Contracting for Collaborative Services

Guillaume Roels, Uday S. Karmarkar
UCLA Anderson School of Management, University of California, Los Angeles, Los Angeles, California 90095
{groels@anderson.ucla.edu, uday.karmarkar@anderson.ucla.edu}
Scott Carr
LECG, Washington, DC 20006, scarr@lecg.com

In this paper, we analyze the contracting issues that arise in collaborative services, such as consulting, financial planning, and information technology outsourcing. In particular, we investigate how the choice of contract type—among fixed-fee, time-and-materials, and performance-based contracts—is driven by the service environment characteristics. We find that fixed-fee contracts contingent on performance are preferred when the service output is more sensitive to the vendor’s effort, that time-and-materials contracts are optimal when the output is more sensitive to the buyer’s effort, and that performance-based contracts dominate when the output is equally sensitive to both the buyer’s and the vendor’s inputs. We also discuss how the performance of these contracts is affected with output uncertainty, process improvement opportunities, and the involvement of multiple buyers and vendors in the joint-production process. Our model highlights the trade-offs underlying the choice of contracts in a collaborative service environment and identifies service process design changes that improve contract efficiency.

Key words: services; consulting; joint production; contracting; principal/agent models

History: Received October 21, 2008; accepted December 22, 2009, by Martin Lariviere, operations and supply chain management. Published online in Articles in Advance March 15, 2010.

1. Introduction

Business services, despite their spectacular growth in the 1990s (Apte and Nath 2006), are currently facing many challenges (The Economist 2007). In particular, the nature of the client-consultant relationship has changed; today it is more focused on delivering value through effective collaboration (Toppin and Czerniawska 2005).

Consider a client preparing a consulting engagement. Which contract should it offer? Fixed-fee contingent on output (FF-O), fixed-fee contingent on effort (FF-E), time-and-materials (TM), or performance-based (PB) pricing? Although it is clear that the choice of contract depends on the service environment characteristics, to date there exist few guidelines for contract selection.

In this paper, we present a formalized model of contracting, identify the key trade-offs underlying the contract choice, and suggest strategies to improve contract effectiveness. We analyze a service provision and delivery process involving collaboration between a buyer and a vendor (Fuchs and Leveson 1968). Because of the joint-production character of the process, the effort levels are complementary; i.e., the buyer exerts high efforts only when the vendor exerts high efforts, and vice versa. We consider three settings, depending on whether the buyer’s and the vendor’s efforts are verifiable. We show that FF-O contracts naturally arise when the buyer’s actions have been contractually specified; that TM and FF-E contracts are optimal when the vendor’s actions are verifiable, e.g., through on-site work and regular meetings; and that PB contracts are optimal when neither the buyer’s nor the vendor’s actions are verifiable, resulting in double moral hazard (Bhattacharyya and Lafontaine 1995). We also relate the optimal contract selection to the service characteristics, such as output sensitivity to the buyer’s and the vendor’s inputs, output measurability, output uncertainty, cost-reduction and value-enhancement opportunities, as well as the presence of multiple buyers or vendors.

The paper is organized as follows. We review the related literature in the next section and present an overview of contracting in the consulting industry. Section 3 introduces the basic model, benchmarks the decentralized decision making to the first-best solution, and relates the choice of contract to the sensitivity of the output to the buyer’s and vendor’s actions. We then discuss in §4 the impact of output uncertainty, process improvement opportunities, and the involvement of multiple buyers and vendors. We present our conclusions and directions for future research in §5. All proofs appear in the electronic companion.1

1 An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.
2. Literature Review

The information service sector, in particular, business services and medical and educational services, was the fastest growing sector of the U.S. economy in 1992–1997 (Apte and Nath 2006). Despite the importance of this sector, research on service operations has primarily focused on the physical aspects of services including queuing models (Allon and Federgruen 2007), logistics ( Bramel and Simchi-Levi 1997), retailing (Fisher and Raman 1996), and revenue management (Talluri and van Ryzin 2004). One could in fact argue that research on “white-collar” service operations is still in its infancy (Hopp et al. 2009).

Building on Karmarkar and Pitblado’s (1995) framework for analyzing service processes, service competition, and its implications on service strategy, this paper analyzes the contractual arrangements arising in a collaborative, business-to-business service environment. The main research question pursued by this paper is to relate the choice of contract—among fixed-fee, time-and-materials, and performance-based contracts—to the service characteristics.

2.1. Contract Theory

Contract theory has mostly focused on adverse selection, when one of the parties has private information about her characteristics, and moral hazard, when one party’s efforts are unverifiable by the other party. Adverse selection can be mitigated by screening the buyers and/or vendors through competitive bidding and menus of contracts. Moral hazard, in contrast, can be addressed by offering the agent a contract that links her compensation to the output of her work. Inefficiencies arise, however, when the agent is risk averse or has limited liability, or when the agent’s performance measure is distorted. See Laffont and Martimort (2002) and Gibbons (2005) for recent overviews of contract theory.

Contract theory has recently been applied to business-to-business settings, from cross-functional coordination (Porteus and Whang 1991, Kouvelis and Lariviere 2000) to supply chain coordination (Lariviere 1999, Corbett 2001, Cachon 2003, Perakis and Roels 2007, Taylor and Plambeck 2007). In contrast to supply contracts, which are based on manufactured quantities, we focus on contracts for projects, such as those arising in construction (Bajari and Tadelis 2001), legal services (Rubinfeld and Scotchmer 1993), movie making (Chisholm 1997), spare-part supply chains (Kim et al. 2007), and call centers (Hasija et al. 2008, Ren and Zhang 2008), among others. Rather than being industry specific, we adopt an abstract representation of the service output to generate insights applicable to most collaborative services, including consulting, financial planning, health care, and information technology (IT) outsourcing services. More importantly, we consider a joint-production function, rather than a traditional principal-agent framework, which fundamentally alters the nature of contracting.

2.2. Collaborative Services

The notion of collaboration in services was first introduced by Fuchs and Leveson (1968) and discussed in an operations management context by Chase (1981). Joint-production processes have become so ubiquitous in services that one could argue that every service involves some degree of value cocreation (Spohrer et al. 2008). Managing collaboration in organizations presents many challenges, such as the deployment of globally distributed work teams (Kumar et al. 2005) or the adoption of interorganizational (information sharing) systems (Chi and Holsapple 2005). Also, the prevalence of collaborative work in complex, dynamic environments makes the traditional top-down approach to process design obsolete. In particular, Kogan and Muller (2006) and Hill et al. (2006) report that most collaborative work takes place on an ad hoc basis and often relies on personal information management tools, whereas prescribed work processes serve only as reference models. Traditional approaches to process analysis are therefore inadequate, and new tools must be developed to model the locus of work in collaborative environments (Kieliszewski et al. 2007).

From a contracting perspective, collaboration may lead to double moral hazard when the parties’ efforts are nonverifiable, which typically makes the first-best effort levels not enforceable (Holmström 1982). The second-best optimal contracts can be linear when the parties are risk neutral (Bhattacharyya and Lafontaine 1995, Corbett et al. 2005) but are in general nonlinear when they are risk averse (Kim and Wang 1998). In supply chain settings, Iyer et al. (2005) study the contracting process for product specification and production, where the product costs depend on both the buyer’s and the supplier’s resources; Plambeck and Taylor (2006) demonstrate the existence of simple relational contracts when the buyer and the supplier interact over multiple periods. Xue and Field (2008) analyze the pricing and effort allocation in collaborative services with information stickiness, similar to our work, but consider substitute, rather than complementary, effort levels. Our model, although similar in spirit, aims instead to relate the choice of contract to the service environment characteristics, such as output elasticities, effort verification costs, output uncertainty, output measurability, and process improvement opportunities.

2.3. Contracting in Practice

Contract choice has also been studied from an empirical perspective, in offshore drilling (Corts and Singh 2004), maintenance of commercial aircrafts (Guajardo et al. 2009), and IT (Banerjee and Duflo 2000, Gopal
et al. 2003, Kalnins and Mayer 2004). Gurbaxani (2007) reviews the empirical literature and proposes a transaction economics framework to analyze outsourcing contracts for information systems. It is generally observed that contract choice depends on project complexity, project size, resource shortage, and a firm’s reputation.

