{\alpha}v(t)^{\beta}.$$

The multipliers $\lambda_1(t)$ and $\lambda_2(t)$ are given by:

$$\lambda_1(t)=-\tfrac{\partial H_c(t)}{\partial q(t)}=-k, \text{ and } \lambda_2(t)=-\tfrac{\partial H_v(t)}{\partial q(t)}=0,$$

Solution of the above differential equations with the boundary conditions $\lambda_1(T)=0$, and $\lambda_2(T)=0$ gives $\lambda_1(t)=-kt+kT$, and $\lambda_2(t)=0$.

The equilibrium behavior necessitates the derivatives of the Hamiltonians' with respect to the control variables to be zero. Hence, I need to solve the following system of equations

$$\tfrac{\partial H_c(t)}{\partial u(t)}=0, \text{ and } \tfrac{\partial H_v(t)}{\partial v(t)}=0.$$

Thus, after substituting the values of $\lambda_1(t)$ and $\lambda_2(t)$, I have:

$$-\gamma c_c u(t)^{-1+\gamma} + \alpha u(t)^{-1+\alpha} v(t)^{\beta}(-kt + kT) = 0, \quad \text{and}$$
$$p - \delta c_v v(t)^{-1+\delta} = 0.$$

The solution of the above equation system results in the presented equilibrium levels. Second order conditions for this differential game is satisfied, because the Hamiltonians are strictly concave. It is also easy to check that the vendor's reservation utility is positive with any p value so that the last constraint in the model is satisfied due to the starting assumption that $\delta > 1$. Besides, I do not need to impose a constraint stating that the utility for the client would be nonnegative. This is because $p = 0$ is already a trivial solution to the differential game that has nonnegative returns to the client. ■

Proof of Lemma 2

If I substitute the equilibrium values of the effort levels into the maximization problem of the client, and take the integrals, I have the following problem:
$\max_p k\frac{T(\gamma-\alpha)\gamma c_c}{k\alpha(2\gamma-\alpha)}\left(\frac{kT\alpha}{\gamma c_c}\left(\frac{p}{\delta c_v}\right)^{\frac{\beta}{\delta-1}}\right)^{\frac{\gamma}{\gamma-\alpha}} - c_c\frac{T^{\frac{2\gamma-\alpha}{\gamma-\alpha}}(\gamma-\alpha)}{2\gamma-\alpha}\left(\frac{k\alpha}{\gamma c_c}\left(\frac{p}{\delta c_v}\right)^{\frac{\beta}{\delta-1}}\right)^{\frac{\gamma}{\gamma-\alpha}} - pT\left(\frac{p}{\delta c_v}\right)^{\frac{1}{\delta-1}}.$

Given this parametric equation that is observed in the equilibrium, the client can maximize over the payment per unit effort, i.e., p. First order condition with respect to p reveals the reported critical value and shows that it is unique. Second order condition is involved to examine. However, it is easy to show that $\exists p$ that results in positive value to the client for all set of parameter values. Moreover, there exists an upper bound for p above which the value for the client is negative. The lower bound for p is obviously zero. Besides, I can disregard the reservation utility constraint for the vendor, because it is always satisfied. Therefore, the constraint set is compact. Hence, due to Weierstrass Theorem (Patrick 2009), the reported p value is the global maximum. ∎

Proof of Lemma 3

The Hamiltonians for the client and the vendor (i.e., $H_c(t)$ and $H_v(t)$, respectively) can be written as below (Arrow and Kurz 1970):

$$H_c(t) = kq(t) - c_c u(t)^\gamma - lq(t) + \lambda_1(t)u(t)^\alpha v(t)^\beta, \quad \text{and}$$
$$H_v(t) = lq(t) - c_v v(t)^\delta + \lambda_2(t)u(t)^\alpha v(t)^\beta.$$

The multipliers $\lambda_1(t)$ and $\lambda_2(t)$ are given by:

$$\lambda_1(t) = -\tfrac{\partial H_c(t)}{\partial q(t)} = -k + l, \text{ and } \lambda_2(t) = -\tfrac{\partial H_v(t)}{\partial q(t)} = -l.$$

Solution of the above differential equations with the boundary conditions $\lambda_1(T) = 0$, and $\lambda_2(T) = 0$ results in $\lambda_1(t) = (k-l)(T-t)$, and $\lambda_2(t) = l(T-t)$. The equilibrium behavior necessitates the derivatives of the Hamiltonians' with respect to the control variables to be zero. Hence, I need to solve the following system of equations

$$\tfrac{\partial H_c(t)}{\partial u(t)} = 0, \text{ and } \tfrac{\partial H_v(t)}{\partial v(t)} = 0.$$

Thus, after substituting the values of $\lambda_1(t)$ and $\lambda_2(t)$, I have:

$$-\gamma c_c u(t)^{\gamma-1} + (k-l)(T-t)\alpha u(t)^{\alpha-1}v(t)^\beta = 0, \quad \text{and}$$
$$-\delta c_v v(t)^{-1+\delta} + l(T-t)\beta u(t)^\alpha v(t)^{\beta-1} = 0.$$

The solution of the above equation system results in the presented equilibrium levels. Second order conditions for this differential game is satisfied, because the Hamiltonians are strictly concave. It is also easy to check that the vendor's reservation utility is positive with any l with $0 < l < k$ so that the last constraint in the model is always satisfied. ∎

Proof of Lemma 4

If I substitute the equilibrium values of the effort levels into the maximization problem of the client, and take the integrals, I have the following objective after I rearrange the terms:

$$\max_l \frac{T^{\frac{\beta\gamma+\alpha\delta}{\beta\gamma+\delta(\alpha-\gamma)}} \frac{(\gamma-\alpha)\delta-\beta\gamma}{(2\gamma-\alpha)\delta-\beta\gamma}}{\left[(k-l)\left(\left(\frac{(k-l)\alpha}{\gamma c_c}\right)^{\alpha\delta}\left(\frac{l\beta}{\delta c_v}\right)^{\beta\gamma}\right)^{\frac{1}{-\beta\gamma-\alpha\delta+\gamma\delta}} - c_c\left(\left(\frac{(k-l)\alpha}{\gamma c_c}\right)^{-\beta+\delta}\left(\frac{l\beta}{\delta c_v}\right)^{\beta}\right)^{\frac{\gamma}{\gamma\delta-\beta\gamma-\alpha\delta}}\right]^{-1}}.$$

Given this parametric equation that is observed in the equilibrium, the client can maximize over the transfer payment per output parameter, i.e., l. First order condition with respect to l reveals the reported critical value, i.e., $l = k\frac{\beta}{\delta}$, and shows that it is unique. Second order condition reveals that $l = k\frac{\beta}{\delta}$ is a local maximum. The upper and lower bounds for l are obviously k and zero. Besides, I can disregard the reservation utility constraint for the vendor, because it is always satisfied. Therefore, the constraint set is compact. Hence, due to Weierstrass Theorem (Patrick 2009), the reported l value is the global maximum. ∎

Proof of Proposition 1

The proof follows from the derivatives of equilibrium effort levels (when the payment term is set to its optimal level) and the optimal payment term with respect to the corresponding parameters. ∎

