Reputation Effects and the Limits of Contracting:
A Study of the Indian Software Industry

Abhijit V. Banerjee* and Esther Duflo†

First Draft: November 1998-This version April 26, 1999‡

Abstract

This paper examines evidence of the role that reputation plays in determining contractual outcomes. We conduct an empirical analysis of the Indian customized software industry. We analyze a data set containing detailed information about 230 projects carried out by 125 software firms that we had previously collected. The evidence supports the view that reputation matters. Ex ante contracts as well as the outcome after ex-post renegotiation vary with firms’ characteristics plausibly associated with reputation. We argue that this pattern is not consistent with optimal risk sharing and propose a model of the industry where reputation determines contractual outcomes, whose predictions are consistent with several facts observed in the data. We argue that there is no obvious alternative explanation to the patterns present in the data.

*Massachusetts Institute of Technology. Email: banerjee@mit.edu
†Massachusetts Institute of Technology. Email: eduflo@mit.edu
‡Preliminary and incomplete.

We thank Mathias Dewatripont, David Genesove Dilip Mookherjee, Jean Tirole and Emmanuel Saez for helpful conversations and Paul Joskow for encouragement and support. We acknowledge financial support from the Alfred P. Sloan Foundation, the Shultz fund and the John D. and Catherine Mac Arthur Foundation. We are especially grateful to the software professionals in India, who patiently answered our many questions, and in particular to Mr N.R. Narayana Murthy, from Infosys Technologies Limited, whose help and insights were crucial throughout this project.
1 Introduction

The idea that there are severe limits to what can be achieved through contracting has had an enormous impact on the way economists now think about firms, markets and governments. Correspondingly, there has been a growing emphasis on the role of reputation as a way of counteracting the problems created by the limitations of contracting. While less often emphasized, a view of the world which gives central importance to issues of contracting, reputation and trust, also has important consequences for the process of growth and development. Most importantly, it suggests that the lack of a proper infrastructure for contract enforcement (which makes contracting less effective) and the difficulty of building a secure reputation are potentially important determinants of success in getting out of poverty, along with the more conventional determinants such as human capital and physical infrastructure.

This paper attempts to quantitatively assess the importance of reputation and, by implication, the seriousness of the limits on contracting, in the context of the Indian customized software industry. Customized software is an obvious place to study such effects since the desired end-product tends to be extremely complex and difficult to describe ahead of time in a way that a third party (such as a court) would understand. In fact, typically the parties to the contract themselves do not fully understand what they want till well into the production process. Therefore it seems naive to expect that they could write a contract enforceable by the courts that would fully cover all contingencies that could arise in the production process. Moreover software production does not require very much fixed capital: indeed most firms nowadays simply own a number of PCs (which are cheap and getting cheaper). The rest, including the premises, access to a mainframe and links to a satellite can all be rented. This limits the possibility of the reputation effects that interest us being confounded with the effects of deferential access to capital or the lack of real competition.

The Indian software industry is suitable for such a study for a number of reasons: First, it is an industry which quite large (employing 140,000 people with a turnover of $1.75 billion in 1997-1

---

1 See for example Greif (1993), Baker, Gibbons and Murphy (1995).
2 Stemming from prejudice, or a history of bad performance, as emphasized by Tirole (1996).
3 In India the government has actually invested heavily and by all accounts fruitfully, to make sure that firms have the option of renting expensive fixed inputs (such as expensive computers, building space and equipment for satellite telecommunication), in virtual “Software Technology Parks”. 
98) and growing fast (at an average annual growth rate of 54% over the past six years). Second, its main focus is on exports (more than 60% of its revenue comes from exports) and a large (over 30%) and fast growing share of the exports is customized software. Moreover the industry’s current focus is on expanding the export of customized software relative to its other businesses on the grounds that this is likely to be its best bet for the near future. Consequently the limits of contracting are a major issue in this industry and one that everyone is clearly concerned about. Finally, the fact the contracts are typically across long distances makes contracting more complicated both by making monitoring somewhat harder and, perhaps more importantly, because of the inherent difficulties of international litigation (combined with the deficiencies of the Indian court system). The data we use in this paper comes from interviews of 125 software companies in three major software development centers in India (Bangalore, Hyderabad and Pune)\(^4\). We collected detailed data on the company and on the two last projects they have completed, including what kinds of contracts were initially arranged between them and how the contract got renegotiated as the project evolved (we have a total of 236 contracts in our data set).

Prima facie, the data supports both the view that contracting is very limited and the view that reputation is important. All contracts in our sample are either fixed price contracts or time and material contracts. In fixed price contracts the software firm - henceforth the firm - gets a fixed price and is supposed to pay for all realized costs. In time and material contracts the software buyer - henceforth, the client- is supposed to pay for all realized costs. A large fraction of the contracts do however get renegotiated ex post: the buyer does not pay the entire cost in almost half the fixed contracts and the client pays less than the full amount in about a quarter of the time and material contracts. There is also a simple pattern in both the kind of contract that get chosen and the sharing of the costs which is a result of the renegotiation. It is shown in Figures 1 and 2. Figure 1 shows the fraction of fixed price projects as a function of the foundation date of the software firm, and Figure 2 shows the share of overrun\(^5\) paid for

\(^4\)In each city, we interviewed half of the firms who belong to the software technology park (all exporters do). We selected the firms randomly, but we oversampled the firms that are not fully owned subsidiaries. No firm refused to meet with us and answer the questionnaire. Some appointments could not be arranged to CEO’s unavailability at the moment we were interviewing, and these firms were replaced.

\(^5\)The amount of the project cost that goes beyond the initial prediction.
by the firm as a function of the age of the software firm.\textsuperscript{6} Both are sharply increasing with the starting date of the firm. In particular, firms created in 1994 or after (half of the sample) bear a substantially larger share of the overrun than older firm on average, and the share of overrun they bear is increasing more sharply with age over this range. Measured both in terms of the ex ante contract and in terms of ex post outcome, young firms bear a larger share of the ‘risk’ of each software project.

This effect of age is perhaps the main empirical finding of the paper. We interpret this as an effect of reputation on the grounds that the firms that started in the industry a long time ago and have survived are more likely to be the kinds of firms that clients can trust - the older firms that cannot be trusted are likely to have already gone out of business (since eventually people would have got to know about them).

To provide further support for our interpretation of the age effect as a reputation effect, in section 4 we show that a similar pattern exists when we use other potential measures of reputation such as whether there has been a previous transaction between the firm and the client, whether it is an internal project (i.e. with a client who either owns the firm or has a long-term arrangement with the firm)\textsuperscript{7}, etc...Further, we show that different kinds of reputation are to some extent substitute. For example, the difference between young and old firms disappear among firms that work for an internal client.

In section 3 of the paper we present a simple model based on our observation of the industry which explains why reputation would have the observed effect on contractual outcomes. The basic idea of the model is that in most cases by the end of the project the firm and the client know who was responsible for cost overruns. While this is not contractible, firms and clients could nevertheless benefit from it if they could commit to always follow a certain norm. The norm we emphasize here - clearly there can be other norms that will also work - is that of being reliable: reliable firms always try very hard to ensure that they do not exceed the cost overrun that they had implicitly promised, and pay for any extra overrun when they fail to do so. The problem is that this is typically not consistent with short-run profit maximization by the firm or the client, and can only be sustained if the firms and clients are either innately reliable or, more

\textsuperscript{6}Because the number of firms per year in the sample is small for firms created before 1988, we have grouped all these firms together.

\textsuperscript{7}We will describe this type of structure below.
conventionally, if the particular equilibrium that they are playing induces them to put some value on their reputation. We look at equilibria where a certain fraction of firms and clients are reliable and the rest are not and investigate the implications of a change in the fraction of those who are reliable (interpreted as a change in the average reputation of the firms). The basic trade-off that governs what happens is that fixed price contracts are best for protecting reliable buyers from unreliable sellers while the reverse is true of time and material contracts. Therefore there should be more fixed price contracts if the share of buyers who are likely to be reliable is smaller, which is consistent with the evidence we describe above. We also argue that a number of other predictions from this model are consistent with what we observe.

While we do provide some evidence supporting the broad premises of our model, it is clear that we cannot provide sharp enough evidence to rule out alternative reputation models: it is possible for example that the relevant reputation is for honesty or for a different form of reliability. The objective of this paper is not to distinguish among different kind of reputation. However our reputation-based story does rule out many alternative explanations. In particular it rules out models where there are no agency problems as well as models of agency problems where there is no learning about the firm's type.

Of course, this is all conditional on establishing that we are in fact correctly interpreting the data when we impose the reputation model on it. In other words, it still remains possible that what we are picking up here is the effect of some other variable which happens to be correlated with these measures of reputation. In section 5, we consider some of these explanations. They fall broadly in two classes. First, a class of alternative explanations, which rule out agency problems, explain the differences in the contracts either in terms of differences in risk-sharing or in terms of differences in the production technology available to the firm. We first argue that it is very implausible that the contractual variations that we observe are a result of optimal risk-sharing. The basic point is that in our data set, firms are usually much smaller than their clients and young firms are especially small. It is therefore very hard to understand why firms bear so much of the risk (57% on average) and why especially the smallest and youngest firms bear the most. In response to the view that there are differences in the production technology (essentially that young firms are more incompetent) we point out that the natural effect of such incompetence should be to lower the price the young firms gets paid rather than to make them

---

8There are of course other determinants of the sharing of the risk. We discuss these issues in section 5
bear a lot risk than them can ill afford. Moreover the evidence does not support the view that the differences in competence between the firms is of a magnitude that can explain the differences in the contracts. For example, we present in figure 3 the average overrun as a function of firm’s foundation date. If the high shares of overrun paid by young firms were a way to make them pay for higher overrun, we should see average overrun increasing with age. If anything, the opposite seems to be true. In response to the second class of competing claim, namely that there are agency problems but no learning about the firm, we point out that this conflicts with the evidence on the effects of sources of reputation other than age. Finally, a number of other potential candidates for an alternative explanation of the data are also examined in section 5.