Although our abstract model encompasses various industries, we will use the consulting industry as a guiding example throughout our analysis. The current state of the consulting industry, its challenges, and future are reviewed in Toppin and Czerniawska (2005), Czerniawska and May (2004), and Czerniawska (2002, 2007).

Because of the inherent diversity of consulting projects, there exists a wide variety of consulting contracts. Løwendahl (2005) reports that the two most common contracts for professional services are fixed-fee contracts—perhaps subject to increases if parts of the contract terms are altered in the process—and payment by the hour. In addition, she notes that more creative forms of contracts are sometimes adopted, such as shared-savings contracts or “no-cure-no-pay” contracts.

Sheedy (2008), from Forrester Research, identifies four types of contracts arising in IT consulting and system integration projects: (i) fixed price, (ii) time and expenses, (iii) risk/reward, and (iv) value based. Our representations of FF-O, TM, and PB contracts are consistent with his definitions of fixed-price, time-and-expenses, and value-based contracts. FF-E contracts are hybrid, because they consist of a fixed fee, similar to Sheedy’s fixed-price contracts, but based on inputs, similar to his time-and-expenses contracts. His risk-reward contract, which complements the simple fixed-price and time-and-expenses contracts with bonuses and penalties for reaching certain performance indicator targets (e.g., speed of delivery), is not captured by our framework, because our model of output is unidimensional and only measures the overall service value. Sheedy also identifies the benefits and challenges of each contract, summarized in Table 1. Our formal analysis will either confirm or circumstantiate these insights.

3. Basic Model

3.1. Model Components

3.1.1. Inputs. Inputs are the buyer’s and the vendor’s efforts, denoted by $x$ and $y$, respectively. We assume linear costs of effort, with $c_B$ and $c_V$, respectively, denoting the marginal costs of effort for the buyer and the vendor, which are effectively the cost of adding workforce on the project multiplied by some effort-to-time ratio, measuring the workforce efficiency (or productivity). The assumption of linear costs, although not essential to our model (any power function of the type $c_B x^m$ and $c_V y^n$ with $m, n \geq 1$ would lead to similar results), is the most relevant to business-to-business engagements, where the cost of adding workforce on a particular project is relatively constant over a reasonable range of values (Corbett et al. 2005).

3.1.2. Output. The output of the collaborative work $V(x, y)$ is modeled with a continuously differentiable, strictly concave function, nondecreasing in the effort levels. We furthermore assume that the output function is supermodular, consistent with the notion of value cocreation in services (Spohrer et al. 2003, Fuchs and Leveson 1968, Bettencourt et al. 2002), with $V(x, 0) = V(0, y) = 0$ and $\lim_{x \to \infty} \partial V(x, y)/\partial x = \lim_{y \to \infty} \partial V(x, y)/\partial y = \infty$. That is, no output is generated if one party does not contribute, and there is a large benefit from exerting more effort when the other party exerts a high effort. Many successful consulting engagements indeed operate “in a spirit of teamwork” (Shenson 1990), where clients and consultants work closely together. If the consultants are bringing a new approach to solving a problem, then the client brings intimate knowledge of the context for that problem, what kind of solution is being sought, whether any attempt to resolve it has already been made. Each side needs the other.

(Czerniawska and May 2004, p. 115)

For instance, the Cobb-Douglas function $V(x, y) = \mu x^\alpha y^\beta$ with $\mu > 0$, $\alpha, \beta > 0$, and $\alpha + \beta < 1$ satisfies these assumptions.

3.1.3. Information and Effort Verifiability. We assume that the project output $V(x, y)$ is contractible but that the effort levels $x$ and $y$ are unverifiable,

<table>
<thead>
<tr>
<th>Table 1 Benetfits and Challenges of Consulting Contracts</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Contracts</strong></td>
</tr>
<tr>
<td>Fixed price (FF-O/FF-E)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Time and expenses (TM/FF-E)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td>Value based (PB)</td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
<tr>
<td></td>
</tr>
</tbody>
</table>

resulting in double moral hazard. Each party is, however, given the opportunity to invest in an effort-reporting/monitoring system that would make its actions verifiable by a third party and to use this information in the contract. Effort reporting/monitoring systems are, however, not free, partly because of information stickiness (Xue and Field 2008). For simplicity, we assume affine costs and assume that it costs the buyer \( \phi_B + \phi_R x \) to make its actions verifiable and costs the vendor \( \phi_V + \phi_V y \) to make its actions verifiable, where \( \phi_B, \phi_V \geq 0 \) for \( i = 0, 1 \). Our results also hold for any other cost allocation. As discussed in §§3.1.1 and 3.1.2, one can interpret these costs as the costs of specifying one party’s roles and responsibilities before the contract is signed, the costs of reporting progress during the contract implementation (e.g., face-to-face meetings, intermediate reports, or presentations), and the costs of monitoring one party’s efforts (e.g., on-site work).

All cost information is assumed to be common knowledge. Otherwise, the moral hazard problem would be assessed with adverse selection, but, as argued by Bajari and Tadelis (2001), adverse selection is typically resolved through competitive bidding or reputation rather than through the choice of contract type and therefore lies beyond the scope of this paper.

3.1.4. Description of the Contracting Game. We model the contracting process as a one-period Stackelberg game, in which a principal—either the buyer or the vendor—offers a contract to the other party and where the choice of effort gives rise to a Nash subgame at the lower level. The timing of the game is modeled as follows: (i) the principal offers the contract; (ii) the agent accepts the contract if its profit is larger than its reservation utility, denoted by \( U > 0 \), or rejects the contract otherwise; (iii) if the agent has accepted the contract, both parties choose their respective effort level; and (iv) payments are collected.

Given our focus on business-to-business transactions, we assume that both parties are risk neutral and have no limited-liability constraints. Under this assumption, it is well established that there exists no contract that induces both parties to choose the first-best (i.e., system-optimal) effort levels as a Nash equilibrium with double moral hazard, at least in the absence of a budget breaker (Holmström 1982).

We will therefore focus our analysis on second-best contracts, under risk neutrality, but shall however discuss their effectiveness under risk aversion.

We further assume that the buyer is the principal and that the vendor is the agent, consistent with the recent change in the balance of power in the consulting industry (Czerniawska 2007, p. 106). This assumption is, however, without loss of generality under the assumptions of risk neutrality and no limited-liability constraint, because any second-best contract that includes a fixed fee can transfer the agent’s surplus to the principal.

3.2. First-Best Contract

As a benchmark, consider the first-best solution, as if the effort levels were verifiable at no additional cost, enabling the principal to choose the agent’s effort so as to maximize the total surplus:

\[
\Pi_{FB} = \max_{x, y \geq 0} V(x, y) - c_B x - c_V y.
\]

Under the assumptions stated in §3.1.2, there exists a unique pair of positive efforts \( (x_{FB}, y_{FB}) \) that maximizes the total surplus. These efforts are further complementary, given that \( V(x, y) \) is assumed to be supermodular. That is, the buyer is more inclined to exert high efforts when the vendor exerts high efforts, consistent with what is observed in practice (Toppin and Czerniawska 2005). In contrast, if \( V(x, y) \) had been submodular (e.g., separable), it would have been optimal to set at least one effort level to zero, and no collaboration would have occurred.