Proof and Threshold Values of Proposition 2

The proof follows from the derivatives of equilibrium effort levels (when the payment term is set to its optimal level) and the optimal payment term with respect to the corresponding parameters. The threshold values are:

$$E_{\alpha c} = e^{\frac{\beta(\beta\gamma+(\alpha-\gamma)\delta)}{\delta^2(2\gamma-\alpha)} - \frac{(\beta\gamma+(\alpha-\gamma)\delta)^2}{\alpha\delta^2(\gamma-\alpha)}} \left(\frac{\gamma c_c}{\alpha(T-t)}\right)^{\frac{(\beta\gamma+(\alpha-\gamma)\delta)^2}{\delta^2(\alpha-\gamma)^2}} \left(\frac{\delta^2 c_v(2\gamma-\alpha)}{T\beta(\gamma-\alpha)}\right)^{\frac{\beta}{\delta}} \left(\frac{\gamma c_c}{T\alpha}\right)^{\frac{\beta\left(\gamma^2(\delta-\beta)+\alpha^2\delta\right)}{\delta^2(\alpha-\gamma)^2}},$$

$$E_{\alpha v} = \frac{1}{T} e^{\frac{-((\gamma-\alpha)\delta-\beta\gamma)(\gamma-\alpha)}{\gamma\delta(2\gamma-\alpha)}} \left(\frac{\delta^2 c_v(2\gamma-\alpha)}{\beta(\gamma-\alpha)}\right)^{\frac{\beta}{\delta}} \left(\frac{\gamma c_c}{\alpha}\right)^{1-\frac{\beta}{\delta}},$$

$$E_{\beta c} = \frac{1}{T} e^{\frac{-((\gamma-\alpha)\delta-\beta\gamma)}{\gamma\delta}} \left(\frac{\delta^2 c_v(2\gamma-\alpha)}{\beta(\gamma-\alpha)}\right)^{1-\frac{\alpha}{\gamma}} \left(\frac{\gamma c_c}{\alpha}\right)^{\frac{\alpha}{\gamma}},$$

$$E_{\beta v} = \frac{1}{T} e^{\frac{-((\gamma-\alpha)\delta-\beta\gamma)(\gamma-\alpha)}{\beta\gamma^2}} \left(\frac{\delta^2 c_v(2\gamma-\alpha)}{\beta(\gamma-\alpha)}\right)^{1-\frac{\alpha}{\gamma}} \left(\frac{\gamma c_c}{\alpha}\right)^{\frac{\alpha}{\gamma}},$$

$$E_{\gamma c} = \frac{e^{\frac{((\gamma-\alpha)\delta-\beta\gamma)\left(\left(\alpha^2+2\gamma^2\right)(\beta-\delta)+\alpha\gamma(3\delta-2\beta)\right)}{\gamma\delta(2\gamma-\alpha)(\gamma-\alpha)(\delta-\beta)}}}{\left(\frac{\gamma c_c}{\alpha(T-t)}\right)^{\frac{((\gamma-\alpha)\delta-\beta\gamma)^2}{-\delta(\gamma-\alpha)^2(\delta-\beta)}} \left(\frac{\delta^2 c_v(2\gamma-\alpha)}{T\beta(\gamma-\alpha)}\right)^{\frac{\delta}{\beta}} \left(\frac{\gamma c_c}{T\alpha}\right)^{\frac{\alpha\beta(2\gamma(\delta-\beta)-\alpha(2\delta-\beta))}{-\delta(\alpha-\gamma)^2(\delta-\beta)}}},$$

$$E_{\gamma v} = \frac{1}{T} e^{\frac{-((\gamma-\alpha)\delta-\beta\gamma)(\gamma-\alpha)}{\gamma\delta(2\gamma-\alpha)}} \left(\frac{\delta^2 c_v(2\gamma-\alpha)}{\beta(\gamma-\alpha)}\right)^{\frac{\beta}{\delta}} \left(\frac{\gamma c_c}{\alpha}\right)^{1-\frac{\beta}{\delta}},$$

$$E_{\delta c} = \frac{1}{T} e^{\frac{2(\beta\gamma+(\alpha-\gamma)\delta)}{\gamma\delta}} \left(\frac{\delta^2 c_v(2\gamma-\alpha)}{\beta(\gamma-\alpha)}\right)^{1-\frac{\alpha}{\gamma}} \left(\frac{\gamma c_c}{\alpha}\right)^{\frac{\alpha}{\gamma}}. \quad \blacksquare$$

Proof of Lemma 5

This result is obtained by substituting the equilibrium effort levels and the optimal payment term presented in Lemmas 1 and 2 into the objective functions of the client and the vendor in the effort dependent model. ■

Proof of Proposition 3

Proof follows from the derivatives of equilibrium effort levels (when the payment term is set to its optimal level $k\frac{\beta}{\gamma}$) with respect to the corresponding parameters. ■

Proof of Proposition 4

The proof follows from the derivatives of equilibrium effort levels (when the payment term is set to its optimal level $k\frac{\beta}{\gamma}$) with respect to the corresponding parameters. The threshold values are:

$$O_{\alpha c} = \frac{1}{T-t} e^{\frac{(\delta-\beta)(\beta\gamma+(\alpha-\gamma)\delta)}{\alpha\delta^2}} \left(\frac{\gamma\delta c_c}{(\delta-\beta)\alpha}\right)^{1-\frac{\beta}{\delta}} \left(\frac{\delta^2 c_v}{\beta^2}\right)^{\frac{\beta}{\delta}},$$

$$O_{\alpha v} = \frac{1}{T-t} e^{\frac{\beta\gamma+(\alpha-\gamma)\delta}{\gamma\delta}} \left(\frac{\gamma\delta c_c}{(\delta-\beta)\alpha}\right)^{1-\frac{\beta}{\delta}} \left(\frac{\delta^2 c_v}{\beta^2}\right)^{\frac{\beta}{\delta}},$$

$$O_{\beta c} = \frac{1}{T-t} e^{\frac{\beta\gamma+(\alpha-\gamma)\delta}{\gamma\delta}} \left(\frac{\gamma\delta c_c}{(\delta-\beta)\alpha}\right)^{\frac{\alpha}{\gamma}} \left(\frac{\delta^2 c_v}{\beta^2}\right)^{1-\frac{\alpha}{\gamma}},$$

$$O_{\beta v} = \frac{1}{T-t} e^{\frac{(\beta\gamma+(\alpha-\gamma)\delta)(\alpha(\beta-2\delta)+2\gamma(\delta-\beta))}{\beta\gamma^2(\delta-\beta)}} \left(\frac{\gamma\delta c_c}{(\delta-\beta)\alpha}\right)^{\frac{\alpha}{\gamma}} \left(\frac{\delta^2 c_v}{\beta^2}\right)^{1-\frac{\alpha}{\gamma}},$$

$$O_{\gamma c} = \frac{1}{T-t} e^{\frac{\beta\gamma+(\alpha-\gamma)\delta}{\gamma\delta}} \left(\frac{\gamma\delta c_c}{(\delta-\beta)\alpha}\right)^{1-\frac{\beta}{\delta}} \left(\frac{\delta^2 c_v}{\beta^2}\right)^{\frac{\beta}{\delta}},$$

$$O_{\delta c} = \frac{1}{T-t} e^{\frac{\beta\gamma+(\alpha-\gamma)\delta}{\gamma\delta}} \left(\frac{\gamma\delta c_c}{(\delta-\beta)\alpha}\right)^{\frac{\alpha}{\gamma}} \left(\frac{\delta^2 c_v}{\beta^2}\right)^{1-\frac{\alpha}{\gamma}},$$

$$O_{\delta v} = \frac{1}{T-t} e^{\frac{(\beta\gamma+(\alpha-\gamma)\delta)(2\beta\gamma+2\alpha\delta-2\gamma\delta-\alpha\beta)}{\delta\gamma(\gamma-\alpha)(\beta-\delta)}} \left(\frac{\gamma\delta c_c}{(\delta-\beta)\alpha}\right)^{\frac{\alpha}{\gamma}} \left(\frac{\delta^2 c_v}{\beta^2}\right)^{1-\frac{\alpha}{\gamma}}. \quad \blacksquare$$