As a final piece of evidence we emphasize the fact that the necessity to build reputation and trust is recognized and is emphasized repeatedly at the industry level as well as by individual firms. For example, the national association of software services companies (NASSCOM) directory of the Indian Software industry has a large section on “quality”. The main element they stress is the number of Indian firms that have ISO 9000 certification or are in the process of acquiring it (ISO certified firms have proven that their software development processes follow approved routines, which is a way for firms to establish a reputation). The association provides technical consulting to any member who wants to get ISO certification. The Indian government provides financial incentives for firms who acquire it. At the individual level, effort to develop a reputation are also obvious.\textsuperscript{9}

This paper is a part of a small but growing number of papers that study the empirics of contractual choice.\textsuperscript{10} Among recent papers Crocker and Reynolds (1993) is most closely related to this work. They examine the determinants of the choice between fixed price contracts and more flexible contracts in US Air force engines procurement. In their view, the key trade off is the following: Fixed cost contracts protect the government against ex post opportunism (in particular it makes it useless for the contractors to claim higher costs) but they require the ability to draft an exhaustive list of requirements (a complete contract), which is possible, but costly. Time and material contracts do not require a truly complete agreement ex ante, but open

\textsuperscript{9}20\% of the firms in our sample already have ISO certification. 13\% are in the process of getting it.

\textsuperscript{10}Monteverde and Teece (1982), Masten and Crocker (1985) and Jossow (1987) are important early papers on this subject. These papers differ from ours in studying settings where there are huge relationship specific investments and very long term relationships are the norm and where the key trade-off is between tightness of contract (or control) and flexibility.
the room for opportunistic behavior by the contractor. Contracts will tend to be fixed costs if
the nature of the engine makes them easy to draft (if the engine is well known or the production
cycle is short), and if the contractor is more likely to behave opportunistically. Their empirical
analysis of a panel of 44 contracts between the government and two contractors confirms these
predictions. Their work shares therefore a central intuition with ours: the reputation of the
contractor does matter for the choice of contracts. The more reputed a firm is, the less likely
it is that the contract will be fixed cost. The central difference is that fixed cost contracts are
not associated with any ex-post cost for the contractor, since fixed cost contracts are “truly
complete agreements”. The government never behaves opportunistically. They don’t discuss
what happens when there are cost overruns in fixed cost contracts, because that is not an
option in their model. In contrast, we recognize the fact that in the software industry the
contract is never complete. Fixed cost contracts need not be more precisely drafted than time
and material contracts. Overrun happens in both types of contracts. The central trade-off is
between containing opportunism by the client and opportunism by the firm.

Lafontaine and Shaw (1996) is another paper that looks at the effect of firm’s age on contracts
(in the context of franchising) and finds that the franchisor’s age has no effect on the contract.
However as they point out (following Mc Affe and Shwartz (1994)), by changing the franchise
contract over time a franchisor runs the risk of hurting its early franchisees (who are locked into
one contract while their competitors get a different contract that perhaps allows them to be more
aggressive). Because of this cross-contract externality, contracts may not change very much over
time, even if the market over time becomes more knowledgeable about the franchisor.

The rest of the paper is organized as follows: In section 2, we describe the institutional
settings in more detail and present a number of basic facts about the production of customized
software. In section 3, we present a model which reflects our understanding of the way the
Indian customized software industry functions, based on two-sided asymmetric information and
reputation effects. The model gives rise to a number of predictions about how reputation shapes
the contractual forms as well as the ex-post renegotiations. In section 4, we provide evidence

\(^{11}\)In their paper, they measure the reputation by the number of litigation conflicts that the contractors had in
the past.

\(^{12}\)This paper is also related to papers such as Barron and Umbeck (1984), Shepard (1993) and Genesove
(1993) which test the implications of theories based on asymmetric information in industrial contexts (but not the
implications for the choice of the contract).
which support the basic assumption and the implications of this model. In section 5, we discuss alternative explanations of the pattern observed in the data. Section 6 concludes.

2 Institutions and basic facts

We begin by describing the sequence of events leading to the off-shore production of a piece of software. The project begins when the client sends a request for proposal to one or more firms. Each interested firm studies the request (this takes the firm 1.25% of the total project cost for the median external project), and submits a proposal, which includes, among other things, a proposed mode of payment and an estimate of how much the client would have to pay. The client chooses a firm, and the firm and the client agree on a contract. The contract specifies an estimate of effort needed to complete the project, a mode of payment, financial details (price, etc...) and a projected schedule for deliverables (which are specific milestones -corresponding to phases of the software development process or to modules of the software- that will be reached in the course of completing the project). The work then starts. The first phase is the writing of specifications. The firm, in collaboration with the user at the client’s end, writes the set of functions that the software will execute. For the median project, it takes 10% of the total project effort to complete this phase. At the end of this part of the project, what the client wants and what it would cost is usually clearer to both the client and the firm and the schedule of deliverables is sometimes amended or clarified.

The second phase of the work is the lower level design, coding and testing of the software. When a specified milestone is reached, the firm sends the deliverable to the client. Each time this happens the client can either acknowledge that it has been delivered (by signing off) or request changes. The firms also send regular status reports to the clients (a little less than once a week on average), keeping the clients up-to-date about the progress of the project.

In terms of project outcomes our main focus will be on overrun: overrun in industry parlance is the difference between the amount of effort actually needed to complete the project and the

---

13 Table 1 shows the descriptive statistics mentioned in this paragraph.

14 Those projects where the client does not own the firm or does not effectively control the part of the firm which is working towards the completion of the project (see below).

15 For some projects, specifications writing and subsequent work are decoupled. One firm—or the client himself—writes the specifications, and another firm completes the project.
estimated effort given in the contract. It is therefore important to be clear about firms mean by an estimate. A standard textbook on software management (Pressman (1997), has an entire chapter on estimation. He describes the process as follows:

'The project planner begins with a bounded statement of software scope and from this statement attempts to decompose software into problem functions that can each be estimated individually. Line of Code or function points (the estimation variable) is then estimated for each function. Alternatively, the planner may choose another component for sizing, such as classes or objects, changes or business processes impacted. Baseline productivity metrics (i.e. line of code per person-month or function point per person months) are then applied to the appropriate estimation variable and cost or effort for the function is derived. Function estimates are combined to produce an overall estimate for the entire project.' (Pressman (1997)).

Our interpretation of this and other material in this book (which is also consistent with what we have learnt from industry sources) is that the estimate is the firm’s best guess about how much effort will needed to complete the project, assuming that the firm’s current understanding of the project is correct and that the firm adheres to its own productivity norms.\(^{16}\) The estimate is therefore clearly not meant to be an unbiased estimate of how much effort the project will actually take. This is important because it tells us that overrun represents the extent of deviation from the firm’s initial plan of action.

This also tells us that overruns ought to be quite common: first, because the needs of the client are typically not very clear at the very beginning of a relationship - even to the client himself. Moreover the client may not put enough effort into understanding and explaining what it wants. Not surprisingly then, the firm often does not understand what the client really wants. When, in the course of the project, the needs of the client eventually become clear, changes have to be made and these are costly. Second, the amount of time and effort needed to design and code a piece of software is difficult to evaluate ex ante, even when the set of functions is well-defined (both for the client and for the firm), and will depend on the type of technology being used, the ability and the experience of the staff of the two firms. Third, with the best of

\(^{16}\)In other words, the presumption behind the estimate is that the firm has understood perfectly what the client wants and that the firm implements the project at its normal level of productivity.
staff and the clearest goals there is also the risk that some unexpected problem arises and delays or destroys the project. Finally, not all firms try their hardest to control costs and delays and one would expect that some projects will end up costing much more than they ought to.

Table 1 shows evidence from our interviews confirming that overruns are indeed common: it turns out that 74% of the projects are completed with a positive overrun. The average overrun amounts to 24% of the initial estimate, and varies a lot (its standard deviation is 34%, and the maximum overrun in the sample is 250%). According to the firms, overruns are due mostly to changes required by the client (these changes cause 48% of the overrun on average). Another 20% of the overrun is due to initial ambiguity in the specifications (i.e. to cases where the firm did not understand what the client really wanted), 8% is due to internal difficulties in the firm (the most frequent one being the loss of the project manager in the middle of the way) and 13% to delay occasioned by the client. Very few projects (less than 5%) are completed with a negative overrun, and that the mean overrun is clearly not zero.

Both firms and clients are, of course, aware of the possibility of overruns. Overruns, apart from being wasteful in themselves (in so much as they could have been avoided by both parties being more diligent), lead to delays which are costly and are a potential source of conflict between the client and the firm (conflicts arise when each side blames the other for the overrun).