In particular, when \( V(x, y) = \mu x^\alpha y^\beta \), the first-best effort levels are equal to

\[
\begin{align*}
x_{FB} &= \mu^{1/(1-\alpha-\beta)} \left( \frac{\alpha}{c_B} \right)^{(1-\beta)/((1-\alpha-\beta)} \left( \frac{\beta}{c_V} \right)^{\beta/(1-\alpha-\beta)}, \\
y_{FB} &= \mu^{1/(1-\alpha)} \left( \frac{\alpha}{c_B} \right)^{\alpha/(1-\alpha)} \left( \frac{\beta}{c_V} \right)^{(1-\alpha)/(1-\alpha-\beta)},
\end{align*}
\]

and the optimal ratio of effort levels, \( x_{FB}/y_{FB} = (ac_V)/(c_B \beta) \), is independent of the project scale \( \mu \). A party will thus exert relatively more effort if its cost of effort is lower and if the output is more sensitive to its effort. Furthermore, the maximum total surplus is equal to

\[
\Pi_{FB} = \mu^{1/(1-\alpha-\beta)} \left( \frac{\alpha}{c_B} \right)^{\alpha/(1-\alpha-\beta)} \cdot \left( \frac{\beta}{c_V} \right)^{\beta/(1-\alpha-\beta)} (1 - \alpha - \beta).
\]

3. By strict concavity of \( V(x, y) \), if there exists a maximum, this maximum uniquely solves the first-order optimality conditions: \( \partial V(x, y)/\partial x = c_B \) and \( \partial V(x, y)/\partial y = c_V \). By supermodularity of \( V(x, y) \), \( \partial V(x, y)/\partial x \) continuously increases as \( y \) increases, from zero when \( y = 0 \) (because \( V(x, 0) = 0 \) for all \( x \)) to infinity when \( y \rightarrow \infty \) by assumption. Hence, for any \( x \geq 0 \), there exists a \( y > 0 \) such that \( \partial V(x, y)/\partial x = c_B \). Similarly, for any \( y \geq 0 \), there exists an \( x > 0 \) such that \( \partial V(x, y)/\partial y = c_V \). Therefore, there exists a global maximum in the positive orthant, and this maximum is unique.
The next proposition further characterizes the behavior of the optimal effort levels when $V(x, y) = \mu x^\alpha y^\beta$. In particular, because $x^\alpha$ is increasing in $\alpha$ if and only if $x \geq 1$, the optimal effort levels are non-monotonic with respect to the effort elasticities.

**Proposition 1.** When $V(x, y) = \mu x^\alpha y^\beta$, the following apply:

1. $x^{FB}(\alpha)$ is either increasing-decreasing or always increasing with $\alpha$. In particular, $x^{FB}(\alpha)$ is always increasing when $x^{FB}(\alpha) \geq e^{-1}$. Similar results hold for $y^{FB}(\beta)$.

2. $y^{FB}(\alpha)$ is either decreasing-increasing or always decreasing with $\alpha$. In particular, $y^{FB}(\alpha)$ is decreasing if and only if $x^{FB}(\alpha) \leq e^{-1}$. Similar results hold for $x^{FB}(\beta)$.

3. $\Pi^{FB}(\alpha, \beta)$ is jointly convex with respect to $\alpha$ and $\beta$. It is increasing with $\alpha$ if and only if $x^{FB} \geq 1$ and increasing in $\beta$ if and only if $y^{FB} \geq 1$.

In contrast to the choice of effort levels $x$ and $y$, which relate to the service execution, the choice of elasticity parameters $\alpha$ and $\beta$ is made at the service design stage. For instance, a firm can choose to develop a custom software application in house with the help of external technology consultants ($\alpha \gg \beta$), or it can decide to completely outsource it, with the client’s role limited to specifying the software requirements ($\beta \gg \alpha$). Information technology may also enable new service designs. For example, Ernst & Young’s automated consulting tool “Ernie,” which gives clients access to previously asked questions, databases, and industry trends and which permits them to ask questions of experts (Czerniawska 2002), effectively makes the initial problem diagnosis more sensitive to the client’s effort.

Because the first-best surplus $\Pi^{FB}(\alpha, \beta)$ is convex in $(\alpha, \beta)$, it attains its maximum over the closure of $\{(\alpha, \beta) : \alpha, \beta > 0, \alpha + \beta < 1\}$ at $(0, 0)$, at $(1, 0)$, or else at $(0, 1)$. That is, the total surplus is maximized when the output is completely insensitive to the effort of at least one party. Hence, in an ideal project, at least one of the parties should exert only an infinitesimal amount of effort or—for more general output functions—just enough effort to get its job done, as any additional effort provision would not increase the service output. In particular, the highest value is generated when the service is completely insourced, i.e., when $\alpha \gg \beta$, such as when a management innovation idea matures (Czerniawska 2002) or when it is completely outsourced, i.e., when $\beta \gg \alpha$, such as for market research or application development (Czerniawska 2007). The sensitivity of the output function to the respective effort levels will directly affect the choice of contracts, as it will be easier to contractually specify the actions and effort levels of the party that has the lowest output elasticity.

Proposition 1 also shows that the buyer’s and the vendor’s efforts $x^{FB}(\alpha)$ and $y^{FB}(\beta)$ evolve non-monotonically with respect to their own elasticities.

Although it is, in general, expected that a party should increase its effort level when the output becomes more sensitive to its effort, the opposite effect can also happen. In particular, if $\alpha$ increases, the vendor may choose to exert less effort, because of its lower relative importance, leading the buyer to adopt the same behavior. This result highlights the complexity of allocating responsibilities in a collaborative environment. However, when $x^{FB}$ and $y^{FB}$ are larger than $1/e$, which happens, from (1), when the project has a high payoff $\mu/(c_y c_B)$, any increase in output elasticities will translate into higher effort levels for both parties. That is, for high-return projects, enhancing one party’s responsibilities will lead to a mutual increase in efforts, yielding higher output.

### 3.3. Optimal Choice of Contract

We now turn our attention to the optimal choice of contract when the actions $x$ and $y$ are not freely verifiable. Three situations must be considered: the buyer’s efforts are verifiable at a cost $\phi_{B0} + \phi_{B1} x$, the vendor’s efforts are verifiable at a cost $\phi_{V0} + \phi_{V1} y$, or no action is verifiable, resulting in double moral hazard. The fourth possible case, where both parties invest in an effort reporting/monitoring system, can easily be seen to be dominated by the first two scenarios.

#### 3.3.1. Verifiable Buyer’s Efforts

When the buyer’s efforts are verifiable at a cost of $\phi_{B0} + \phi_{B1} x$, the problem reduces to a standard principal-agent model. The transfer payment $t$ can only be a function of what is contractible, namely the buyer’s effort $x$ and the output $V$. The buyer chooses the contract that maximizes its total profit, subject to the vendor’s participation (or individual rationality) constraint $(IR_V)$, ensuring that the vendor obtains a profit larger than its reservation utility $U$, and the incentive compatibility constraints, $(IC_V)$ and $(IC_B)$, which define the noncooperative Nash game of effort choice. Accordingly, the buyer’s problem can be formulated as follows:

$$\max_{t(x, y)} \quad V(x, y) - (c_B + \phi_{B1})x - \phi_{B0} - t(x, V(x, y))$$

s.t. $t(x, V(x, y)) - c_y y \geq U$  

$y = \arg \max \underset{\tilde{y}}{t(x, V(x, \tilde{y}))} - c_y \tilde{y}$  

$x = \arg \max \frac{V(\bar{x}, y) - (c_B + \phi_{B1})\bar{x} - \phi_{B0}}{-t(\bar{x}, V(\bar{x}, y))}$.

The next proposition shows that, when the buyer’s efforts are verifiable, the vendor’s effort can be inferred from the output and the buyer’s effort, eliminating the double moral hazard problem.
Proposition 2. When the buyer’s efforts are verifiable, the following fixed-fee contract with a payment contingent on output (FF-O) is optimal:

\[
t(x, V) = \begin{cases} 
0 & \text{if } V < V^{\text{FF-O}} \text{ and } x \geq x^{\text{FF-O}}, \\
\mathcal{c}_V y^{\text{FF-O}} + U & \text{if } V \geq V^{\text{FF-O}} \text{ and } x \geq x^{\text{FF-O}}, \\
\infty & \text{if } x < x^{\text{FF-O}},
\end{cases}
\]

where \( V^{\text{FF-O}} = V(x^{\text{FF-O}}, y^{\text{FF-O}}) \) and \((x^{\text{FF-O}}, y^{\text{FF-O}}) = \arg\max_{x, y} V(x, y) - (c_{\beta} + \phi_{\beta})x - c_{\gamma}y - \phi_{\beta} \).

When \( \phi_{\beta} > 0 \), both parties will exert less effort than optimal, i.e., \( x^{\text{FF-O}} \leq x^{FB} \) and \( y^{\text{FF-O}} \leq y^{FB} \). Moreover, if \( V(x, y) = \mu x^\alpha y^\beta \), the buyer exerts proportionally less effort than the vendor, i.e., \( x^{\text{FF-O}}/y^{\text{FF-O}} = (\alpha c_V)/\beta(c_{\beta} + \phi_{\beta}) \leq x^{FB}/y^{FB} \).