Proof of Lemma 6

This result is obtained by substituting the equilibrium effort levels and the optimal payment term presented in Lemmas 3 and 4 into the objective functions of the client and the vendor in the output dependent model. ■

Proof of Proposition 5

One can derive this result by solving the output dependent payment model with $\alpha = 0$, and checking the sensitivity of the client's value with respect to β. ■

Proof of Lemmas 7, 8, 9, and 10

These proofs are similar to the proofs of Lemmas 1, 2, and 3. ■

APPENDIX B

PROOFS OF LEMMAS AND PROPOSITIONS IN SECTION 3

Proof of Lemma 11

The Hamiltonian for the optimal control problem can be written as below (Arrow and Kurz 1970):

$$H(t) = kq(t) - c_c u(t)^\gamma - c_v v(t)^\delta + \lambda(t)u(t)^\alpha v(t)^\beta.$$

Control variables $u(t)$ and $v(t)$ that are presented in the lemma are derived from solving the following set of equations.

$$\begin{aligned}
\frac{\partial H(t)}{\partial u(t)} &= -\gamma c_c u(t)^{\gamma-1} + \alpha u(t)^{\alpha-1} v(t)^\beta \lambda(t) = 0,\\
\frac{\partial H(t)}{\partial v(t)} &= -\delta c_v v(t)^{\delta-1} + \beta u(t)^\alpha v(t)^{\beta-1} \lambda(t) = 0,\\
\dot{\lambda}(t) &= -\frac{\partial H(t)}{\partial q(t)} = -k, \quad \text{and}\\
\lambda(T) &= 0.
\end{aligned}$$

Finally, since the Hamiltonian is strictly concave in u and v, the solution presented in the lemma is global maximum. ■

Proof of Lemma 12

Total value is calculated by substituting the optimal effort levels presented in Lemma 11 into the objective function presented in Equation (3.3). ■

Remark 2

Remark 2 *The following condition satisfies:* $(\gamma - \alpha)\delta - \beta\gamma > 0$.

Proof: $(\gamma-\alpha)\delta-\beta\gamma=\gamma\delta-\alpha\delta-\beta\gamma>\gamma\delta-(1-\beta)\delta-\beta\gamma\geq\gamma\delta-\max(\gamma,\delta)>0.$■

Proof of Lemma 13

The Hamiltonians for the client and the vendor (i.e., $H_c(t)$ and $H_v(t)$, respectively) can be written as below (Arrow and Kurz 1970):

$$H_c(t)=kq(t)-c_cu(t)^{\gamma}-lq(t)-F_d+\lambda_1(t)u(t)^{\alpha}v(t)^{\beta},\quad \text{and}$$
$$H_v(t)=lq(t)+F_d-c_vv(t)^{\delta}+\lambda_2(t)u(t)^{\alpha}v(t)^{\beta}.$$

The multipliers $\lambda_1(t)$ and $\lambda_2(t)$ are given by:

$$\dot{\lambda_1}(t)=-\frac{\partial H_c(t)}{\partial q(t)}=-k+l,\ \text{and}\ \dot{\lambda_2}(t)=-\frac{\partial H_v(t)}{\partial q(t)}=-l.$$

Solution of the above differential equations with the boundary conditions $\lambda_1(T)=0$ and $\lambda_2(T)=0$ results in $\lambda_1(t)=(k-l)(T-t)$, and $\lambda_2(t)=l(T-t)$. The equilibrium behavior necessitates the derivatives of the Hamiltonians with respect to the control variables to be zero. Hence, I need to solve the following system of equations

$$\frac{\partial H_c(t)}{\partial u(t)}=0,\ \text{and}\ \frac{\partial H_v(t)}{\partial v(t)}=0.$$

Thus, after substituting the values of $\lambda_1(t)$ and $\lambda_2(t)$, I have:

$$-\gamma c_cu(t)^{\gamma-1}+(k-l)(T-t)\alpha u(t)^{\alpha-1}v(t)^{\beta}=0,\quad \text{and}$$
$$-\delta c_vv(t)^{-1+\delta}+l(T-t)\beta u(t)^{\alpha}v(t)^{\beta-1}=0.$$

The solution of the above equation system results in the presented equilibrium levels. The Second order conditions for this differential game are also satisfied, because the Hamiltonians are strictly concave. ■

Proof of Lemma 14

The client does not leave any value to the vendor above his reservation utility. Hence, $F_d = -\int_0^T lq(t)dt + \int_0^T c_v v(t)^\delta dt + R_v$. After I substitute F_d and the equilibrium effort levels into the maximization problem of the client, and take the integrals, I have the following objective after rearranging the terms:

$$\max_l T^{\frac{(2\gamma-\alpha)\delta-\beta\gamma}{(\gamma-\alpha)\delta-\beta\gamma}} \frac{(\gamma-\alpha)\delta-\beta\gamma}{(2\gamma-\alpha)\delta-\beta\gamma} \left(\left(\frac{(k-l)\alpha}{\gamma c_c} \right)^{\alpha\delta} \left(\frac{l\beta}{\delta c_v} \right)^{\beta\gamma} \right)^{\frac{1}{(\gamma-\alpha)\delta-\beta\gamma}} \left(k - \frac{(k-l)\alpha}{\gamma} - \frac{l\beta}{\delta} \right) - R_v.$$

Given this parametric equation that is observed in the equilibrium, the client can maximize over the transfer payment per output parameter, i.e., l. First order condition w.r.t. l reveals that there are two critical values, i.e., $k\frac{\beta(\gamma-\alpha)-\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}}{\beta\gamma-\alpha\delta}$ and $k\frac{\beta(\gamma-\alpha)+\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}}{\beta\gamma-\alpha\delta}$. I show below that the first root is between 0 and k.

Assume first that $\beta\gamma - \alpha\delta < 0$. Then, $k\frac{\beta(\gamma-\alpha)-\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}}{\beta\gamma-\alpha\delta} < k$ iff $\alpha\beta - \beta\gamma + \sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)} < \alpha\delta - \beta\gamma$. This inequality further reduces to the condition $\beta(\gamma-\alpha) < \alpha(\delta-\beta)$ that is satisfied because of the assumption $\beta\gamma - \alpha\delta < 0$. On the other hand, $k\frac{\beta(\gamma-\alpha)-\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}}{\beta\gamma-\alpha\delta} > 0$ iff $\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)} > \beta(\gamma-\alpha)$. This inequality further reduces to the condition $\alpha(\delta-\beta) > \beta(\gamma-\alpha)$ that is satisfied due to the assumption $\beta\gamma - \alpha\delta < 0$. Therefore, under the assumption that $\beta\gamma - \alpha\delta < 0$, the first root is between 0 and k. If $\beta\gamma - \alpha\delta > 0$, then the proof that shows the first root is between 0 and k is almost identical. Hence, I omit that part due to brevity.