Vertical integration and contracts are two ways of limiting the waste due to overrun. Many foreign companies have set up 100% owned subsidiaries in India. These subsidiaries are 100% export oriented, and carry outwork for their mother company and in some cases, for other clients as well. A number of Indian software firms have also entered into arrangements under which the firm dedicates a part of its employees, office space, and computers to a single foreign client. This is what is called an “Off shore software development center” (OSDC). The client sends a

---

17For example, the template of a firm’s contract specifies that “the effort estimates provided for the conversion and testing phases of this project have been provided by the software firm on a best estimate basis. If the scope of the effort changes as a result of discussions during the detailed design phase, the software firm will analyze the impact of changes on the project and may present revised schedules and costs. Changes in schedules and costs resulting from such changes will be reflected by an amendment to this contract.”

18Delays, while rarer than overrun, are far from uncommon in our sample: there are delays in 19% of the cases, and 25% of the cases where there was an overrun.

19Including AT&T, IBM, Microsoft, INTEL, ORACLE, Fujitsu and Motorola.
steady fraction of his software development need to the firm, and is responsible for making use of the facilities devoted to him. This is an effect a type of vertical integration: the OSDC becomes virtually a unit of the client for whom it works regularly. In such cases the interests of the firm and the client are clearly better aligned and while there may be overrun, there is much less reason why the overrun should be wasteful.\footnote{Indeed there may be more overrun in such cases than in general precisely because overrun entails less waste}.\footnote{For example, the client may not need to be very precise about what he wants since he knows that the firm will be happy to do whatever is asked of it.}

Since we are interested in contracts rather than vertical integration our focus in this paper is mainly on external contracts (i.e. contracts that are performed neither within OSDC, nor for the mother companies of the firm). We observe the following types of external contracts: Under \textit{fixed price contracts}, a fixed price is agreed upon up-front, before the specification analysis. These contracts are by far the most frequent: 58\% of external contracts are fixed price contracts. Under \textit{mixed contracts} the price is fixed for the specification phase only at the beginning of the process. The price for the complete project is fixed only when specifications are written and more is known. Typically in such cases the requirement analysis is paid for on a time and material basis, though this is not necessarily the case. Under \textit{time and material contracts}, the entire product is paid for on a time and material basis. This contracts are the least frequent among external contracts (15\%). A striking fact is that there appears to be no “intermediate” contracts: all contracts belong to one of these three categories.\footnote{Or their variants: in some cases property rights in the product substitutes for cash payments.} For example, there are no contracts where the client and the firm agree on sharing the costs.\footnote{Such contracts are observed, albeit rarely, among the procurement contracts for airplane engines studied by Crocker and Reynolds (1993).}

While these contracts predict extreme outcomes in terms of cost-sharing, we actually do not always observe this. It turns out that a large fraction of contracts get renegotiated ex post. This is evident from Table 2, which shows the fraction of overrun paid for by the firm and the proportion of firms that pay all or nothing of the overrun for the three types of contracts. Even in fixed price contracts, the actual overrun is often shared between the client and the firm (in 46\% of the cases) while firms with time and material contracts sometimes pay for overrun (in 22\% of the cases). However it is also clear from the figures in table 2 that the initial contract has an clear influence on which party bears the risk of the project: in fixed cost contracts, firms
bear on average 63% of the overrun, while they bear on average 51.5% in mixed contracts and 15.5% in time and material contracts. Since fixed price contracts dominate our sample, this evidence also implies that firms bear a lion’s share of the overrun (57% on average of external project, 76% for the median external project). Since firms are typically much smaller than their clients this is at least somewhat surprising.

There are several potential explanations for the pervasiveness of renegotiation. First, even when a firm faces a fixed price contract it may have some bargaining power because it usually has the option of walking off the job. If it does, it will not get paid for work that it has already done, but it will also avoid the overrun and at least at early stages of the job, the second effect may dominate. Second, the court system in India is extremely inefficient and going to court is very costly. Firms and clients will therefore prefer to make some concessions in order to avoid going to court. In fact, from our conversations with industry people we have the impression that people go to court very rarely and therefore we ought to expect some renegotiation.

Finally, firms and clients may voluntarily pay for any overrun that is of their own making, because they care about their reputation for being reliable. We had a number of conversations where the CEO of the firm told us ‘it was our fault and we paid for it’. We also have some more indirect evidence that this is at least sometimes the case: as mentioned above, we asked firms questions about who was responsible for the overrun. In what follows, we assume that the firm is responsible for what it described as changes due to ambiguities and overrun caused by internal difficulties. Changes required by the client and delays coming from the client’s side are taken to be caused by the client’s responsibility. Table 3 shows the share of overrun paid by the firm when the overrun is entirely due to the client (column (1)), entirely due to the firm (column (3)), or due partly to both (column (2)). In column (4), we present the coefficient of an OLS regression of the share of overrun paid by the firm on the share of overrun which it caused. In all types of contracts, firms always pay more of the overruns entirely caused by their own mistakes compared to the overrun entirely due to the client. Moreover, in all cases but one, the share of overrun paid by the firm lies in between these two numbers when the overrun is partly caused by each side. Furthermore, the OLS regressions indicate that, regardless of the initial contract, the larger the fraction of the overrun that a firm has caused, the larger the share it has to pay (if a firm causes one additional percent of the overrun, it bears approximately 0.20 percent more of it).
In the next section we present a model of the industry which is based on the picture that emerges from the above discussion. The main elements we wish to capture in our model are the following:

- the high levels of overrun,
- the fact that both sides are responsible for overrun,
- the fact that the software firms end up bearing a large part of the overrun,
- the use of simple ex ante contracts,
- the fact that the contracts get renegotiated ex post,
- the fact that the ex ante contract continues to influence the renegotiated outcome,
- the fact that firms and clients care about their reputation for being reliable and will often voluntarily pay for overrun that is of their won making,
- and the fact that young firms bear, on average, more of the overrun than old firms

3 A Model of the Software Industry

The model we propose in this section is an attempt to capture in as simple way as possible what, on the basis of our experience in the industry, we see as the fundamental structures and conflicts in the Indian customized software industry. The contracting outcomes that will be predicted by the model will, as we shall see, match up reasonably well with what is observed in the data. However, one could come up with other models, or at least combinations of other models, which also explain the data. We will discuss some alternative explanations in the next section. In the end, however, it remains plausible that elements of these other models could also be a part of any comprehensive story of the software industry in combination, perhaps, with the story we tell. In this sense, the model is meant to be illustrative rather than definitive.

The premise of the model is that software projects are prone to cost overruns and that the main conflicts are over the apportioning of these cost overruns. Overruns can happen for two reasons. First, the client could have been insufficiently diligent in delineating his requirements or he could have made a mistake. As a result, when the firm comes up with a product he might
realize that this is not what he wants and demand changes. The firm is, of course, happy to make the changes – since they are Pareto improving – but only if it is adequately compensated. The issue is whether the client will be willing to compensate it enough. Second, overruns could also happen because the firm was either lazy or unlucky in the way it carried out the project.

Since the overrun could come from either side, when there is an overrun, there is a real possibility that each side will blame the other for it. This would not, of course, be a problem if outsiders and specifically the courts can observe who was really responsible. Our assumption will be that this is not possible in most cases.

This is clearly something of a caricature of reality: firms and clients clearly do try to set up systems to ensure that it is clear, ex post, who was to blame for any overrun. The procedure of defining deliverables and having the client sign off on each deliverable is one such system. Once a client signs off on a deliverable, he is to a large extent committed to admit that at least up to that point the firm had done what it was supposed to do. This clearly limits the scope for future disagreements. Nevertheless, there seem to be lots of disagreements and this is presumably ascribable to the fact that even after many milestones have been reached, there remains substantial ambiguity about what exactly needs to be done.

We capture the possibility of this kind of disagreement as follows. The client ($C$) wants the firm ($F$) to build a piece of software that will be worth $V$ to the client. It should normally cost an amount $\bar{y}$ (i.e. the estimate is $\bar{y}$). However with some probability there is an overrun and the total cost is $\bar{y}$. We adopt the normalization that $\bar{y} = 0$ so that all of the cost of the project is overrun.

Overrun is the sum of overrun caused by the firm ($y_F$) and overrun caused by the client ($y_C$). The amounts of the overrun, $y_F$ and $y_C$, are chosen by the firm and the client respectively. Both firms and clients get some private benefits from generating high levels of overrun - this may be because controlling overrun takes effort\(^{23}\) or because the firm (or the client) gets to keep a part of the overrun it has generated (cost padding).\(^{24}\) These private benefits are given by $B_C(y_C)$ for the client and $B_F(y_F)$ for the firm. Both are assumed to have the usual increasing concave

---

\(^{23}\)It requires that the client puts effort into defining its requirements and the firms puts effort into understanding them.

\(^{24}\)In other words, the firm’s effort should thought of as the effort spent by the management on properly organizing the work on the project and not as simply labor time.
shape.