The performance of FF-O contracts critically depends on the buyer’s ability to credibly commit to the effort level \( x^{\text{FF-O}} \) in the contract. For instance, standard outsourcing contracts include clauses regarding the client’s roles and responsibilities, the staffing of the project, the transfer of employees and/or assets, and the provision of certain services or assets (Halvey and Melby 2007). Specifying the client’s efforts, however, goes beyond contractual agreements. In particular, Bettencourt et al. (2002) highlight the importance of holding information sessions or kickoff meetings so as to inform the client of the significance of its contribution in the service delivery and to agree on common protocols for information exchange, flexibility, and solidarity. Specifying the buyer’s role is critical, as “clients are often the most unprofessional party...[they] may be dealing with a large project for the first time and be much less clear about their role” (Toppin and Czerniawska 2005, p. 54). In fact, with an incomplete specification of its roles and responsibilities, the buyer would fail to exert the required efforts, given that only the vendor is accountable for the project, ultimately reducing the output value. It is indeed common that FF-O contracts fail to deliver their promises because of a client’s lack of involvement. In the particular case of offshore contracts (which are typically based on FF-O contracts; see, for instance, O’Connell and Loveman 1995), Czerniawska (2002, p. 171) observes that “organizations were wrong if they thought that they could simply bundle up a piece of work, subcontract it to a company in India and just leave them to deliver without too much ongoing co-ordination.”

Although the performance of FF-O contracts critically depends on the specification of the buyer’s efforts \( x^{\text{FF-O}} \), the accuracy of the buyer’s effort measurements is less crucial. In particular, FF-O contracts can still enforce the first-best effort levels \((x^{\text{FF-O}}, y^{\text{FF-O}})\) when the measurement of the buyer’s efforts is subject to a random noise \( \varepsilon \), unobservable to the parties. If the range of \( \varepsilon \) is bounded from below, i.e., \( \varepsilon \geq \varepsilon_0 \), and if the contract specifies that the buyer must pay a large penalty (e.g., the liquidated damages when the contract is breached) if the measured effort, \( x + \varepsilon \), is smaller than \( x^{\text{FF-O}} + \varepsilon_0 \), then the buyer will never take the chance of exerting a smaller effort than agreed upon, for fear of having to pay the indemnity. Therefore, the pair of effort levels \((x^{\text{FF-O}}, y^{\text{FF-O}})\) remains enforceable with noisy effort measurements, as long as the noise is bounded from below and the liquidated damages are large enough, even if either party is risk averse (given that payments are deterministic).

Besides the buyer’s roles and responsibilities, the contract must also specify a performance measure on which the fee transfer will be contingent. In fact, the contract’s effectiveness is highly sensitive to the alignment of this performance measure with the project output (Baker 2002). For instance, if the vendor must perform two complementary tasks, \( y_1 \) and \( y_2 \), both associated with a unit cost \( c_{\gamma} \), additional inefficiencies arise when the performance measure \( \hat{V}(x, y_1, y_2) = \mu x^\alpha (y_1^{\delta/\delta - 1})^\beta \) is a distorted metric of the output \( V(x, y_1, y_2) = \mu x^\alpha (y_1^{\delta - 1})^\beta \), with \( \delta \neq \delta_0 \), and \( \delta_0 \in (0, 1) \). In particular, when \( \phi_{\beta} = \phi_{\gamma} = 0 \), the ratio of the total surplus under a FF-O contract to the first-best total surplus turns out to be proportional to \( (\delta_0/\delta) ((1 - \delta)/\delta) (1 - \delta)^{-\delta_0} < 1 \). Hence, even if FF-O contracts guarantee a clear outcome (Table 1), this outcome may not correspond to what the buyer actually wants.

FF-O contracts are also sensitive to measurement errors in outputs when either party is risk averse, because the transfer payments would then be subject to uncertainty. In particular, a risk-averse vendor would inflate the fixed fee by a risk premium to be covered against the output variability (Seshadri 2005). When both parties are risk neutral, however, the contract parameters can be adjusted to still enforce the effort levels \((x^{\text{FF-O}}, y^{\text{FF-O}})\). Given the difficulty of establishing undistorted and accurate performance measures, projects are often broken down in stages.

---

4 Suppose, for instance, that the performance measure is a noisy measure of the output, i.e., \( V(x, y) + \varepsilon \), where \( \varepsilon \) is a random variable unobservable to both parties, such that the buyer pays the vendor a fixed fee whenever the performance measure \( V(x, y) + \varepsilon \) is larger than a certain threshold \( V^{\text{FF-O}} \). One could still induce the vendor to exert an effort level equal to \( V^{\text{FF-O}} \) by dividing the contingent fee, which is transferred when \( V(x, y) + \varepsilon \geq V^{\text{FF-O}} \), by \( P(\varepsilon \geq V^{\text{FF-O}} - V(x^{\text{FF-O}}, y^{\text{FF-O}})) \), so as to satisfy its participation constraint, and by adjusting the threshold \( V^{\text{FF-O}} \) to ensure that \( y^{\text{FF-O}} \) maximizes its expected profit.
with intermediate milestones, checkpoints, and prototypes to track progress.

To summarize, FF-O contracts need to clearly specify the scope of the project, define the client’s roles and responsibilities, and establish key performance indicators and service levels. The specification of these contract components is typically done in a preliminary “disambiguation” stage. For instance, HCL America requires its clients to “define clearly the capabilities and functions of the program they wanted” so as to reduce the ambiguity about the output, and to develop a “work plan process with checkpoints to serve as deadlines” to specify each party’s (and in particular the buyer’s) responsibility in the project (O’Connell and Loveman 1995). Additionally, FF-O contracts must be associated with a series of workshops or kick-off meetings that emphasize the importance of teamwork, creative problem solving, effective communication, and functional conflict resolution (Bettencourt et al. 2002).

The disambiguation and coordination processes may, however, consume valuable time, which we represent by the costs \( \phi_{bo} + \phi_{b1}x \). Clearly, the higher the client’s responsibilities, the more time the consultant and the client should invest in defining the client’s roles and responsibilities. In fact, the disambiguation process may be so lengthy that it may not be justified for small projects (O’Connell and Loveman 1995). The costs of making the buyer’s efforts verifiable can be incurred upfront—i.e., during the disambiguation stage—under the form of kickoff meetings (Bajari and Tadelis 2001) or extensive documentation (Bajari and Tadelis 2001), as well as during the project implementation, e.g., with the appointment of a liaison officer or the setup of onsite facilities to coordinate offshore contracts (O’Connell and Loveman 1995). The results of our analysis also hold when these costs are incurred by the vendor when both parties are risk neutral. To remain competitive, vendors should therefore develop economical specification and coordination processes, such as Ernst & Young’s online self-diagnosis tools (Czerniawska 2002) or HCL America’s dual project manager system (O’Connell and Loveman 1995).

### 3.3.2. Verifiable Vendor’s Efforts

When the vendor’s efforts are verifiable at a cost of \( \phi_{V0} + \phi_{V1}y \), the problem also reduces to a standard principal-agent model. In this case, the transfer payment \( t \) is a function of the vendor’s effort and the output, i.e., \( t(y, V) \), given that both are contractible. The buyer chooses the contract that maximizes its total profit subject to the vendor’s participation constraint \( (IR_y) \) and the incentive compatibility constraints, \( (IC_y) \) and \( (IC_b) \)—that is,

\[
\begin{align*}
\max_{(y, V)} & \quad V(x, y) - c_b x - t(y, V) \\
\text{s.t.} & \quad t(y, V) - (c_v + \phi_{V1})y - \phi_{V0} \geq U & (IR_y) \\
& \quad y = \arg\max_{\bar{y}}(\bar{y}, V(x, \bar{y})) \\
& \quad - (c_v + \phi_{V1})\bar{y} - \phi_{V0} & (IC_y) \\
& \quad x = \arg\max_{\bar{x}}(\bar{x}, y) - c_b \bar{x} - t(y, V(\bar{x}, \bar{y})). & (IC_b)
\end{align*}
\]

**Proposition 3.** When the vendor’s efforts are verifiable, the following fixed-fee contract with a payment contingent on effort \((FF-E)\) is optimal:

\[
t(y, V) = \begin{cases} 
0 & \text{if } y < y_{FF-E}, \\
(c_v + \phi_{V1})y_{FF-E} + U + \phi_{V0} & \text{if } y \geq y_{FF-E},
\end{cases}
\]

where \( (x_{FF-E}, y_{FF-E}) = \arg\max_{x, y} V(x, y) - c_b x - (c_v + \phi_{V1})y - \phi_{V0} \).