In a similar manner, the second root can be shown to be either less than 0 or more than k. In addition, the second order condition reveals that the first root, i.e., $l = k\frac{\beta(\gamma-\alpha)-\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}}{\beta\gamma-\alpha\delta}$, is a local maximum. The upper and lower bounds for l are obviously k and zero. Besides, I can disregard the reservation utility constraint for the vendor, because it is always satisfied. Therefore, the constraint set is compact. Hence, due to Weierstrass Theorem (Patrick 2009), the reported l value is the global maximum. ■

Proof of Proposition 6

I provide the proof for $\frac{dl^*}{d\alpha} < 0$. Other proofs are similar, hence omitted. First observe that

$(\beta\gamma - 2\alpha\beta + \alpha\delta)^2 - 4\alpha\beta(\gamma - \alpha)(\delta - \beta) = (\beta\gamma - \alpha\delta)^2 > 0$. Hence,

$(\beta\gamma - 2\alpha\beta + \alpha\delta)^2 > 4\alpha\beta(\gamma - \alpha)(\delta - \beta)$.

Positive root of $(\beta\gamma - 2\alpha\beta + \alpha\delta)^2 = 0$ is $\beta\gamma - 2\alpha\beta + \alpha\delta$, because $\beta\gamma - 2\alpha\beta + \alpha\delta > \beta^2 - 2\alpha\beta + \alpha^2 = (\alpha - \beta)^2 \geq 0$. Therefore, I have:

$$\beta\gamma - 2\alpha\beta + \alpha\delta > 2\sqrt{\alpha\beta(\gamma - \alpha)(\delta - \beta)}. \tag{5.1}$$

I use this inequality to show that $\frac{dl^*}{d\alpha} < 0$. Now, let us write
$\frac{dl^*}{d\alpha} = \frac{k\beta\gamma(\beta-\delta)\left(\beta\gamma-2\alpha\beta+\alpha\delta-2\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}\right)}{2\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}(\beta\gamma-\alpha\delta)^2}$.
Here, $\frac{k\beta\gamma(\beta-\delta)}{2\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}(\beta\gamma-\alpha\delta)^2} < 0$, because $\alpha < \gamma$ and $\beta < \delta$. Therefore $\frac{dl^*}{d\alpha} < 0$ iff
$-\left(\beta\gamma - 2\alpha\beta + \alpha\delta - 2\sqrt{\alpha\beta(\gamma - \alpha)(\delta - \beta)}\right) < 0$.
Hence, due to inequality (5.1), I conclude that $\frac{dl^*}{d\alpha} < 0$. ■

Proof and Threshold Values of Proposition 7

The equilibrium effort levels with the optimal payment parameter (i.e., $l^* = k\frac{\beta(\gamma-\alpha)-\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}}{\beta\gamma-\alpha\delta}$) is presented below:

$$u(t) = \left(\frac{(T-t)^{\delta}\left(\frac{k\alpha\left(\alpha(\beta-\delta)+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\gamma(\beta\gamma-\alpha\delta)c_C}\right)^{\delta-\beta}}{\left(\frac{k\beta\left(\alpha\beta-\beta\gamma+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\delta(\alpha\delta-\beta\gamma)c_V}\right)^{-\beta}}\right)^{\frac{1}{(\gamma-\alpha)\delta-\beta\gamma}},$$

$$v(t) = \left(\frac{(T-t)^{\gamma}\left(\frac{k\alpha\left(\alpha(\beta-\delta)+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\gamma(\beta\gamma-\alpha\delta)c_C}\right)^{\alpha}}{\left(\frac{k\beta\left(\alpha\beta-\beta\gamma+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\delta(\alpha\delta-\beta\gamma)c_V}\right)^{-\gamma+\alpha}}\right)^{\frac{1}{(\gamma-\alpha)\delta-\beta\gamma}}.$$

The threshold values presented below follow from the derivatives of these effort levels with respect to the corresponding parameters.

$$O_{\alpha c} = \frac{1}{T-t}\frac{e^{\frac{((\gamma-\alpha)\delta-\beta\gamma)\left(2\alpha^2(\beta-\delta)\delta+\gamma\left(3\beta\gamma(\beta-\delta)+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\delta\right)+2\alpha\gamma\left(\delta^2-\beta^2\right)\right)}{2\alpha\delta^2(\alpha-\gamma)(\alpha\delta-\beta\gamma)}}}{\left(\frac{\alpha\left(\alpha(\beta-\delta)+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\gamma(\beta\gamma-\alpha\delta)c_c}\right)^{\frac{\delta-\beta}{\delta}}\left(\frac{\beta\left(\beta(\alpha-\gamma)+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\delta(\alpha\delta-\beta\gamma)c_v}\right)^{\frac{\beta}{\delta}}},$$

$$O_{\alpha v} = \frac{1}{T-t}\frac{\frac{e^{\frac{(\beta\gamma+(\alpha-\gamma)\delta)\left(2\alpha^3\delta-\gamma^2\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}+\alpha\gamma^2(2\beta+\delta)-\alpha^2\gamma(2\beta+3\delta)\right)}{2\delta\gamma\ \alpha(\alpha-\gamma)(-\beta\gamma+\alpha\delta)}}}{\left(\frac{\beta\left(\alpha\beta-\beta\gamma+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\delta(-\beta\gamma+\alpha\delta)c_v}\right)^{\frac{\gamma-\alpha}{\gamma}}}}{\left(\frac{\alpha\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\delta c_v}{\beta^2(\gamma-\alpha)\gamma c_c}\right)^{\frac{\beta\gamma+(\alpha-\gamma)\delta}{-\delta\gamma}}\left(\frac{\alpha\left(\alpha\beta-\alpha\delta+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\gamma(\beta\gamma-\alpha\delta)c_c}\right)^{\frac{\alpha}{\gamma}}},$$

$$O_{\beta c} = \frac{1}{T-t}\frac{\frac{e^{\frac{(\beta\gamma+(\alpha-\gamma)\delta)\left(2\beta^3\gamma-\beta^2(2\alpha+3\gamma)\delta+\beta(2\alpha+\gamma)\delta^2-\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\delta^2\right)}{2\beta(\beta-\delta)(\beta\gamma-\alpha\delta)\gamma\delta}}}{\left(\frac{\alpha\sqrt{\alpha\beta(\gamma-\alpha)(\delta-\beta)}\delta c_v}{\beta^2(\gamma-\alpha)\gamma c_c}\right)^{\frac{\beta\gamma+(\alpha-\gamma)\delta}{\gamma\delta}}}}{\left(\frac{\alpha\left(\alpha\beta-\alpha\delta+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\gamma(\beta\gamma-\alpha\delta)c_c}\right)^{\frac{\delta-\beta}{\delta}}\left(\frac{\beta\left(\alpha\beta-\beta\gamma+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\delta(\alpha\delta-\beta\gamma)c_v}\right)^{\frac{\beta}{\delta}}},$$

$$O_{\beta v} = \frac{1}{T-t}\frac{e^{\frac{(\beta\gamma+(\alpha-\gamma)\delta)\left(2\beta^2\gamma(-\alpha+\gamma)+2\beta(\alpha-\gamma)(\alpha+\gamma)\delta-\gamma\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\delta+3\alpha(\gamma-\alpha)\delta^2\right)}{2\beta\gamma^2(\beta-\delta)(\beta\gamma-\alpha\delta)}}}{\left(\frac{\alpha\left(\alpha(\beta-\delta)+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\gamma(\beta\gamma-\alpha\delta)c_c}\right)^{\frac{\alpha}{\gamma}}\left(\frac{\beta\left(\alpha\beta-\beta\gamma+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\delta(-\beta\gamma+\alpha\delta)c_v}\right)^{\frac{\gamma-\alpha}{\gamma}}},$$