Given these assumptions, the first best outcome has \( y_C \) and \( y_F \) being chosen to satisfy

\[
1 = B'_C(y_C) \quad \text{for the client,}
\]

and

\[
1 = B'_F(y_F) \quad \text{for the firm.}
\]

However, we will assume that both \( y_C \) and \( y_F \) are private information: third parties such as the courts only observe total overrun (\( y_C + y_F \)). Therefore the only enforceable contracts that do not involve money being thrown away ex post can never give the first best.\(^{25}\)

It is however possible to improve on this outcome if the behavior of the firms and the clients is at least partly norm-governed. Specifically assume that there are two types of firms and two types of clients. Of these, one type of firm and one type of client observes a norm of being reliable. Assume, pending discussion, that reliable firms and clients choose whatever level of overrun that they have ex ante promised to deliver. Assume also that by contrast, unreliable firms and clients always choose \( y_C \) (or \( y_F \)) to maximize their current profits. We will later make assumptions and that both \( \overline{y_C} \) and \( \overline{y_F} \) are much higher than the first best level and indeed more than anything that a reliable type would ever choose.\(^{26}\) Moreover this high level of overrun is accompanied by delays and these delays cost the other party an amount \( D \). We assume that \( D \) is large enough that no one will want to contract with someone who is known to be unreliable. Finally let the fraction of reliable firms be denoted by \( \theta_F \) and the fraction of reliable clients be \( \theta_C \) and let the actual type of the firm (whether it is reliable or not) be private information.\(^{27}\)

In this setting, since the reliable firms and clients are going to be self-regulated, the function of the contract is to protect reliable clients against opportunism by an unreliable firms and vice


\(^{26}\)In other words, the private benefits of generating a high level of overrun are extremely high for unreliable firms.

\(^{27}\)We emphasize here the importance of reputation for reliability. Firms, of course, have reputations not just for reliability but also for competence, cooperativeness or honesty. In principle one could build reputation stories based on the idea that it is these characteristics and not reliability that are imperfectly observed that could also explain the data. We focus on this mechanism in writing the model here, because we think it is an accurate description of the most important conflicts in the industry, but it is not to say that we think these other kinds of reputation are unimportant.
versa. Assume for the time being that in the event of an encounter between an unreliable firm and a client (who may or may not be reliable) or between an unreliable client and a firm, the outcome is governed by the ex ante contract. This is the natural outcome since in this case renegotiation can only lead to a redistribution between the two parties and neither party has any reason to give up something for the benefit of the other. To simplify the analysis we assume that the ex ante contract is always linear, i.e. it takes the form of a fixed payment \( P \) to the firm and a share \( s \) of total overrun that is paid for by the firm. When \( s = 1 \) we will describe the resulting contract as a fixed price contract and when \( s = 0 \) we will call it a time and material contract.\(^{28}\)

It remains to say what determines payoffs in a situation where both parties have acted reliably. We will assume that the price \( P \) specified in the contract is always paid so that the only issue that remains is to decide how the overrun gets split.\(^{29}\) If, as we have assumed, reliable firms always follow the norm and deliver the contracted outcome, the contracted division of the overrun is irrelevant in the sense that it has no effect on their choices of \( y_F \) and \( y_C \). However it is clearly easier to enforce the norm if the overrun is split in such a way as to give the two parties the incentive to follow the norm. We will presently show that a rule that has this property is one which says that a reliable firm or reliable client always pays for the overrun it has generated. From conversations with industry people it is clear that this is a rule that a lot of firms do follow. Moreover, in the next section, we will provide evidence that this rule is applied.

\(^{28}\)The actual time material contracts do not have the fixed price component that we have assumed here. Instead firms get paid a markup on realized costs. Taking this into account complicates the analysis (the convenient transferable utility assumption can longer be made) but the comparative statics remain very similar.

\(^{29}\)This is not an entirely innocuous assumption. One could imagine a contract where \( P \) is set very low and then once it is established that both parties are reliable (i.e. at the end of the production process) they renegotiate the price upwards. This would discourage unreliable firms from bidding for the contract. Alternatively one could set a very high \( P \) in order to discourage unreliable clients. In practice we do not observe any instances of this kind of arrangement. This may be because setting a low \( P \) also expands the scope for opportunism. In the setting of our model this does not matter because both parties are risk-neutral but if, as is plausible, they are at least somewhat risk-averse, this kind of contract may be too risky. Moreover there is a serious question as to whether courts will enforce such contracts - the courts may feel that the agreed upon price is unfairly low (or high). Finally, and perhaps most importantly, this kind of contracts relies on a much more inclusive notion of what it means to be reliable: reliable clients are now expected to give up a chance to make a huge amount of money (by refusing to raise the price ex post). It is not clear whether this kind of expectation is realistic.
If we accept that this is the rule for splitting the overrun, it is clear that the ex ante contract will frequently be renegotiated. Moreover renegotiation here will actually enhance ex ante efficiency.

We prefer to remain agnostic about the source of norm-governed behavior that is at the heart of this model. It could be that it is simply an outcome of a repeated game: even in cases where the client and the firm may not expect to transact again they will remain part of the same industry and as long as the information about their past behavior becomes public with some probability, we can expect there to be repeated game equilibria where all sufficiently patient agents honor the norm. Norm-governed behavior could also be innate (see, for example, Bowles (1998) for an evolutionary model of social norms). Finally it is possible that there is a fraction of innately reliable people and a much larger population of opportunists who in equilibrium imitate the behavior of the reliable agents (as in the reputation literature (Kreps et al. (1982), Fudenberg and Maskin (1986))).

Here we will focus exclusively on equilibria where all the people we have called reliable never deviate from the norm - in other words both those who are innately reliable and those who choose to be reliable will follow the exact same norms. By contrast those we have designated unreliable will always follow the behavior we assigned to them above.

These assumptions allow us to treat the types of the firms and clients as fixed parameters and to analyze what happens in a one-shot interaction between the firm and the client. However a complication still remains: the choice of the contract can be used as a signaling device - a firm that plans to be unreliable will prefer a contract where it pays very little of the overrun and therefore, by choosing to absorb most of the overrun a firm may be able to signal that it is reliable. Given that we are in a signaling environment, we will expect that there will be many equilibria. However all such equilibria will involve pooling since in a separating equilibrium all the unreliable firms and clients will never get a contract and therefore will prefer the contract chosen by the reliable type.

\[30\] The simplest way to model this would be to assume that at some point after the transaction between the firm and the client is completed, the fact that one of the parties had acted unreliably becomes public with some probability (a disgruntled employee reports what really happened or an incriminating document gets to the wrong hands). The future play of the game would then be made contingent on such public information. For modeling of social norms along these lines see Kandori (1990).
Among the set of pooling equilibria we focus on the contractual outcome where the utility of the client of the reliable type is maximized given that the firm is getting at least its outside option if it is of the reliable type. This is always a Bayesian-Nash equilibrium (sustained by the belief that only opportunists deviate). The fact that it is also Pareto optimal from the point of view of the reliable types makes it an obvious focal outcome.

3.1 Analysis of the Basic model

Given that we are in a transferable utility setting, maximizing the utility of a client of the reliable type under the constraint that a reliable firm is getting at least its outside option amounts to maximizing the total social surplus calculated from the point of view a pair of reliable types.

This expression for joint surplus is:

\[
W = V + B_C(y_C) + B_F(y_F) - \theta_F y_C - (1 - \theta_F)(1 - s)(y_C + \overline{y_F}) - (1 - \theta_F)D - \theta_C y_F - (1 - \theta_C)s(y_F + \overline{y_C}) - (1 - \theta_C)D.
\]

The fifth terms and sixth in this expression give the total surplus that is lost because in a pooling equilibrium a reliable client must allow for the possibility that the firm is unreliable while last two terms are the surplus that is lost because the firm must allow for the possibility that the client is unreliable.

It is immediate that the choice of \(y_F\) and \(y_C\) that maximize this expression must satisfy, respectively,

\[
\theta_C + (1 - \theta_C)s = B_F'(y_F), \tag{1}
\]

and

\[
\theta_F + (1 - \theta_F)(1 - s) = B_C'(y_C). \tag{2}
\]

The levels of \(y_F\) and \(y_C\) that will be expected of a reliable firm and a reliable client will therefore satisfy these two equations (recall that reliable firms and clients always do what is expected of them and these are the levels that maximize the joint surplus).

Note that these are also the levels of \(y_F\) and \(y_C\) that would be chosen by a reliable firm and a reliable client if there were no explicit expectation about the \(y_F\) and \(y_C\) they chose but they were expected to pay for the share of overrun that they had generated (as long as the other
part y has behaved reliably. To see this note that under these rules they will choose $y_F$ and $y_C$ to maximize, respectively,

$$W_F = P - \theta_C (y_F - B_F(y_F)) - (1 - \theta_C) \left( s(y_F + \overline{y_F}) - B_F(y_F) \right) - (1 - \theta_C) D$$

(3)

and

$$W_C = V - P - \theta_F (y_C - B_C(y_C)) - (1 - \theta_F) \left( (1 - s)(y_C + \overline{y_C}) - B_C(y_C) \right) - (1 - \theta_F) D.$$  

(4)

Maximizing these expression yields exactly the same expression for $y_F$ and $y_C$ as the maximization of the total surplus.

**Claim 1** If we require firms and clients to pay for any overrun that they have caused and reliable firms follow this rule when they deal with other reliable firms but stick to the initial contract when they are dealing with unreliable firms, the level of $y_F$ and $y_C$ that will be the same as those that maximize the joint surplus defined above.

On the strength of this last claim we will henceforth assume that reliable firms and clients actually follow the rule of paying for their own misdeeds unless the other party is unreliable. This gives us a specific rule for sharing the overrun following a renegotiation which (as we have seen) is optimal while being empirically plausible. Moreover it explains the frequency of renegotiation: generically there will be some renegotiation of the initial contract whenever two reliable parties get paired.