Similar to the case where the buyer’s efforts are verifiable, one can show that, when \( \phi_{V1} > 0 \), the parties exert less effort, i.e., \( x_{FF-E} \leq x_{FB} \) and \( y_{FF-E} \leq y_{FB} \). Moreover, when \( V(x, y) = \mu x^\beta y^\alpha \), the buyer exerts proportionally more effort, i.e., \( x_{FF-E}/y_{FF-E} = (\alpha(c_v + \phi_{V1})/(\beta c_b) \geq x_{FB}/y_{FB} \).

The same total surplus can alternatively be achieved through a TM contract. Given that the vendor’s effort is contractible, the vendor can transfer its decision rights to the buyer, i.e., let the buyer decide how much effort it should exert. For instance, HCL America reports that some of its clients who have opted for a TM contract use “HCL America engineers instead of hiring a permanent programming staff of their own” (O’Connell and Loveman 1995, p. 4). From Proposition 3, the optimal TM contract is a linear compensation scheme \((c_v + \phi_{V1})y \) consisting of the vendor’s hourly rate \( c_v \) and direct expenses of the verification of its effort \( \phi_{V1} \), to which is added some overhead cost \( U + \phi_{V0} \) (Shenson 1990).

In contrast to FF-O contracts, FF-E and TM contracts are not based on output and therefore remain unaffected by noisy output measurements. As a result, specifying the scope and performance measures of a project is not critical for FF-E/TM contracts. HCL America indeed recommends TM contracts for “projects that have to be designed as work goes along and new decisions arise, projects that cannot be specified precisely in advance” (O’Connell and Loveman 1995, p. 4). Therefore, if the vendor is risk averse and

5 If the vendor’s verification costs were incurred by the buyer instead, the optimal compensation would be equal to \( U + \alpha y \); i.e., the buyer would just reimburse the vendor for its costs of effort.
the output is uncertain at the time the contract is written, the buyer will be more likely to select a TM contract rather than a FF-O contract, so as to reduce the vendor’s risk premium (Seshadri 2005).

Effective TM contracts should, however, be associated with economical verification processes (e.g., on-site work, frequent meetings) to provide visibility on the vendor’s work at low cost. For instance, typical outsourcing contracts require the vendor to develop a management procedure manual as well as a manual detailing the vendor’s day-to-day procedures. The manual must also include a clause that allows the client to audit the vendor’s operations and fees (Halvey and Melby 2007). Alternatively, the vendor can increase the visibility of its actions by submitting progress or interim reports or working on site (O’Connell and Loveman 1995). In addition, these verification processes should be as accurate as possible when either party is risk averse, because the uncertainty in transfer payments would further reduce their incentives to exert efforts.

The costs of verifying the vendor’s actions can be reduced by an effective use of technology, such as by holding TelePresence rather than face-to-face meetings (The Economist 2009) or by building a certain degree of trust between the clients and the vendors, as it arises when they are engaged in a long-term relationship: “With interaction, there is more trust and less bureaucracy” (Toppin and Czerniawska 2005, p. 58). Gopal et al. (2003) and Corts and Singh (2004) indeed observed that TM contracts are more likely to be adopted when clients have had prior experience with a vendor. The recent development of client account teams indeed aims at ensuring continuity in the client relationship, by developing a single point of contact between the client and the consulting firm (Czerniawska 2007).

3.3.3. Double Moral Hazard. Without effort monitoring/reporting, no action is verifiable, resulting in double moral hazard. In this case, the transfer payment \( t \) is a function of the output only, i.e., \( t(V) \). The next proposition, which appears in Bhattacharyya and Lafontaine (1995) and Kim and Wang (1998), shows that, with double moral hazard, the second-best contract is a PB contract, according to which the agent receives a fixed fee \( s \) and a share \( b \) of the output, such as a proportion of the increased revenue following the implementation of a customer relationship management system (Sheedy 2008). In the following, we will refer to \( b \) as the bonus rate.

**Proposition 4.** When no action is verifiable, the following PB contract is (second-best) optimal:

\[
t(V) = bV + s,
\]

where the bonus rate \( b \) and the fixed fee \( s \) are optimally chosen.

Because of the double moral hazard, both parties exert less effort, i.e., \( x^{\text{FB}} \leq x^{\text{PB}} \) and \( y^{\text{FB}} \leq y^{\text{PB}} \), similar to FF-O and FF-E/TM contracts. If \( V(x, y) = \mu x^\alpha y^\beta \), the relative allocation of effort \( x^{\text{FB}}/y^{\text{FB}} \) can, however, be smaller or larger than \( x^{\text{PB}}/y^{\text{PB}} \), depending on whether \( \alpha < \beta \).

When the parties are risk neutral, PB contracts are not affected by the accuracy of the output measures, as long as the noises are unobservable to the parties, given that the bonus \( bV \) is linear with the output. When the vendor is risk averse, however, the optimal bonus is no longer linear (Kim and Wang 1998) because it must account for the risk aversion as well as the incentive provision.

In the case where \( V(x, y) = \mu x^\alpha y^\beta \), the optimal bonus rate is equal to

\[
b = (-\beta(1-\alpha) + \sqrt{\alpha\beta(1-\alpha)(1-\beta)})/(\alpha - \beta),
\]

and its contours are depicted in Figure 1. The bonus rate is independent of the project value \( \mu \) and costs \( c_V \) and \( c_B \), which, according to Bhattacharyya and Lafontaine (1995), explains that the same split rates are applied to different markets. The next proposition characterizes the sensitivity of the bonus rate to the effort elasticities as well as the sensitivity of the profit function to the bonus rate.

**Proposition 5.** When \( V(x, y) = \mu x^\alpha y^\beta \), the following apply:

1. \( \beta \leq b \leq 1 - \alpha \).
2. \( b(\alpha) \) is decreasing from 1 when \( \alpha = 0 \) to \( \beta \) when \( \alpha \rightarrow 1 - \beta \), convexly if \( \beta \geq 1/2 \) or in a convex-concave manner if \( \beta \leq 1/2 \), with the inflection point lying in the interval \( (\beta, 1 - \beta) \). In contrast, \( b(\beta) \) is increasing from 0 when
\( \beta = 0 \) to \( 1 - \alpha \) when \( \beta \to 1 - \alpha \), concavely if \( \alpha \geq 1/2 \) or in a concave-convex manner if \( \alpha < 1/2 \), with the inﬂection point lying in the interval \((\alpha, 1 - \alpha]\).

3. Let \( \Pi^{PB}(b) = V(x^{PB}(b), y^{PB}(b)) - c_{x}x^{PB}(b) - c_{y}y^{PB}(b) \) be the total surplus under a PB contract with bonus rate \( b \). When \( b = 1/2 \), \( \Pi^{PB}(b) \) is inelastic, i.e., \(|(b/\Pi^{PB}(b))(d\Pi^{PB}(b)/db)| \leq 1 \) if \( \alpha \leq 1/2 \) and \( \beta \leq 1/2 \).

Proposition 5 shows that more incentives are needed when the service output becomes less elastic to the buyer’s effort and more elastic to the vendor’s effort. When the service is completely outsourced (i.e., \( \alpha = 0 \)), the vendor has high-powered incentives. In contrast, when the service is completely done in house (i.e., \( \beta = 0 \)), the contract reduces to a fixed-fee contract. Finally, when \( \alpha = \beta, b = 1/2 \) and the risk is equally shared among both parties.