$$O_{\gamma c} = \frac{1}{T-t}\frac{e^{\frac{(\beta\gamma+(\alpha-\gamma)\delta)\left(2\alpha^2(\beta-\delta)\delta+\gamma\left(3\beta\gamma(\beta-\delta)+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\delta\right)+2\alpha\gamma\left(\delta^2-\beta^2\right)\right)}{2(\alpha-\gamma)\gamma\delta(\beta-\delta)(-\beta\gamma+\alpha\delta)}}}{\left(\frac{\alpha\left(\alpha(\beta-\delta)+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\gamma(\beta\gamma-\alpha\delta)c_c}\right)^{\frac{\delta-\beta}{\delta}}\left(\frac{\beta\left(\alpha\beta-\beta\gamma+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\delta(-\beta\gamma+\alpha\delta)c_v}\right)^{\frac{\beta}{\delta}}},$$

$$O_{\gamma v} = \frac{1}{T-t}\frac{\frac{e^{\frac{(\beta\gamma+(\alpha-\gamma)\delta)\left(-\gamma^2\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}+2\alpha^3\delta+\alpha\gamma^2(2\beta+\delta)-\alpha^2\gamma(2\beta+3\delta)\right)}{2(\alpha-\gamma)\alpha\delta\gamma(-\beta\gamma+\alpha\delta)}}}{\left(\frac{\alpha\left(\alpha\beta+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}-\alpha\delta\right)}{\gamma(\beta\gamma-\alpha\delta)c_c}\right)^{\frac{\alpha(\delta-\beta)}{\alpha\delta}}}}{\left(\frac{\beta\left(\alpha\beta-\beta\gamma+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\delta(-\beta\gamma+\alpha\delta)c_v}\right)^{\frac{(\delta-\beta)(\gamma-\alpha)}{\alpha\delta}}\left(\frac{\beta\left(\alpha\beta-\beta\gamma+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\delta(-\beta\gamma+\alpha\delta)c_v}\right)^{\frac{(\beta\gamma+(\alpha-\gamma)\delta)}{\alpha\delta}}},$$

$$O_{\delta c} = \frac{1}{T-t} \frac{\frac{e^{\frac{(\beta\gamma+(\alpha-\gamma)\delta)\left(2\beta^3\gamma-\beta^2(2\alpha+3\gamma)\delta+\beta(2\alpha+\gamma)\delta^2-\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\delta^2\right)}{2\beta\gamma(\beta-\delta)\delta(\beta\gamma-\alpha\delta)}}}{\left(\frac{\alpha\left(\alpha(\beta-\delta)+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\gamma(\beta\gamma-\alpha\delta)c_c}\right)^{\frac{\beta\gamma+(\alpha-\gamma)\delta}{\beta\gamma}}}}{\left(\frac{\alpha\left(\alpha(\beta-\delta)+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\gamma(\beta\gamma-\alpha\delta)c_c}\right)^{\frac{(\gamma-\alpha)(\delta-\beta)}{\beta\gamma}} \left(\frac{\beta\left(\alpha\beta-\beta\gamma+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\delta(\alpha\delta-\beta\gamma)c_v}\right)^{\frac{\gamma-\alpha}{\gamma}}},$$

$$O_{\delta v} = \frac{1}{T-t} \frac{e^{\frac{(\beta\gamma+(\alpha-\gamma)\delta)\left(2\beta^2\gamma(\gamma-\alpha)+2\beta(\alpha-\gamma)(\alpha+\gamma)\delta-\gamma\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\delta+3\alpha(\gamma-\alpha)\delta^2\right)}{2(\alpha-\gamma)(\beta-\delta)\delta\gamma(-\beta\gamma+\alpha\delta)}}}{\left(\frac{\alpha\left(\alpha(\beta-\delta)+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\gamma(\beta\gamma-\alpha\delta)c_c}\right)^{\frac{\alpha}{\gamma}} \left(\frac{\beta\left(\alpha\beta-\beta\gamma+\sqrt{\alpha\beta(\alpha-\gamma)(\beta-\delta)}\right)}{\delta(-\beta\gamma+\alpha\delta)c_v}\right)^{\frac{\gamma-\alpha}{\gamma}}}. \qquad \blacksquare$$

Proof of Lemma 15

Client's value is derived by substituting the equilibrium effort levels and the optimal payment terms (presented in Lemmas 13 and 14, respectively) into the objective function of the client. ■

Proof of Proposition 8

In this contract, the client covers all the costs due to exerting efforts and the monitoring of vendor's effort. However, the client leaves no value in addition to these costs and the reservation utility of the vendor. ■

Proof of Lemma 16

The equilibrium effort levels are derived by solving the optimal control problem **Problem P**. The proof is similar to the earlier proofs, and therefore the details are omitted. ■

Proof of Proposition 9

These results follow from the derivatives of equilibrium effort levels (when the payment term is set to its optimal level) with respect to the corresponding parameters. ■

Proof and Threshold Values of Proposition 10

These results follow from the derivatives of equilibrium effort levels with respect to the corresponding parameters. The threshold values are:

$$E_{\alpha c} = e^{\frac{-(\delta-\beta)((\gamma-\alpha)\delta-\beta\gamma)}{\alpha\delta^2}} \frac{\gamma c_c}{(T-t)\alpha} \left(\frac{\alpha\delta c_v}{\beta\gamma c_c}\right)^{\frac{\beta}{\delta}},$$

$$E_{\alpha v} = e^{\frac{-((\delta-\beta)\gamma-\alpha\delta)}{\gamma\delta}} \frac{\delta c_v}{(T-t)\beta} \left(\frac{\beta\gamma c_c}{\alpha\delta c_v}\right)^{\frac{\delta-\beta}{\delta}},$$

$$E_{\beta c} = e^{\frac{-((\delta-\beta)\gamma-\alpha\delta)}{\gamma\delta}} \frac{\gamma c_c}{(T-t)\alpha} \left(\frac{\alpha\delta c_v}{\beta\gamma c_c}\right)^{\frac{\gamma-\alpha}{\gamma}},$$

$$E_{\beta v} = e^{\frac{-(\gamma-\alpha)((\gamma-\alpha)\delta-\beta\gamma)}{\beta\gamma^2}} \frac{\delta c_v}{(T-t)\beta} \left(\frac{\beta\gamma c_c}{\alpha\delta c_v}\right)^{\frac{\alpha}{\gamma}},$$

$$E_{\gamma c} = e^{\frac{-((\delta-\beta)\gamma-\alpha\delta)}{\gamma\delta}} \frac{\gamma c_c}{(T-t)\alpha} \left(\frac{\alpha\delta c_v}{\beta\gamma c_c}\right)^{\frac{\beta}{\delta}},$$

$$E_{\gamma v} = e^{\frac{-(\gamma(\delta-\beta)-\alpha\delta)}{\gamma\delta}} \frac{\delta c_v}{(T-t)\beta} \left(\frac{\beta\gamma c_c}{\alpha\delta c_v}\right)^{\frac{\delta-\beta}{\delta}},$$

$$E_{\delta c} = e^{\frac{-((\delta-\beta)\gamma-\alpha\delta)}{\gamma\delta}} \frac{\gamma c_c}{(T-t)\alpha} \left(\frac{\alpha\delta c_v}{\beta\gamma c_c}\right)^{\frac{\gamma-\alpha}{\gamma}},$$

$$E_{\delta v} = e^{\frac{-((\gamma-\alpha)\delta-\beta\gamma)}{\gamma\delta}} \frac{\delta c_v}{(T-t)\beta} \left(\frac{\beta\gamma c_c}{\alpha\delta c_v}\right)^{\frac{\alpha}{\gamma}}. \qquad \blacksquare$$