Recall next that the condition that determines $y_F$ is:

$$\theta_C + (1 - \theta_C)s = B'_F(y_F)$$

(5)

This gives us $y_F(s, \theta_C)$, with $\frac{\partial y_F}{\partial s} < 0$ and $\frac{\partial y_F}{\partial \theta_C} < 0$.

Likewise, we can derive $y_C(s, \theta_F)$ from the equation:

---

31The initial contract described here is therefore an incomplete contract in the sense that it leaves a lot to be determined through ex post renegotiation. This is in fact our understanding of what the actual contracts look like. One could however imagine a complete contract which mimics this contract: it simply has to say that whenever two parties agree to renegotiate, they report the part of the overrun that they have caused and pay for that part. If at least one party does not want to renegotiate, the initial contract is enforced.
\[ \theta_F + (1 - \theta_F)(1 - s) = B'_C(y_C) \]  

\[ y_C(s, \theta_F) \text{ has the property that } \frac{\partial y_C}{\partial s} > 0 \text{ and } \frac{\partial y_C}{\partial \theta_F} < 0. \]

We state this as:

**Claim 2** The amount of overrun generated by the firm is decreasing in the share of overrun borne by the firm. The amount of overrun generated by the client is decreasing in the client’s share of the overrun. The amount of overrun generated by the client is decreasing with the reputation of the firm, and the converse is true as well.

This result implies among other things that those clients facing time and material contracts should generate less overrun than in a fixed price contract, while firms in the same situation have the opposite reaction. That ought to be intuitive: time and material contracts give high powered incentives to the client while fixed price contracts give high powered incentives to the firm. It is also worth noting that if \( \theta_F = 1 \) (all firms are reliable) then the client will choose the first best level of overrun. If \( \theta_C = 1 \) (all clients are reliable), the firm will choose the first best level of effort. If both parties were reliable, the outcome of the renegotiation is always efficient, which, in a sense, justifies the procedure of renegotiating the contract.

To find the optimal contract we need to maximize the expression for \( W \) given above. Differentiating \( W \) with respect to \( s \) and using the envelope theorem gives us the expression:

\[ (1 - \theta_C)(y_F + y_C) - (1 - \theta_F)(y_C + y_F) \]

If there is to be an interior optimum for \( s \), we must have:

\[ (1 - \theta_C)(y_F + y_C) - (1 - \theta_F)(y_C + y_F) = 0 \]

and further:

\[ (1 - \theta_C) \frac{\partial y_C}{\partial s} - (1 - \theta_F) \frac{\partial y_F}{\partial s} < 0 \]

We have however already seen that \( \frac{\partial y_C}{\partial s} > 0 \) and \( \frac{\partial y_F}{\partial s} < 0 \), so this condition cannot be satisfied. This means that \( s \) cannot have an interior optimum. Intuitively, the advantage of a high \( s \) is that it protects reliable clients by passing off a large part of the overrun onto unreliable
firms. Increasing $s$ raises the level of the overrun chosen by reliable clients and therefore increases the advantage of passing overrun to unreliable firms. The disadvantage of a high $s$ is that it forces reliable firms to pay for the overrun generated by unreliable clients. This disadvantage becomes smaller when $s$ becomes larger because the firm itself generates less overrun. For both these reasons, once $s$ is high, the benefit of increasing it even further goes up and therefore an interior optimum cannot exist.

Given that an interior optimum does not exist, the optimum will be either $s = 1$ or $s = 0$. To see which dominates the other, we need to compare:

$$W(1) = V - \theta_{CYF}(1, \theta_C) - (1 - \theta_C)(y_F(1, \theta_C) + \overline{y_C}) - B_F(\overline{y_F}(1, \theta_C)) - \theta_{FYC}(1, \theta_F) - B_C(y_C(1, \theta_F))$$

and:

$$W(0) = V - \theta_{CYF}(0, \theta_C) - B_Fy_F(0, \theta_C) - \theta_{FYC}(0, \theta_F) - (1 - \theta_F)(\overline{y_F} + y_C(0, \theta_F)) - B_C(y_C(0, \theta_F))$$

While either of these could be larger, it is clear that the difference $(W(1) - W(0))$ is increasing in $\theta_C$ and decreasing in $\theta_F$. This observation gives us the following result.

**Claim 3** The optimal contract is always either a fixed price or a time and material contract. It is a fixed price contract when most clients are reliable while firms are more likely to be opportunists, and a time and material contract in the reverse situation.\(^{32}\)

\(^{32}\)The derivative is $\overline{y_F} + y_F(0, \theta_C)$.

\(^{33}\)The derivative is $-y_C(1, \theta_F) - \overline{y_F}$.

\(^{34}\)This result clearly depends on the assumption of risk neutrality. The extreme contracts predicted here clearly do not make for optimal risk-pooling and if the firm and the client were sufficiently risk-averse they would surely want a more intermediate contract. In the face of this, the fact that no intermediate contracts are observed suggests either that the people are relatively risk-tolerant or that the contract is also influenced by things that we have not modeled. For example, as soon as the client agrees to pay a share of the overrun it would have to set up a system for determining the total overrun. Likewise, if the firm pays a positive share it would have to invest in a system for measuring overrun. If this kind of measurement has a fixed cost, the optimal contract may well be an extreme contract.
This confirms the intuition given in the introduction that fixed price contracts are instituted to protect clients against opportunism, while time and material contracts protect firms. Firms that have high reputation will get time and material contracts while the rest of the firms will not.

3.2 Extensions

3.2.1 Introducing some discretion at the level of the court

The model of the previous section implicitly assumes that, when a contract goes to court, the court simply enforces the basic sharing agreement and ignores any other clauses written into the contract. In reality, courts certainly exercise a fair amount of discretion and this gets reflected (for obvious reasons) in out of court settlements as well. This possibility can easily be introduced into our model and, as we will see, does not change any of the results reported so far - though it adds some additional nuances to our analysis of the level and variability of cost overruns. We introduce this possibility by assuming that when the contract goes to court, the court implements an outcome which is a convex combination of the original contract and some “fair outcome” that the court determines.\(^\text{35}\) We will assume that the share paid by the firm in the fair outcome is a decreasing function of the firm’s reputation and an increasing function of the client’s reputation. In other words, the share of overrun paid by the firm when they go to court is written as:

\[
s^*(s, \theta_C, \theta_F) = s(1 - \rho) + \rho s_c,
\]

where \(s_c = s_c(\theta_C, \theta_F)\) with \(0 < s_c < 1\) and \(\frac{\partial s_c}{\partial \theta_C} > 0\) and \(\frac{\partial s_c}{\partial \theta_F} < 0\).

This formulation recognizes that when courts have to decide which of the two disputant is lying they give weight to their past reputation.

It is easily checked that equations (1), (2), (5) and (6) remain valid with \(s^*\) in place of \(s\). In other words we now have the functions \(y_F(s^*, \theta_C)\) and \(y_C(s^*, \theta_F)\) which exactly parallel the functions \(y_F(s, \theta_C)\) and \(y_C(s, \theta_F)\) that we had before. Claim 3 therefore continues to hold. Moreover the problem for the choice of an optimal contract continues to generate only extreme

\(^{35}\)Or equivalently, the court sticks for the original contract with some probability and chooses some other “fair” repartition in the remaining cases.
solutions. The only difference is that while the optimal \( s \) is still either 1 or 0, the resulting \( s^* \) will be \( s^*(1) = (1 - \rho) + \rho s_c(\theta_C, \theta_F) \) or \( s^*(0) = \rho s_c(\theta_C, \theta_F) \).

We now define:

\[
W^*(s^*, \theta_C, \theta_F) = V + B_C(y_C) + B_F(y_F) - (1 - \theta_F)(1 - s^*(s, \theta_C, \theta_F))(y_C + y_F) - \nonumber
(1 - \theta_F)D - (1 - \theta_C)s^*(s, \theta_C, \theta_F)(y_F + y_C) - (1 - \theta_C)D.
\]

These are the expressions for total surplus for a given value of \( s^* \). Now,

\[
\frac{\partial(W^*(s^*(1), \theta_C, \theta_F) - W^*(s^*(0), \theta_C, \theta_F))}{\partial \theta_F} = -y_C - y_F + \frac{\partial W^*(s^*(1), \theta_C, \theta_F)}{\partial s^*} \frac{\partial s^*}{\partial \theta_F} - \frac{\partial W^*(s^*(0), \theta_C, \theta_F)}{\partial s^*} \frac{\partial s^*}{\partial \theta_F}.
\]

Of these the first two terms are of course negative. The third term is also negative, since increasing \( \theta_F \) reduces \( s^* \) and reducing \( s^* \) reduces \( W_F + W_C \) when \( s \) is in the neighborhood of 1 (by the quasi-convexity of the \( W_C + W_F \) function). The last term is also negative because increasing \( \theta_F \) reduces \( s^* \) and reducing \( s^* \) increases \( W_C + W_F \) when \( s \) is close to 0. Therefore increasing \( \theta_F \) reduces \( W_C + W_F \) when \( s \) is close to 0. Therefore, increasing \( \theta_F \) favors moving to a time and material contract. A parallel argument shows that raising \( \theta_C \) favors moving to a fixed price contract.