The second part of Proposition 5, together with Figure 1, shows that the optimal bonus rate, when \( V(x, y) = \mu x^{\alpha} y^{\beta} \), is relatively insensitive to the effort elasticities, provided that \( \alpha \) and \( \beta \) are bounded away from \( 0 \) and \( \alpha + \beta \) is bounded away from \( 1 \). Moreover, the last part of Proposition 5 shows that the total surplus is relatively flat around \( b = 1/2 \), provided that elasticities \( \alpha \) and \( \beta \) are no larger than \( 1/2 \). As a result, the 50%-50% split contract will be near optimal whenever the output sensitivity to each party’s efforts is moderate, which explains why it is so common in joint ventures (Bhattacharyya and Lafontaine 1995, Bai et al. 2004).

In contrast to FF-O and FF-E/TM contracts, PB contracts may involve a payment transfer from the vendor to the buyer (i.e., \( s \leq 0 \)). If the vendor has a limited-liability constraint, the buyer may thus need to adjust the parameters of the PB contract to avoid negative transfer payments, which could reduce the buyer’s ability to capture the total surplus. The buyer will therefore be more tempted to offer PB contracts to vendors that have no limited-liability constraints, and not necessarily to the most competent service providers, as shown in Table 1.

3.3.4. Optimal Contract Selection. As a result of our analysis, one can rank the different types of contracts according to their proﬁtability. In the particular case of the Cobb-Douglas joint-production function \( V(x, y) = \mu x^{\alpha} y^{\beta} \), the optimal choice of contract depends on the output elasticities with the respective effort levels \( \alpha \) and \( \beta \), the costs of effort relative to the veriﬁcation costs, \( \phi_{B0}/c_{B} \) and \( \phi_{V1}/c_{V} \), and the buyer’s and the vendor’s fixed costs of effort veriﬁcation \( \phi_{B0} \) and \( \phi_{V0} \), and is independent from the project value \( \mu \).

Figure 2 illustrates the contract ordering in the elasticity space \( \{(\alpha, \beta): \alpha > 0, \beta > 0, \alpha + \beta < 1 \} \) for speciﬁc values parameters. In the dashed region, the total surplus is lower than the vendor’s reservation utility \( U \), making collaboration infeasible. Although the

![Figure 2](image_url)

Notes. The dashed region corresponds to infeasible contracts. The parameters in this example were \( V(x, y) = \mu x^{\alpha} y^{\beta}, \phi_{B1} = 1, \phi_{V1} = 2, \phi_{B0} = \phi_{V0} = 0, \mu = 2, c_{B} = c_{V} = 1 \), and \( U = 0.075 \).

exact boundaries of the different regions depend on the parameter values, the next proposition, which proof is based on standard operations on convex sets (Boyd and Vandenberge 2004), demonstrates that the shapes of the regions depicted in Figure 2 hold in general. Let us denote by \( \mathcal{R}^{FF-O}, \mathcal{R}^{FF-E/TM}, \) and \( \mathcal{R}^{PB}(b) \) the total surpluses under FF-O, FF-E/TM, and PB contracts, respectively. For tractability, we assume that the bonus rate of PB contracts is set to \( 1/2 \) (in contrast, the bonus rate \( b \) was optimally chosen when drawing Figure 2), which is a minor simpliﬁcation given that PB contracts seem to dominate in the vicinity of the line \( \alpha = \beta \), on which setting \( b = 1/2 \) is optimal, and that \( \Pi^{PB}(b) \) is relatively flat around \( b = 1/2 \) by Proposition 5.

Proposition 6. Suppose \( V(x, y) = \mu x^{\alpha} y^{\beta} \) and \( \phi_{B0} = \phi_{V0} = 0 \). The \((\alpha, \beta)\) space can be divided into three connected regions: \( \mathcal{R}^{FF-O} = \{(\alpha, \beta): \Pi^{FF-O} \geq \max[\Pi^{FF-E/TM}, \Pi^{PB}(1/2)]\}; \mathcal{R}^{FF-E/TM} = \{(\alpha, \beta): \Pi^{FF-E/TM} \geq \max[\Pi^{FF-O}, \Pi^{PB}(1/2)]\}; \) and \( \mathcal{R}^{PB} = \{(\alpha, \beta): \Pi^{PB}(1/2) \geq \max[\Pi^{FF-O}, \Pi^{FF-E/TM}]\}\).

1. \( \mathcal{R}^{FF-O} \supseteq \{(\alpha, \beta): \alpha = 0, 0 \leq \beta \leq 1\}, \mathcal{R}^{FF-E/TM} \supseteq \{(\alpha, \beta): \beta = 0, 0 \leq \alpha \leq 1\}, \) and they have a linear boundary with each other.

2. \( \mathcal{R}^{PB} \) is convex and anchored at the origin.

3. The set of infeasible contracts \( \{(\alpha, \beta): \max[\Pi^{FF-O}, \Pi^{FF-E/TM}, \Pi^{PB}] < U\} \) is convex.

According to Proposition 6, FF-O contracts are preferred when the output is more sensitive to the vendor’s effort, i.e., \( \beta \gg \alpha \), for projects that do not critically depend on the efficiency of the systems and procedures that specify the buyer’s roles and
responsibilities (e.g., liaison officer, disambiguation process). HCL America, for instance, recommends FF-O contracts for projects that require little buyer involvement, i.e., projects that are “well defined and structured” (O’Connell and Loveman 1995). FF-E/TM contracts, in contrast, are preferred when the output is more sensitive to the buyer’s effort, i.e., \( \alpha \gg \beta \), for projects that do not critically depend on the efficiency of the systems and procedures that make the vendor’s efforts verifiable (e.g., on-site work, intermediate reports). HCL America, for instance, uses TM contracts for projects with a high degree of buyer involvement, i.e., “projects that had to be designed as work went along and new decisions arose” (O’Connell and Loveman 1995). In contrast, PB contracts dominate when the output is equally sensitive to the buyer’s and the vendor’s contribution, i.e., \( \alpha \approx \beta \), when the negative externality from splitting the output and therefore diluting the parties’ incentives is the smallest. Although PB contracts have recently grown in popularity, not only in consulting (Shenson 1990) but also in other industries, such as law firms or private schools (The Economist 2008a, b), because they give vendors the opportunity to link their work to the actual value they generate, it is important to note that their “greater likelihood for project success” (Table 1) is only relevant when \( \alpha \approx \beta \).

It is interesting to draw the parallel between Figure 2 and Løwendahl’s (2005) taxonomy of professional services. Løwendahl classifies professional firms in three categories, according to their strategic focus, with implications on the resources from which they derive their competitive advantage. Specifically, she distinguishes on one hand, firms that focus on delivering superior solutions, adaptable to a wide variety of clients, and which therefore rely on efficient organizational resources (e.g., Ernst & Young), and on the other hand, firms that build their competitive advantage on a few key long-term client relationships and which therefore rely on developing individual competences to be responsive to their clients’ needs (e.g., small partnerships). In between, there are firms that are specialized in delivering unique solutions to complex problems and that therefore must support expert professionals with strong organizations (e.g., engineering design firms). The client’s involvement in the service delivery is arguably the weakest in the first type of projects, which often consists of implementing standard solutions, and the strongest in the second type of projects, where consultants sell their services on an ad hoc basis, according to their client’s agenda. In the third type of projects, the client’s input is as important as the consultant’s; in fact, the client is often considered as a partner. Figure 2 therefore suggests that solution-driven firms should mostly adopt FF-O contracts, firms that derive their competitive advantage on their client relationships should work on a TM basis (especially given that verification costs could decrease as the relationship becomes more established), and the third type of firms would be more inclined to use PB contracts (e.g., Parsons Brinckeroff’s shared saving contract; Løwendahl 2005).

The nonconvexity of the feasible contracting region highlights the complexity of task allocation in a collaborative environment. In particular, an increase in the buyer’s elasticity (e.g., through greater responsibilities in the project) may not only decrease its effort level in the project (Proposition 1), but it could also lower the total surplus so far below the vendor’s reservation utility that the vendor would no longer be willing to participate.

In addition to the input verification costs, the choice of contract also depends, as discussed above, on the alignment between performance measure and output (FF-O contracts) and the vendor’s limited-liability constraint (PB contracts). Risk aversion would also impact the performance of all contracts if measurement errors affected the output (FF-O and PB contracts) or the vendor’s verification of efforts (TM/FF-E contracts). We analyze additional factors in the next section.