Proof of Lemma 17

The value is derived by substituting the equilibrium effort levels in Lemma 16 and the function $H(.)$ in Proposition 8 into the objective function of the client. ■

Proof of Proposition 11

Here I show the proof for the effort dependent contract, and for only with respect to α. Other parts of the proof are similar, and therefore I omit them. It is easy to show that, when the vendor's effort is monitored, client's net value presented in Lemma 17 increases in α if and only if:

$$\gamma\delta(\beta\gamma+(\alpha-\gamma)\delta)+(\beta\gamma+(\alpha-2\gamma)\delta)\ln\left[(kT)^{-\gamma\delta}\left(\frac{\alpha}{\gamma c_c}\right)^{\beta\gamma-\gamma\delta}\left(\frac{\beta}{\delta\left(c_v+c_m\right)}\right)^{-\beta\gamma}\right]>0.$$

This condition can be rewritten as:

$$\left(\frac{\alpha}{\gamma c_c}\right)^{\delta-\beta}(kT)^{\delta}\left(\frac{\beta}{\delta\left(c_v+c_m\right)}\right)^{\beta}>e^{\delta\frac{(\gamma-\alpha)\delta-\beta\gamma}{(2\gamma-\alpha)\delta-\beta\gamma}}. \tag{5.2}$$

Hence, it is apparent that left side of the inequality $\left(\frac{\alpha}{\gamma c_c}\right)^{\delta-\beta}(kT)^{\delta}\left(\frac{\beta}{\delta(c_v+c_m)}\right)^{\beta}\rightarrow 0$ as $\alpha\rightarrow 0$. However, right hand side $e^{\delta\frac{(\gamma-\alpha)\delta-\beta\gamma}{(2\gamma-\alpha)\delta-\beta\gamma}}\rightarrow e^{\delta\frac{\gamma(\delta-\beta)}{\gamma(2\delta-\beta)}}>0$ as $\alpha\rightarrow 0$. Therefore with low values of α, Inequality (5.2) would not satisfy. This implies that the client's net value decreases in α when α is low. ■

Proof of Proposition 12

The condition in this proposition is derived by comparing the client's value in both contracts that are presented in Lemmas 15 and 17. ■

Proofs of Lemmas 18, 19, and 20

These proofs are similar to the proofs of Lemmas 14, 15, and 16. Hence, I omit them due to brevity.

■

Proof of Proposition 13

I derive this result by a direct comparison of the expressions in Lemmas 14 and 18. ■

APPENDIX C

LINEARIZED MODEL, AND PROOFS OF LEMMAS AND PROPOSITIONS IN SECTION 4

Linearized Model

For linearization, I replace constraint sets (4.14) and (4.15) by the following sets of constraints when $R_{max} = 5$. Similar constraints can be easily written for any other value of R_{max}.

$$
\begin{aligned}
b_{ijt} + 2c_{ijt} + 4q_{ijt} &\leq R_{ij} + (R_{max} - R_{ij})\, x_{ijt}E + 0.5;\ i \in U,\ j \in J,\ t = 1, 2, \ldots, T \\
b_{ijt} + 2c_{ijt} + 4q_{ijt} &\geq R_{ij} + (R_{max} - R_{ij})\, x_{ijt}E - 0.5;\ i \in U,\ j \in J,\ t = 1, 2, \ldots, T \\
b_{ijt} + 2c_{ijt} + 4q_{ijt} &\leq R_{max};\ i \in U,\ j \in J,\ t = 1, 2, \ldots, T \\
\sum_{s \in J} f_{ijst} + 2\sum_{s \in J} g_{ijst} &+ 4\sum_{s \in J} h_{ijst} - b_{ijt} - 2c_{ijt} - 4q_{ijt} = 0;\ i \in U,\ j \in J,\ \forall t \\
f_{ijst} &\leq b_{ist};\ i \in U,\ j \in J, s \in J,\ t = 1, 2, \ldots, T \\
b_{ist} + A_{ijt} - 1 &\leq f_{ijst} \leq A_{ijt};\ i \in U,\ j \in J, s \in J,\ t = 1, 2, \ldots, T \\
g_{ijst} &\leq c_{ist};\ i \in U,\ j \in J, s \in J,\ t = 1, 2, \ldots, T \\
c_{ist} + A_{ijt} - 1 &\leq g_{ijst} \leq A_{ijt};\ i \in U,\ j \in J, s \in J,\ t = 1, 2, \ldots, T \\
h_{ijst} &\leq q_{ist};\ i \in U,\ j \in J, s \in J,\ t = 1, 2, \ldots, T \\
q_{ist} + A_{ijt} - 1 &\leq h_{ijst} \leq A_{ijt};\ i \in U,\ j \in J, s \in J,\ t = 1, 2, \ldots, T \\
b_{ijt}, c_{ijt}, q_{ijt} &\in \{0, 1\};\ i \in U,\ j \in J,\ t = 1, 2, \ldots, T
\end{aligned}
$$

Proof of Proposition 14

I first prove the proposition for a single period and then extend it to the multiple period case. Hence, I omit the time index in the discussion below.

Sufficiency: *If at least one of the conditions in parts (a), (b), or (c) is satisfied, then the current industry practice is optimal.*

In both parts (a) and (b), the left hand side indicates the expected demand for a movie cluster after the recommendations are made using the all-inclusive policy. Hence, the condition in part (a) implies that expected demand for all of movie clusters is greater than or equal to their corresponding available inventories after the recommendations. Therefore, the firm is utilizing its entire inventory using the all-inclusive policy and cannot satisfy more demand in any other policy. Hence, the all-inclusive policy is optimal. On the other hand, the condition in part (b) implies that the expected demand for the movie clusters that are in the set of recommendable movie clusters for at least one user cluster (movie clusters that have at least one pair above the threshold) is below or equal to their corresponding available inventories after the recommendations. Hence, in this case, all-inclusive policy satisfies the demand for such movie clusters. On the other hand, because all of the movies above the threshold are recommended, the demands for the movies for which all rating values are below the threshold level are minimized. Hence, although the demands for these movie clusters cannot be all satisfied, they are at their minimum anyways. Therefore, the all-inclusive policy is again optimal. In part (c), no recommendations can be made in any policy. Hence, every policy are the same and optimal. Extending the argument for the multiple period case is now easy. In each period, if either the entire inventory is shipped or all of the demands are satisfied as much as possible using the all-inclusive policy, then this policy is optimal. Hence, the conditions are sufficient in the multiple period case as well.