This is as before. However, the adoption of discretion on the part of the court changes things by allowing the share of overrun in the case of a dispute, conditional on having a time and material or fixed cost contract, to depend on the firm’s (and the client’s) reputation. The share of the overrun paid by the average firm with reputation \( \theta_F \) that works for a client of reputation \( \theta_C \) and has a fixed price contract, is:

\[
\theta_F \theta_C \frac{y_F(\theta_F, \theta_C)}{y_F(\theta_F, \theta_C) + y_C(\theta_F, \theta_C)} + (1 - \theta_F \theta_C)(\rho s_c(\theta_F, \theta_C) + (1 - \rho))
\]

As \( \theta_F \) goes up there are three effects: First, \( \frac{y_F}{y_F + y_C} \) increases, for two reasons: \( y_C \) goes down when \( \theta_F \) increases, and raising \( \theta_F \) reduces \( s^* \) and thus increases \( y_F \). Second, \( s_c(\theta_F, \theta_C) \) goes down. Third, when \( \theta_F \) goes up, it shifts weight from the second term to the first term in the
above expression. Since we think of \( s^* \) as being close to \( s \), and therefore close to 1, we expect it to be higher than \( \frac{y_F(\theta_F, \theta_C)}{y_F(\theta_F, \theta_C) + y_C(\theta_F, \theta_C)} \) and therefore the net effective should be negative.

The last two effects imply that when \( \theta_F \) goes up, the share of overrun that the firms pays goes down while the first effect goes the other way. The negative effect will dominate as long as the \( y_F(\theta_C, \theta_F) \) and \( y_C(\theta_C, \theta_F) \) functions are not too elastic, which seems plausible.

**Claim 4** The firm’s share of the overrun (conditionally on having a fixed price contract) is likely to be decreasing in its reputation.

If we now look at the mean overrun generated by a firm of reputation \( \theta_F \) and a client of reputation \( \theta_C \), this is given by the expression:

\[
\theta_F y_F(\theta_F, \theta_C) + (1 - \theta_F) \overline{y_F}.
\]

When \( \theta_F \) goes up, \( y_F \) goes up but since \( \overline{y_F} \) is much greater than \( y_F(\theta_C, \theta_F) \), the shift in weight from \( \overline{y_F} \) to \( y_F(\theta_C, \theta_F) \) has a negative effect on the overrun. Unless \( y_F \) is very elastic with respect to \( \theta_F \) we would expect the second effect to dominate: this tells us that the average amount of overrun caused by the firm should fall as the firm’s reputation gets better. On the other hand, since \( \overline{y_F} \) is fixed while \( y_F(\theta_C, \theta_F) \) goes up when \( \theta_F \) goes up, the dispersion of overrun must be lower for more reputable firms and as a result the variance of overrun generated by the firm will also tend to be lower for more reputed firms.\(^{36}\)

**Claim 5** Mean overrun caused by the firm is lower for more reputed firm. The variance of overrun caused will tend to be lower as well. A parallel result holds for more reputed clients.

Note however that since \( y_C(\theta_F, \theta_C) \) is decreasing in \( \theta_F \), the average overrun of clients dealing with more reputed firms, \((1 - \theta_C)\overline{y_C} + \theta_C y_C(\theta_F, \theta_C)\) is lower when the firm is more reputed. Moreover since the dispersion is larger and the mean overrun is lower, the variance of overrun generated with the client will grow with firm’s reputation.

**Claim 6** The mean overrun generated by a client goes down with the firm’s reputation but its variance goes up. The same hold for the overrun generated by the firm when the client becomes more reputable.

\(^{36}\)A sufficient condition for this to happen is that \( \theta_F \) is greater than 1/2, which seems plausible.
It then follows that:

**Claim 7** The mean total overrun goes down when the firm becomes more reputable, however the variance can go up or down or remain constant.

### 3.2.2 Choice of Projects

The fact that firms with low reputation pay for most of the overrun, should clearly influence their choice of projects. This can be introduced into our model by making the plausible assumption that the most rewarding projects (the ones with the highest $V$) will also have the highest possibility of large overruns ($\overline{y_C}$ and $\overline{y_F}$ are going to be large). It is easy to show by introducing this assumption into the model of the previous sub-section, that keeping the reputation of the client fixed, firms that are facing a fixed price contracts will be more willing to trade off a lower $V$ for a lower $\overline{y_C}$ than firms which face time and material contracts. Therefore low reputation firms will want projects where client side opportunism is limited even at the cost of a lower $V$.

It does not immediately follow however that the low reputation firms will get these projects: low $\overline{y_C}$ projects are probably also low $\overline{y_F}$ projects and low $\overline{y_F}$ projects are attractive to clients who are facing time and material contracts (i.e. clients who are working with high reputation firms) for precisely the same reason that the firm wants a low $y_C$ project. Therefore high reputation firms may be able to bid away these projects from low reputation firms. In equilibrium however this is unlikely to be true: the difference lies in the fact that for a fixed $\theta_C$, time and material contracts will only be chosen when $\theta_F$ is relatively large and overruns caused by the firm are relatively unlikely. Therefore the saving in cost by choosing a low payoff project is going to be relatively small.

**Claim 8** Low reputation firms will tend to be specialized in projects which have low potential for client-side opportunism.

In terms of what we observe, this seems to suggest that low reputation firms will choose projects which are simple and well-understood so that the client does not have to do very much work to make clear what he wants. These projects can be either short projects\textsuperscript{37}, or projects

\textsuperscript{37}The complexity of a software project increases sharply with its size, see Pressman (1997).
where the main goal is easily defined. Year 2000 projects are typical in this respect.\footnote{Other projects where the objectives are relatively easily defined include CAD projects and migration of an existing software from a platform to another.}

Adding the possibility of choosing projects however makes it more likely that the results on the relation between reputation and the mean and variance of overrun will be ambiguous. Because firms with low reputation can choose projects so as to limit the overruns, both the mean and the variance of the overrun generated in the projects of low reputation firms may be lower than that generated in the projects of higher reputation firms.

3.2.3 Reputation building

If seems plausible that most firms with a low reputation will want to build a reputation for reliability. One can imagine them trying to signal their reliability by being extremely careful about not generating any overrun. This would have the effect of reducing the mean overrun generated by low reputation firms and to increase its variance, since the dispersion between the reliable and the unreliable goes up. It would also weaken the relationship between reputation and the share of the overrun generated by the firm.

3.2.4 Incentives for opportunists

We have assumed so far that opportunists cannot respond to incentives. This is clearly extreme, especially since the incentives can be quite strong. If we allow the opportunistic firms to reduce the overrun that they cause when \( s \) is high, both the mean and the variance caused by the firm will have less of a tendency to fall as the firm’s reputation improves.

4 Evidence

In this section, we document that the central implications of the model are consistent with the data, by showing that contractual forms as well as the actual sharing of the overrun vary with characteristics likely to be correlated with the reputation of the firm. We then examine how the other predictions of the model match with the software data. Finally we consider some obvious alternative explanation of the patterns observed in the data.
4.1 Sources of reputation

We begin by describing alternative sources of reputation for the firm. There are a number of ways in which a firm can acquire a reputation for being reliable. We list them below (the relevant data is presented in table 4:39

First, we have already argued (in the introduction) that we will think of age as an important source or reputation. Firms which manage to survive in the market for a long time are more likely to be “good firms”. In the framework of our model, it means that they are less likely to act unreliably. While we do not formalize the process which leads to the elimination of unreliable firms from the pool of firms, it would be straightforward to do so. Essentially, if unreliable firms are discovered with positive probability and nobody wants to deal with them after they are found out, then as firms age, the probability that they are unreliable will decline.

Second, ISO certification can potentially also give a reputation to a firm. ISO certification is awarded by international or Indian agencies, themselves accredited, which examine that the processes of software production in the firm follows some approved routines. In particular, the firm must follow specified procedures to report on the progress of the software and to perform the tests. Consequently, the software development process should be easier to monitor for ISO certified firms. Moreover, ISO certified firms are monitored every once in a while, and lose the certification if they cannot prove that they followed the approved methods. This should give every incentive to the ISO certified firms to stick to standard procedures and report problems reliably. As we noted in the introduction, firms in the industry are currently very keen to acquire ISO certification, precisely because they think that it will improve their reputation. 19% of the external contracts in the sample were done by ISO certified firms. ISO certified firms tends to be older firms (only 9% of the young firms have ISO certification).

Third, trust established in repeated relationships can play the same role within a relationship

39Note that we think of the reputation as being a attribute of the firm, more than of the individuals who compose it. It could be that an experienced professional leaving his job to create a software firm takes his individual reputation with him. It turns out that individual reputation seems difficult to transport (we asked what the past career of the person who founded the software firm was, and examined whether this was related with sharing of the overrun, but did not find that this was the case). The main reason is that the important input the boss of a software firm has to provide is the management of the team, which may or may not be related to his ability as a software professional.
that reputation plays on the broader market. When a client has gone through a project with a firm once, he has gathered information about it, and therefore, if he decides to work with the firm again, he must have a better opinion of it than the average opinion it has on a firm of the same type. This will include an opinion about the general ability of the firm (which will certainly affects contracts and the choice of projects in many ways), but also a better idea of its reliability, which will play within the relationship the same role as a reputation. 41% of the contracts in the sample involved a client with whom the firm had worked already. This proportion is roughly the same among young and old firms.