4. Extensions

In this section, we consider three extensions to our basic model. We first investigate the impact of output uncertainty on the choice of contract, then look at process improvement opportunities, and finally investigate contract performance in the presence of multiple buyers and sellers.

4.1. Output Uncertainty

The output value is often uncertain at the contracting stage because the service value may be customer specific or complex to describe (Karmarkar and Pitblado 1995, Frei 2006). To capture output complexity, we model the output as \( \epsilon_0 + \epsilon_1 V(x, y) \), where \( \epsilon_0 \) and \( \epsilon_1 \) are random variables. However, we assume that the effort levels can be adapted to different levels of output, i.e., that the parties observe \( \epsilon_0 \) and \( \epsilon_1 \) before choosing their effort levels.\(^6\) For instance, in an enterprise

\[^6\] Alternatively, one could assume that the noises \( \epsilon_0 \) and \( \epsilon_1 \) are unobservable to the parties, as in Bhattacharyya and Lafontaine (1995), such as when a product with intrinsic value \( V(x, y) \) has uncertain market success \( \epsilon_0 + \epsilon_1 V(x, y) \). When both parties are risk neutral, one can show that the effectiveness of the contracts presented in Propositions 2–4 remains unaffected by the unobservable output uncertainty, using a similar argument to our discussion of the impact of output measurement errors in §3. If the vendor were risk averse, however, output uncertainty would make the vendor command a risk premium with FF-O and PB contracts, thus making TM contracts more attractive.
resource planning system implementation, although preliminary effort could specify the scope of the implementation and the degree of customization, many decisions must be made during the implementation phase (e.g., Austin et al. 2002).

As a benchmark, we evaluate the sensitivity of the first-best profit function to the output uncertainty. Because (2) is convex increasing in $\mu$, the expected total surplus when $V(x, y) = \epsilon_1 x^\alpha y^\beta$ is larger when the distribution of $\epsilon_1$ has a larger mean residual life (Müller and Stoyan 2002). Specifically, consider two projects, 1 and 2, with uncertain value $\epsilon_i$, respectively, distributed according to probability distributions $\Phi_i$, $i = 1, 2$. If $\int_{0}^{\infty} (1 - \Phi_i(\xi)) d\xi \geq \int_{0}^{\infty} (1 - \Phi(\xi)) d\xi$, for all $\epsilon_1 > 0$, the expected total surplus is larger under distribution 1 than under distribution 2. That is, the project associated with the fattest right tail will generate the highest expected surplus. In particular, if the random output $\epsilon_1$ is normally or log-normally distributed, the expected total surplus is increasing with its variance, keeping its mean constant (Levy 1992). Hence, an ideal project, besides having the output sensitive to only one party (Proposition 1), would also have it highly uncertain, everything else being equal. Intuitively, uncertain projects that have a slight chance of leading to breakthrough outcomes are more valuable than standard projects with the same expected, but more predictable, outcomes.

We next characterize the adaptability of the different contracts to the output uncertainty. For simplicity, we focus on the Cobb-Douglas production function and consider additive and multiplicative noises. Let us define $E[I(\epsilon_0, \epsilon_1)]$ as the expected total surplus under contract $k = FF-O$, FF-E, TM, PB, or FB, when the output is uncertain; in particular, with no output uncertainty, the type of output is equal to $I(0, \mu) = I^o$, for any contract type $k$, as in §3.

**Proposition 7.** Let $V(x, y) = \epsilon_0 + \epsilon_1 x^\alpha y^\beta$, with $E[\epsilon_0] = 0$, $E[\epsilon_1] = \mu \geq 1$, and $\mu(\epsilon_1 > 0) = 1$, and let $\phi_0 = \phi_0 = 0$. If $\epsilon_0$ and $\epsilon_1$ are unverifiable but observed by the parties before they choose their effort levels, then

$$\frac{E[I^{FF-O}(\epsilon_0, \epsilon_1)]}{E[I^{TM}(\epsilon_0, \epsilon_1)]} \leq \frac{E[I^{FF-E}(\epsilon_0, \epsilon_1)]}{E[I^{TM}(\epsilon_0, \epsilon_1)]} \leq \frac{E[I^{FB}(\epsilon_0, \epsilon_1)]}{E[I^{TM}(\epsilon_0, \epsilon_1)]}.$$ 

Therefore, fixed-fee contracts are the least adaptable to changes in output, consistent with Table 1, because the effort level of at least one party must be specified before the noises are observed, whereas TM and PB contracts postpone the choice of effort. In particular, TM contracts dominate FF-E contracts in the presence of output uncertainty, even when both parties are risk neutral. Among the fixed-fee contracts, FF-O contracts are the least adaptable because the vendor must bear the entire risk associated with the output uncertainty, although its compensation is based on a flat fee. In contrast, FF-E contracts allocate both risks and rewards on the buyer, making the buyer’s incentives better aligned. To mitigate the lack of adaptability of fixed-fee contracts, projects are often broken down in multiple stages, with intermediate milestones and checkpoints.

Much recent business press has advocated for measuring performance with outcomes (i.e., business drivers) rather than with outputs (i.e., service levels); see Toppin and Czerniawska (2005). For instance, London’s strategic transport authority contracted with the PA Consulting Group to “reduce crime and fear of crime on buses, reduce the number of illegal private hire vehicles, and improve bus flow demonstrably” (Czerniawska and May 2004, p. 63) rather than just focusing on systems delivered and deadlines met. To better appreciate the value of measuring outcomes instead of outputs, we model the outcome as the value obtained by the buyer—i.e., $\epsilon_0 + \epsilon_1 x^\alpha y^\beta$, which subject to external factors—and the output as the intrinsic result of the collaborative process, namely $\mu x^\alpha y^\beta$. Assuming that both outcome and output are contractible, one can show that, consistent with what is advocated in the business press, PB contracts perform better whenever the vendor receives a share of outcome than whenever the vendor receives a share of output, because it better aligns the parties’ incentives. (The proof of this result exploits the fact that the bonus rate (3) is independent of the project value ($\epsilon_1$ or $\mu$) when the output has a Cobb-Douglas function.) Measuring outcomes is, however, not always valuable. In fact, using Jensen’s inequality, one can show that FF-O contracts perform better when the fixed fee is contingent on output rather than on outcome, because the vendor would bear the risk of the outcome without being properly remunerated (i.e., share the risk without the rewards). Hence, the recent push for outcome-based measurements should be considered with caution, as different contracts call for different performance measures.

### 4.2. Process Improvement

We next evaluate the impact of process improvement opportunities on the optimal contract choice. We assume that the vendor can spend $z$ hours on process improvement, at a private, i.e., noncontractible, unit cost $c_{RT}$. We consider three types of process improvement opportunities, corresponding to the three “E’s” of consulting projects (Toppin and Czerniawska 2005, p. 50): (1) the vendor can improve the project economy by reducing its cost of effort $c\epsilon v$ to $c\epsilon v g(z)$, where $g(z) < 0$; (2) the vendor can improve the project efficiency by working only $g(z)y$ hours to provide $y$ units of effort, where $g(z) < 0$; and (3) the vendor can improve the
project effectiveness by increasing the project value from $V(x, y)$ to $V(x, y)g(z)$, where $g(z) > 0$.

**Proposition 8.** When the vendor can spend $z$ hours on process improvement, at a private cost $c_Vz$, the following results apply:

1. If the vendor’s unit cost of effort is equal to $g(z)c_V$, where $g(0) = 1$, $g'(z) < 0$, and $g''(z) > 0$, then $z^k$—if it is positive—solves $-c_\nu y^p g'(z^k) = c_\nu$ for $k = FB, PB, FF-E/TFM, FF-O$.

2. If the vendor works $g(z)y$ hours to provide $y$ units of effort, where $g(0) = 1$, $g'(z) < 0$, and $g''(z) > 0$, then $z^k$—if it is positive—solves $c_\nu y^p g'(z^k) = c_\nu$ for $k = FB, PB, FF-O$, whereas $z^{FF-E/TFM} = 0$.

3. If the service output is equal to $V(x, y)g(z)$, where $g(0) = 1$, $g'(z) > 0$, and $g''(z) < 0$, then $z^k$—if it is positive—solves $c_\nu g(z^k)(\partial V(x^k, y^k)/\partial y) = c_\nu y^p g'(z^k)V(x^k, y^k)$ for $k = FB, PB, FF-O$, whereas $z^{FF-E/TFM} = 0$.