Necessity: *If the all-inclusive policy is optimal, then at least one of the conditions in parts (a), (b), or (c) is satisfied.*

For this part of the proof, I use contraposition. First, I divide the set of movie clusters into three subsets B, C, and D as follows: (i) Set B contains those movie clusters for which expected demand is more than inventory after the recommendations are made using the all-inclusive policy, (ii) Set C contains those movie clusters for which expected demand is less than inventory after the recommendations are made using the all-inclusive policy, and (iii) Set D contains those movie clusters for which expected demand is equal to inventory after the recommendations are made using the all-inclusive policy. Then, the only condition not covered in parts (a), (b), and (c) of the proposition is the following:

i. $|B| \geq 1$ and $\exists(\gamma, b) \in V : \gamma \in U; b \in B$.

ii. $|C| \geq 1$.

These conditions state that (i) the set B is not empty, and at least one movie cluster in this set is recommendable to at least one user cluster, and (ii) there is at least one movie cluster in set C. If the all-inclusive policy is used to make the recommendations, then the demand for a movie cluster $j \in J$ can be written as

$$\sum_{i \in U:(i,j) \in V} \frac{R_{ij} + (R_{max} - R_{ij})Ex_{ij}}{\alpha + \beta + \eta} n_i + \sum_{i \in U:(i,j) \notin V} \frac{R_{ij}}{\alpha + \beta + \eta} n_i, \tag{5.3}$$

where $\alpha = \sum_{p \in B:(i,p) \in V} (R_{ip} + (R_{max} - R_{ip})Ex_{ip}) + \sum_{p \in B:(i,p) \notin V} R_{ip}$,

$\beta = \sum_{q \in C:(i,q) \in V} (R_{iq} + (R_{max} - R_{iq})Ex_{iq}) + \sum_{q \in C:(i,q) \notin V} R_{iq}$,

$\eta = \sum_{v \in D:(i,v) \in V} (R_{iv} + (R_{max} - R_{iv})Ex_{iv}) + \sum_{v \in D:(i,v) \notin V} R_{iv}$.

Consider a policy where the recommendations for the movie clusters that belong to set B and that have at least one (i, j) pair in set V is lowered by a positive amount ϵ, i.e., $x_{ib} = 1 - \epsilon$, $b \in B$, $\forall \{i, b\} \in V$; and the recommendations for the other movie clusters that belong to sets C and D are not changed. Clearly, the value of α will decrease with ϵ unless $R_{ib} = R_{max}$. Since R_{ib} is defined in the continuous domain and it represents an aggregate measure, it is reasonable to consider that R_{ib} values are at least marginally smaller than R_{max}. In summary, one can state that the value of α will decrease with ϵ, whereas the values of β and η will remain unchanged. Therefore, the denominator in Equation (5.3) will decrease with ϵ.

Furthermore, the numerator of Equation (5.3) will remain unchanged with ϵ for the movie clusters in sets C and D. Therefore, the demand for movie clusters in sets C and D will increase with ϵ. It is also easy to see that the demand for movie clusters in set B will decrease with ϵ. Therefore, there exists an ϵ such that the demand for the movie clusters in set B is decreased no less than their corresponding inventory level. In this scenario, a portion of the demand for the movie clusters in Set B is shifted to sets C and D. The unsatisfied demand shifted from Set B to Set D is again not satisfied, however, at least some portion of the demand shifted to Set C is satisfied. Therefore, total dissatisfaction is less in this policy. Hence, I have shown that the presented conditions are necessary and sufficient for the optimality of the all-inclusive policy in the single period problem.

Now let us consider the multiple period case for this part of the proof. The proof for the necessity part again follows from contraposition. Suppose, in the all-inclusive policy, neither of the conditions presented in the proposition is satisfied for at least one time period, say this period is s. The proof so far implies that another recommendation policy, say new-policy, is optimal in period s. Hence, if utilized, the new-policy reduces the total dissatisfaction in period s by a positive amount, say Δz. Therefore, the all-inclusive policy saves some inventory in the amount of Δz in period s compared to this new policy. However, this same amount, i.e., Δz, is the additional number of customers not satisfied in period s compared to the new-policy. Out of the extra shipments made with the new-policy, there will be returns next period by the amount of $Q_{ss}\Delta z$. Therefore, because of the returns, the available inventory in the all-inclusive policy is more than its counterpart in the new policy by only $(1 - Q_{ss})\Delta z$ in period $s + 1$. Thus, if $s = 1, 2, \ldots, T - 1$, depending on the demand realization in period $s + 1$, the all-inclusive policy is inferior to the new policy in the range of Δz to $Q_{ss}\Delta z$. If $s = T$, then the all-inclusive policy is inferior to the new policy by Δz. Hence the new-policy is better than the all-inclusive policy, and therefore the all-inclusive policy is not optimal. This implies that the conditions presented in the proposition are necessary for each and every period. ■

Proof of Proposition 15

The upper bound presented in the proposition is actually the minimum of three different values. The first one is the total dissatisfaction when the recommendation threshold θ is set to R_{max} in which case nothing is recommended. This is based on the fact that dissatisfaction is a nonnegative number, i.e., the expected amount of demand that cannot be satisfied. Hence, no policy can decrease the dissatisfaction level below zero. Moreover, if the recommendation threshold is reduced, then some movie clusters enter into the set of recommendable movies to different user clusters. The second expression states that by being able to recommend a specific movie, the dissatisfaction can at most be decreased by the amount of corresponding unused inventory in the case when θ is set to R_{max}. Finally, for this specific movie, the maximum increase in demand for the users for which the recommendation can be made is calculated in the last term. ■

Proof of Proposition 16

This proof is similar to the proof of Proposition 15, and hence omitted. ■

Proof of Proposition 17

Part (a): I use an example to show the sub-optimality of $x_{ij} = 1\ \forall (i, j) \in V$ for a movie cluster j which has more inventory than its demand. In this example, $|J| = 2$, $|U| = 1$, $T = 1$, $n > K_1$, and R_1,$R_2 > \theta$. Hence, I can omit the indices for both user clusters and periods whenever it does not cause any confusion. Furthermore, the demand for movie 1 (resp., movie 2) is less (resp., more) than its available inventory before any recommendations are made. For an arbitrary instance of this problem, assume that $x_1 = 1$ (if $(1, 1) \in V$) is the optimal policy. However, I show later in Proposition 18 that the optimal x_1 is given by $x_1 = \min\left[\frac{(R_1+R_2)K_1-R_1n}{E(R_{max}-R_1)(n-K_1)}, 1\right]$. Therefore, when $\frac{(R_1+R_2)K_1-R_1n}{E(R_{max}-R_1)(n-K_1)} < 1$ is satisfied, the optimal x_1 is less than 1. This completes the proof for part (a).

Part (b): Again, I construct a counter-example to show the sub-optimality of $x_{ij} = 0\ \forall (i, j)$. In this example, $|J| = 3, |U| = 1, T = 1, n > (K_1 + K_2)$, and R_1, R_2, $R_3 > \theta$. I ignore the indices for user clusters and time periods in this proof whenever it does not cause any confusion. Moreover, the movie cluster 1 has surplus inventory, and movie clusters 2 and 3 have more demand than corresponding inventories before any recommendations are made. I set $x_2 = x_3 = 0$, and find the x_1 that makes the demand for movie cluster 1 equal to its inventory: $x_1 = \min\left[\frac{(R_1+R_2+R_3)K_1-R_1n}{E(R_{max}-R_1)(n-K_1)}, 1\right]$.