Finally, we contrast internal (projects for OSDC and mother companies) and external projects. The informational problems that we mentioned are greatly mitigated in internal relationships. Therefore we should find the same kind of differences (but perhaps even stronger) between internal and external projects as between projects of old and young firms. However we need to be sensitive to the fact that companies working for internal clients are potentially very different from other companies. In particular, OSDC will be established only after the client has spent a very long time studying the firm. Fully owned subsidiaries are often run by people who had been previously working in the US office of the firm. We therefore restrict the comparisons to firms that perform some internal projects (e.g. subsidiaries that works for their mother company and also for external client). This insures that the selection of firms for internal projects does not invalidate the comparison (since all firms in this subsample have been selected for some internal work).

4.2 Choice of contract and sharing of the overrun

4.2.1 Structure of the contracts

An implication of the model is that contractual forms will be restricted to contracts where the ex ante rule is that firms will bear either all or nothing of the cost overruns. This implication rests on the particular assumption we have made (in particular on the fact that we don’t allow for risk aversion), but it matches well with the observed pattern.

As we describe in section 2, there are three major types of contracts: fixed cost, time and material and “mixed” contracts. Fixed cost contracts are linear contracts with \( s = 1 \). As we discuss above time and material contracts are similar to such contracts, with \( s = 0 \). In mixed
contracts, the initial agreement specifies a payment for the specifications only. At the end of the specification phase, another agreement is specified for the development and testing phases. This kind of contracts effectively splits the projects into two subprojects. For each of them, a separate sharing rule is chosen, which is either 1 or 0 (often, time and material for the specification phase and fixed cost for the subsequent work).

In other words, mixed contracts are a juxtaposition of fixed cost and time and material contracts. It is easy to understand why, when the project is broken into these two phases, specifications tend to be written on time and material and the rest of the work tends to be done using a fixed cost contract. In the specification phase, the potential for the client to generate an overrun is extremely large. In particular when the firm first sends the specifications, he can pretend that the specifications written do not correspond to what he wanted. The whole effort of the firm until that point becomes in effect useless. Therefore, it is important to give the client higher powered incentive. On the other hand, at the time the second sub-contract is written, a large part of the uncertainty about what the client really wants is resolved, since it has agreed (in writing) to the specifications. Therefore a fixed cost contract, which give better incentives to the firm, can become optimal from that point on. In practice the choice of the contract for the second phase of the project is often endogenous: if the firm feels that a substantial amount of uncertainty remains, it can in general insist on getting a second time and material contract. Mixed contracts are therefore ex ante more constraining for the client than for the firm.

4.2.2 Reputation and the choice of contract

The reputation of the firm determines both which contract it will get (choice of $s$) and what share of the overrun it will end up paying (actual $s^*$). Firms without a reputation will be more likely to have fixed cost contracts than time and material or mixed contracts. Conditionally on having fixed cost contracts, firms without a reputation will bear more of the overrun than firms with a reputation. The combined effect of the two is of course that firm without a reputation will bear a larger share of the overrun.

This subsection presents data related to these implications.

We presented evidence that age does matter in the introduction, as a motivation for this project. It is illustrated in figure 1 and 2. The proportion of fixed cost contracts and the share of the overrun borne by the firm are increasing with the foundation date of the firm. Table
5 shows the means of the firm’s share of the overrun for each type of firm, and the difference between low and high reputation firms. In column (1), we report the mean for the sample of external firms. In columns (2) to (4), we show the contrast between young firms (created in 1994 or after) and old firms (created in 1993 or after). Young firms are significantly more likely to have fixed cost contracts (the probability is 26% higher). They bear substantially more of the overrun (19%). This holds within the projects with fixed cost contracts (the difference is 13%).

The pattern is less clear for ISO certification: ISO-certified firms are not less likely to get fixed cost contracts and they do not pay for a lower fraction of the overrun in general. However, conditionally on doing fixed cost contracts however, they bear less of the overrun (20.4%).

A relationship with a client has the same effect of a general reputation. Firms engaged in a repeated relationship with their client are about as likely to than other firms to have fixed contracts, but they pay significantly less of the overruns (20% less).

Finally, among firms who have internal contracts, firms pay for more of the overrun when they deal with external clients than when they deal with internal clients. Almost half of their external contracts are fixed cost contracts (a number close to the proportion of fixed cost contracts among old firms), whereas only 23% of the internal contracts are fixed price contracts. They pay a much smaller share of the overrun (20% instead of 47%) in internal contracts than in external contracts. The difference conditional on doing fixed cost projects is not significant, but this is probably due to the small number of fixed price contracts among internal projects.

In summary, it seems that young firms, firms working with a new client and firms working with an external client bear a larger share of the overrun compared respectively to older firms, firms engaged in a repeated relationship and firms working for an OSDC or their mother company. We interpret these results as showing that reputation does influence the way the overruns are shared between the client and the firm. We will address some alternative explanations below, but the first possible caveat to this interpretation is that these firms do different types of projects, which require different types of incentives or entail different types of risk. For example if old firms do mostly projects that entails the possibility of very large overruns, they may refuse doing the project unless they know they will be covered in case this happens.\footnote{We will comment more on the choice of project per se below.} In particular, table 4 shows that young firms, non ISO-certified firms, and firms working for external clients do
on average smaller and simpler projects than old firms, ISO firms and firms working for internal clients. It is therefore important to check that the simple contrast between the groups is not an artifact of the different composition of their contracts.

In table 6, we show the differences between the overrun paid for by each type of firms in project-size cells (panel B) and complexity cells (panel D). The first panel reproduces the uncontrolled difference of table 5. In the panel C, we show the “controlled contrast”: this is simply a weighted average of the differences between the young and old firms in the project size cells, where the weights are given by the fraction of projects falling into this project size cell. This is a crude way to take into account the two facts that different types of firms choose different type of projects and that the differences across young and old firms are not necessarily the same for all project sizes.

Firms tend to bear less of the overrun when they do complicated project than when they do simple projects. There is also a weak relationship between the size of the project and the share of the overrun paid for by the firm. Young firms pay a larger share of the overrun than old firms for small and for large projects, but not for medium-sized projects. The controlled contrast between young and old firms is slightly smaller than the simple difference, but still high. The controlled contrast becomes positive, though insignificant, for ISO-certified firms, mainly because that ISO-certified firms doing small projects don’t pay any of the overrun.

Controlling for project size does not affect the difference between repeated and new clients and between internal and external contracts. Whatever the complexity of the project, young firms bear more overrun than young firms, firms working with new client bear more overrun than firms working with a repeated client, and firms bear less overrun when they do internal projects. The evidence for ISO certification is, once again, mixed.

In summary, even after taking into account the size of the projects, firms with low reputation bear more of the overrun than other firms (although the evidence in favor of ISO certification remains less than overwhelming). A final piece of evidence is presented in table 7. In this table,

---

41We used the subjective complexity measure given by the firms.
42Following is a simple example: if old firms pay the same share of the overrun than old firms for small and for large projects, but both pay more for small projects than for large projects, and if small firms are more likely to do small projects, then the simple contrast would be positive due to the selection, whereas the controlled contrast would be 0.
43This number should be taken with caution, as very few ISO firms do small projects.
we examine whether different kinds of reputation are good substitutes. Namely, we ask in panel A whether young firms still bear more of the overrun when they benefit from another kind of reputation. In the table we present the difference between young and old firms in the proportion of fixed cost contracts (line 1) and in the share of overrun they pay (line 2) within groups of ISO-certified /non ISO-certified firms, repeated/new clients, internal/external contracts (for firms who do some jobs for internal firms). The contrasts are interesting. Non ISO-certified young firms bear 27% more of the overrun than non ISO-certified old firms, but among ISO-certified firms, there is no difference. Young firms are significantly more likely than old firms to have fixed cost contracts if they work with a new client, but not if they have already worked with this client. Among firms that do some internal contracts, the same contrast appears: young firms are more likely to have fixed cost contracts when they work with an external client, but not when they work with an internal client or an OSDC. Finally, the difference between young and old firms in the share of the overrun the firm pays is larger for contracts with new clients compared to contracts with repeated clients and for contracts with external clients compared to contracts with internal clients. In panel B, we perform the same exercise, but we look at how the difference between the share of the overrun paid for by firms working with a new rather than repeated client varies across different kind of firms. Interestingly, a very different pattern emerges. The difference between new and repeated client subsists for old firms and for ISO certified firms, and does not decline. It suggests that the mechanism of reputation formation is rather inefficient: even after a firm has been in the market for some time, much remains to be learnt about the it.

We have documented systematic differences in the way cost overruns are shared across young and old firms, contracts with repeated and new clients, ISO certified firms and other firms, and firms in internal and external contracts. This evidence is consistent with a model where reputation is an important determinant of the contracts and the sharing of the overrun. In the next subsection, we examine the whether the other empirical predictions of the model also hold.
4.3 Further results

4.3.1 Choice of project

A simple extension to our model also predicts that the firms with a low reputation will tend to choose simpler projects, where the objectives are easier to define, which will tend to limit the over-run generated by the client. We present evidence relating to the choice of project in table 4, and figure 5, 6 and 7. Young firms do smaller projects, which have smaller overruns (even expressed in proportion of predicted costs).\textsuperscript{44} They also tend to carry more often “simple” projects (Y2K, CAD, Data manipulation), which generate lower overruns, are easily defined, and are easier to monitor. We have also asked them to subjectively rate the complexity of the project, and even according to this subjective measure, young firm do more simple projects. As a result of these two combined facts, the returns to each project (cost multiplied by markup) is smaller on average for young firms than old firms.