The first result states that the vendor has the same incentives, under all contracts, to reduce its costs of effort, in the sense that the optimal effort in process improvement must satisfy the same optimality conditions. In contrast, the second and third results state that, under FF-E and TM contracts, the vendor has no incentive to improve its efficiency (ratio of effort to time), because it is paid by the hour, or to improve the overall project value, because its compensation does not depend on the output, in contrast to the other contract types. Consistent with Proposition 8, Guajardo et al. (2009) report that TM contracts are associated with lower product reliability than PB contracts in the maintenance, repair, and overhaul industry for commercial aircrafts. The lack of incentives with TM contracts to improve the project efficiency is, according to Seshadri (2005, p. 3), their main drawback, as it may lead a “rampant change-order temptation” (Table 1), where “consultants are encouraged to allow scope creep—the longer the project runs, the more money it will make” (Sheedy 2008, p. 3). To overcome this challenge, TM contracts could be assorted with bonuses and penalties if certain performance targets have been reached or not (Seshadri 2005).

### 4.3. Multiple Buyers and Multiple Vendors

We finally investigate the impact of having multiple buyers or multiple vendors involved in the joint-production process. Multivendor (or multisourcing) arrangements have become more common in the consulting industry as clients want to have access to world-class expertise and to make advisory work independent from system implementation (Toppin and Czerniawska 2005). Multibuyer arrangements arise, for instance, in supply chain projects, such as when a manufacturer and a retailer jointly hire a consultant to streamline their supply chain.

For simplicity, let us consider a service involving one buyer and two symmetric vendors, with unit cost of effort $c_V/2$, and a process output equal to $\bar{V}(x, y_1, y_2) = V(x, g(y_1, y_2))$, where $y_i$ refers to the effort level of vendor $i$, $i = 1, 2$, and $g(y_1, y_2)$ is a generalized mean function, i.e., $g(y_1, y_2) = ((y_1^p + y_2^p)/2)^{1/p}$ for any real nonzero number $p \leq 1$. We assume that the buyer offers the vendors identical contracts, and we only consider the symmetric Nash equilibrium in the lower-level effort-choice game; that is, in equilibrium $y_1 = y_2$.

Let $\Pi^b$, respectively, denote the total surplus, obtained with contract $k = FF-O, FF-E, TM, PB, or FB$, with two symmetric vendors. For comparison purposes, we also consider $\Pi^b$, the corresponding total surplus obtained with one vendor with cost $c_V$ and output function $V(x, y)$. It turns out that, by construction, if $(x^{FB}, y^{FB})$ maximizes the first-best total surplus with one vendor $\bar{V}(x, y) - c_\nu x - c_\nu y$, the triplet $(x^{FB}, y^{FB}, y^{FB})$ will maximize the first-best total surplus with two vendors $\bar{V}(x, y_1, y_2) - c_\nu x - (c_\nu/2)(y_1 + y_2)$. Hence, $\Pi^{FB} = \Pi^{FB}$. The next proposition shows that having multiple vendors involved in the joint-production process may create a negative externality with PB contracts, similar to free riding.

**Proposition 9.** Let $\bar{V}(x, y_1, y_2) = V(x, g(y_1, y_2))$ with $g(y_1, y_2) = ((y_1^p + y_2^p)/2)^{1/p}$, $p \neq 0$, $p \leq 1$. Let the two vendors be symmetric with effort provision cost $c_\nu y$, and effort verification cost $\phi_V y_i$, $i = 1, 2$. If the buyer offers identical contracts to both vendors and if the vendors choose their effort levels symmetrically, then $\Pi^b = \Pi^b$, for $k = FF-O, FF-E, and TM$; in contrast, $\Pi^{FB} < \Pi^{FB}$.

Similarly, one can show that having multiple buyers in a joint-production process could also create a negative externality under PB contracts. The performance of PB contracts is lowered when multiple buyers and vendors are involved in the production process because, as the profit gets split among more parties, the effort incentives are diluted. Hence, even though PB contracts have the benefit of tying consulting firms to results (Toppin and Czerniawska 2005), they give rise to a free-riding effect when more parties are involved in the project, potentially creating conflicts among vendors (Table 1). This free-riding effect arising in multisourcing arrangements could, however, be mitigated by a broker or orchestrator firm or by a vendor strategic alliance without diminishing the firms’ ability to demonstrate the value they generate (Czerniawska 2002).

### 5. Conclusions

In this paper, we have developed an analytical framework for contract selection in a collaborative service
Table 2 Contract Challenges and Risk Mitigation Strategies

<table>
<thead>
<tr>
<th>Context</th>
<th>Optimal contract</th>
<th>Challenges</th>
<th>Remedies</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\alpha &lt; \beta$</td>
<td>FF-O</td>
<td>Lack of adaptability</td>
<td>Break down project</td>
</tr>
<tr>
<td>Specification of buyer’s role and responsibilities</td>
<td>Specification of buyer’s role and responsibilities</td>
<td>Hold training workshops/kickoff meetings</td>
<td></td>
</tr>
<tr>
<td>Output measurability and accuracy</td>
<td>Output measurability and accuracy</td>
<td>Specify milestones and checkpoints</td>
<td></td>
</tr>
<tr>
<td>Vendor’s risk aversion</td>
<td>Vendor’s risk aversion</td>
<td>Measure performance with predictable outputs</td>
<td></td>
</tr>
<tr>
<td>$\alpha \gg \beta$</td>
<td>FF-E</td>
<td>Lack of adaptability</td>
<td>Break down project</td>
</tr>
<tr>
<td>Visibility on vendor’s effort</td>
<td>Visibility on vendor’s effort</td>
<td>Improve reputation; develop account teams</td>
<td></td>
</tr>
<tr>
<td>Low incentives on efficiency and effectiveness</td>
<td>Low incentives on efficiency and effectiveness</td>
<td>Add performance incentives</td>
<td></td>
</tr>
<tr>
<td>TM</td>
<td>TM</td>
<td>Visibility on vendor’s effort</td>
<td>Improve reputation; develop account teams</td>
</tr>
<tr>
<td>Low incentives on efficiency and effectiveness</td>
<td>Low incentives on efficiency and effectiveness</td>
<td>Add performance incentives</td>
<td></td>
</tr>
<tr>
<td>$\alpha \approx \beta$</td>
<td>PB</td>
<td>Low incentives in joint production</td>
<td>Contract with large vendors</td>
</tr>
<tr>
<td>Vendor’s limited liability</td>
<td>Vendor’s limited liability</td>
<td>Appoint orchestrator or broker firm</td>
<td></td>
</tr>
<tr>
<td>Outcome measurability</td>
<td>Outcome measurability</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Free riding with multiple vendors and buyers</td>
<td>Free riding with multiple vendors and buyers</td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

The richness of our simple model to understand complex contracting issues.

We also discussed strategies to improve customer satisfaction. In particular, reducing project variability is not necessarily desirable because it may prevent the development of breakthrough ideas. In terms of service design, one should ideally allocate all responsibilities to one party and require the other party to provide enough efforts “to get the job done,” rather than having the output highly sensitive to both parties’ contributions. However, we noticed that task allocation in a collaborative environment could be extremely complex, because of the complementarity of efforts, making the feasible set of contracts nonconvex.

In the future, we plan to analyze the contracting dynamics arising in competitive bidding environments and their implications on service design. In particular, we would like to understand how a service provider such as HCL America (O’Connell and Loveman 1995) should position its service offering (on-site work, off-site work in the United States, off-site work in India) when it is bidding against other vendors so it can win the contract and deliver high value to its clients. Additionally, we would like to decompose the contracting process into multiple phases, such as it arises in IT system implementation, and characterize the optimal contracting dynamics. Our model could also be extended in a multiperiod model to help understand the impact of repeated interaction on contract selection.

6. Electronic Companion

An electronic companion to this paper is available as part of the online version that can be found at http://mansci.journal.informs.org/.

Acknowledgments

The authors thank the associate editor, the referees, Christian Hellwig, Marty Lariviere, and the participants of the Operations Management seminar at the Sloan School of Management for their insightful comments, which have significantly improved this paper.

References