Now, consider the case where $\frac{(R_1+R_2+R_3)K_1-R_1n}{E(R_{max}-R_1)(n-K_1)} < 1$. Therefore, demand for movie cluster 1 is equal to K_1. By setting $x_1 = \frac{(R_1+R_2+R_3)K_1-R_1n}{E(R_{max}-R_1)(n-K_1)}$, $x_2 = 0$, $x_3 = 0$, demands for movie clusters 2 and 3 are $\frac{(n-K_1)R_2}{R_2+R_3}$ and $\frac{(n-K_1)R_3}{R_2+R_3}$, respectively. For optimality of this policy, one of the following two conditions should hold $\frac{(n-K_1)R_2}{R_2+R_3} \geq K_2$ and $\frac{(n-K_1)R_3}{R_2+R_3} \geq K_3$; or $\frac{(n-K_1)R_2}{R_2+R_3} \leq K_2$ and $\frac{(n-K_1)R_3}{R_2+R_3} \leq K_3$.

In the first case, demands for movie clusters 2 and 3 are more than their inventories. Hence, they are decreased as much as possible by shifting the demand to movie cluster 1 by shipping all K_1 units. In the second case, there are no more demand shortages for movie clusters 2 and 3. In both cases, the optimal is obtained by setting $x_1 = \frac{(R_1+R_2+R_3)K_1 - R_1 n}{E(R_{max}-R_1)(n-K_1)}$, and $x_2 = x_3 = 0$. However, consider a third case where $\frac{(n-K_1)R_2}{R_2+R_3} \leq K_2$ and $\frac{(n-K_1)R_3}{R_2+R_3} \geq K_3$. In this case, by recommending movie cluster 2 to some user clusters, the demand shortage of movie cluster 3 is decreased. The policy, if there is any, that will ship entire inventories of movie clusters 1 and 2 clearly performs better than the above policy. Hence, one needs to solve the following two equations simultaneously: $\frac{[R_1+(R_{max}-R_1)x_1 E]n}{R_1+R_2+R_3+(R_{max}-R_1)x_1 E+(R_{max}-R_2)x_2 E} = K_1$ and $\frac{[R_2+(R_{max}-R_2)x_2 E]n}{R_1+R_2+R_3+(R_{max}-R_1)x_1 E+(R_{max}-R_2)x_2 E} = K_2$. The solution of this system results in: $x_1 = \frac{K_1(R_1+R_3)-R_1(K_2-n)}{E(R_{max}-R_1)(n-K_1-K2)}$ and $x_2 = \frac{K_2(R_2+R_3)-R_2(K_1-n)}{E(R_{max}-R_1)(n-K_1-K2)}$.

Now consider the case where $K_1(R_1+R_3)+R_1(K_2-n) > 0$, and $K_2(R_2+R_3)+R_2(K_1-n) > 0$. The assumption at the beginning that $n > (K_1+K_2)$ ensures the non-negativity of x_1 and x_2. Moreover consider the case where $0 < \frac{K_1(R_1+R_3)-R_1(K_2-n)}{E(R_{max}-R_1)(n-K_1-K2)} < 1$, and $0 < \frac{K_2(R_2+R_3)-R_2(K_1-n)}{E(R_{max}-R_1)(n-K_1-K2)} < 1$. In this case, the optimal solution is: $x_1 = \frac{K_1(R_1+R_3)-R_1(K_2-n)}{E(R_{max}-R_1)(n-K_1-K2)}$, $x_2 = \frac{K_2(R_2+R_3)-R_2(K_1-n)}{E(R_{max}-R_1)(n-K_1-K2)}$, and $x_3 = 0$. Therefore, I find a contradicting example for the optimality of the statement in part (b). ■

Proof of Proposition 18

Part (a): If expected demand for both of the movies are less than the respective inventories before any recommendations are made, i.e., $\frac{R_1 n}{R_1+R_2} - K_1 \leq 0$ and $\frac{R_2 n}{R_1+R_2} - K_2 \leq 0$, then one of the optimal policies is to recommend nothing, i.e., $x_1 = x_2 = 0$, because all of the demand is already satisfied without any recommendations.

Part (b): When expected demand for both of the movie clusters are more than their inventories before any recommendations are made, i.e., $\frac{R_1 n}{R_1+R_2} - K_1 > 0$ and $\frac{R_2 n}{R_1+R_2} - K_2 > 0$, then it is again optimal to have $x_1 = x_2 = 0$, because the entire inventory is already utilized and no other policy can ship more DVDs.

Part (c): In this case, the optimal policy should maximize the demand for movie cluster 1 and minimize it for movie cluster 2. No policy can increase the demand for movie cluster 1 more than its inventory given that $n \leq K_1$. Hence, an optimal level is found by recommending movie cluster 1 to everybody and movie cluster 2 to nobody, i.e., $x_1 = 1$ and $x_2 = 0$.

Part (d): In this case, one should aim to utilize the surplus inventory of movie cluster 1 and equivalently reduce that of movie cluster 2. Demand for movie cluster 2, i.e., $A_2 n$, strictly increases in x_2, and therefore $x_2 = 0$. Substituting $x_2 = 0$ in the constraint sets (4.14) and (4.15) from the original formulation, one gets $A_1 = \frac{R_1+(R_{max}-R_1)x_1 E}{R_1+(R_{max}-R_1)x_1 E+R_2}$ and $A_2 = \frac{R_2}{R_1+(R_{max}-R_1)x_1 E+R_2}$. Now, substituting the above into the constraint set (4.11) gives $Z_1 \geq \frac{R_1+(R_{max}-R_1)x_1 E}{R_1+(R_{max}-R_1)x_1 E+R_2} n - K_1$ and $Z_2 \geq \frac{R_2}{R_1+(R_{max}-R_1)x_1 E+R_2} n - K_2$.

At the optimal level, it might not be best to set $x_1 = 1$, because the demand for movie cluster 1 may become more than its available inventory K_1. This may still be one of the optimal policies, however not in the case where Z_2 becomes 0. On the other hand, if I try to utilize movie cluster 1 as much as possible without making $(A_1 n_1 - K_1) \geq 0$, I have one of the optimal policies. Therefore, the value of x_1 that satisfies the demand balance equation $\frac{R_1+(R_{max}-R_1)x_1 E}{R_1+(R_{max}-R_1)x_1 E+R_2} n = K_1$ is given by $x_1 = \frac{(R_1+R_2)K_1 - R_1 n}{E(R_{max}-R_1)(n-K_1)}$.

The given conditions in part (d), i.e., $R_1 n_1/(R_1 + R_2) - K_1 \leq 0$, and $n_1 > K_1$ guarantee that x_1 is nonnegative. The above expression also gives the interval in which $x_1 = 1$. For that to happen, $\frac{(R_1+R_2)K_1 - R_1 n}{E(R_{max}-R_1)(n-K_1)} \geq 1$ must hold. It holds when $E \leq \frac{(R_1+R_2)K_1 - R_1 n}{(R_{max}-R_1)(n-K_1)}$. Therefore, if $E \leq \frac{(R_1+R_2)K_1 - R_1 n}{(R_{max}-R_1)(n-K_1)}$, then $x_1 = 1$, $x_2 = 0$; else if $E > \frac{(R_1+R_2)K_1 - R_1 n}{(R_{max}-R_1)(n-K_1)}$, then $x_1 = \frac{(R_1+R_2)K_1 - R_1 n}{E(R_{max}-R_1)(n-K_1)}$, $x_2 = 0$. ■

Proof of Proposition 19

This proof is similar to the proof of Proposition 18, and hence omitted. ■