This could be at least partly explained by the fact that young firms are on average less competent (and that therefore clients do not want entrust them with large or complex projects). However the same contrast holds between internal and external projects. Since we have restricted the comparison to firms that do at least some internal work, the difference between internal projects and external projects is not tainted by this bias. This confirms that part of this difference between young and old is due to difference in behavior.

4.3.2 Mean and variance of the overrun

As a result of the different choices of projects, the model’s predictions about mean and variance of the overrun are not unambiguous. The mean of the overrun generated by the firm and of the overrun generated by the client could decrease or increase with reputation. The basic model implies that overrun generated by the firm should, on average, be larger for firms with lower reputation (and so should total overrun), but this could be compensated by the fact that firms with low reputation choose on average simpler projects. The variance of the overrun generated by the firm is more likely to decrease with the reputation of the firm (especially if reliable firms with low reputations try to signal their type), but the model does not make any prediction about

\textsuperscript{44}Moreover by doing that they keep the share of overrun accounted for by each project more or less similar across young and old firm: this could therefore be explained by adding risk aversion to the model.
the variance of the total overrun. As we will see below, alternative models would on the other hand have strong predictions on the mean or the variance of the overruns.

In table 8, we present the mean and the variance of the overruns caused by the firm and of the total project overrun, unconditional and controlling for project size. The mean and the variance of the total overrun do not seem to vary with firms reputation in any systematic way. If anything, overruns are actually smaller for young firms, non-ISO firms, and external projects (but except for the last one, these differences are not significant). The variance of the total overruns are similar. If we restrict attention to the share of the overrun due to the firm (internal difficulties and ambiguities), young firms do seem to cause slightly more of these overruns. The difference between overrun caused by young and old firms, controlling for project size, is 2.6 %. (the sample mean of overrun caused by the firm is 7.67 %). Overruns caused by the firm are also smaller for projects realized for a repeated client. The variance of the overrun caused by the firm is also larger for young firms and firms working with a repeated client. These difference are significant. These two facts (small differences in mean of the overrun caused by the firm but large difference in variance) are consistent with our model, extended to take into account reputation building concern and choice of project.\[45\]

The data seems consistent with our model of how reputation determines contractual outcomes. In the next section, we examine the most obvious alternative explanation to the observed pattern.

5 Alternative interpretations of the data

This section reviews alternative explanations to the pattern observed in the data (in particular to the main result that young firms bear a larger share of the overrun than old firms).

5.1 Pure Risk Sharing

One possible interpretation of what is going on this industry is pure risk sharing. However, as explained in much greater detail in a previous version of this paper, this interpretation very quickly runs into trouble. In the case where we assume CRRA preferences (which is standard

\[45\] Once again, we find that ISO firms do not perform particularly well: they generate large overruns than non ISO firms (although the variance is similar).
in cases, like this, where there is substantial variation in the size of the contracting parties) we showed in the previous version of this paper that

Claim 9 If the firm and the client have CRRA preferences, for a fixed project size, the share of the risk that they each bear will be approximately in the inverse proportion of their coefficients of relative risk-aversion, keeping fixed the ratio of their total revenues. It will also be approximately in the direct proportion of their total revenues, keeping fixed the ratio of their risk-aversions.

Given that the client’s revenue are much bigger than that of the firm, an implication of his proposition is that the client should bear most of the risk unless the client is much more risk averse than the firm. In fact, the firm bears on average more than half of the cost overrun, suggesting that the client’s coefficient of risk-aversion must be very large relative to the firm’s. It is however difficult hard to think of a basis for such differences in risk aversion.

Moreover, this result has systematic predictions about the relationship between firm size and the share of the risk that it bears, controlling for client size and project size.

Table 9 presents the share of overrun paid by the firm by client size, project size and firm age. In all project size-client size cells, old firms bear less of the overrun than young firms. Since old firms are on average larger (this is shown in figure 4 and in columns 2 to 4 in table 4: old firm’s turnover is larger by $3.7 million, or more than 100%, than young firms), this contradicts the basic implication of the risk-sharing model.

One might also speculate that old and new firms generate different risk profiles and that this explains why old firms bear less risk: perhaps old firms simply generate less risk. However, the evidence on the standard deviation of total overrun presented in table 8 shows that this is not the case. The standard deviations of total project overrun are very similar across all types of firms. There is therefore no evidence to support the view that young firms are systematically

---

46 More than half of the contracts in the sample are with "large" clients, 26% are with medium sized client, and 19% are with small clients. Large clients are in general fortune 500 companies or equivalent. Small clients are firms with turnover below $10 million. In contrast, the median turnover of the software companies in the sample is only $1.2 million, and the largest firm had a turnover of $47 million. Only 27% of the firms have a turnover above $10 million. Among the firms engaged in contracts with small clients, the median firm has a turnover of $0.5 million.

47 To maintain legibility and avoid constructing cells few observations, we have divided the projects into larger and greater than median, grouped small and medium clients, and separated the firms between "young" (created in 1994 or later) and old (created in 1993 or before).
more risky to deal with. Another possibility, however, would be that the underlying distribution of overrun is different for young and old firms (despite the fact that mean and standard deviation are different), and that the particular form the the risk faced by old firms made this particular risk sharing rule optimal. We examined the entire distribution of overrun generated for both young and old firms. The two distributions are very similar, except for four old firms which generated very large overrun (150% and higher). These four old firms are however not driving the results, since all of them paid 100% of the overrun. Moreover, as we have shown above, the difference between old and young firms is maintained when we control for project type (complexity or size), which are presumably good indicators of project-specific risk.

The evidence we give above strictly only applies to the case of CRRA preferences. There are, of course, many classes of risk preferences which do not fall into this category. However, there are two basic intuitions which suggest that these other preferences will not work particularly well either: on the one side, if the coefficient of absolute risk aversion falls faster than a CRRA, it is very hard to explain why the firm bears any risk at all. On the other side, if the coefficient falls slower than a CRRA (so that the preferences approach the CARA model), it can be shown that project size and client size effects also become smaller and this leaves very little to explain the inter-firm differences.\footnote{In the extreme case of CARA preferences neither project size nor client size affects risk-sharing.}

### 5.2 Varying Levels of competence

Our model has assumed that both clients and firms are risk-neutral. Suppose we now assume that firms and clients make mistakes which lead to overrun but that these mistakes can be contracted upon. In this case, one possible first best contract is one in which firms take the full responsibility of any mistake that they make.\footnote{Of course the actual contracts do not say anything about dividing the overrun. Therefore what we are referring to here is a fully efficient implicit contract.} Now if young firms\footnote{Or more generally, firms that we have called so far “low reputation” firms.} are on average less competent, then it is to be expected that they would pay on average for more of the overrun.

The first point against this explanation is simply that risk-neutrality is a very extreme assumption. If the firm was at all risk averse then the optimal contract would try to insure the firm against all sources of risk that are beyond its control. Therefore, since young firms do not choose to be incompetent, they should be insured against overrun that results from their
mistakes. Of course, the extent of such insurance may well be limited by the client’s willingness to bear risk, but as we have already argued, the client is in a much better position to bear risk than the firm, and in particular small firms should only bear a small part of the risk. Of course, this assumes that the mistakes are not made deliberately. The case of varying levels of moral hazard will be examined below.

There are also some simple empirical arguments against this view: first (and most importantly), firms pay much more of the overrun than the share for which they are responsible: as shown in table 3, even when the client is fully responsible for the overrun, the firms still pay on average 51% of it. Second, to explain the differences in the share of overrun paid by the firms entirely by differences in the share of overrun that is caused by the firm, it would have to be the case that young firms cause substantially more overrun than old firm do. Recall that the total overrun are, if anything, larger for old firms than young firms. The difference between the share of the overrun due to young firms and the share of overrun due to old firms is clearly not large enough to explain difference in the sharing of these additional costs. Differences in sources of the overrun are shown in table 8. If we look at the difference between the fraction of overrun due to young firms and that due to old firms, it is only 1.18 (and not significant)\(^{51}\), but they pay 20% more of the overruns. Finally, even within firms that do some internal work (and are therefore more homogeneous), it is the case that firms bear more of the overrun in external contracts than in internal contracts. Therefore it does not seem to be the case that the differences in the share of the overrun borne by young firms can be explained by systematic differences in competence between young and old firms.

5.2.1 Underbidding by young firms

One could imagine that even in a world where contracts are effectively complete, young firm might systematically underbid (quote a price based on intentionally low estimates) to win the project. Of course the client knows this, and in the optimal contract corrects for it by holding them responsible for the extra overrun resulting from the underbidding.\(^{52}\) Young firms therefore end up paying for a higher share of the overrun. However it should be easy to see that the same

\(^{51}\)This is obtained by adding the difference in the share of overrun due to ambiguity and the share of overrun due to internal difficulties.

\(^{52}\)Again we are referring here to a fully efficient implicit contract.
objections that we list above to the competence-based explanation also apply in this case.

5.2.2 Varying level of moral hazard

Of course, this assumes that the mistakes (or prediction errors) are not made deliberately. Once we allow for such moral hazard, the client may well not be willing to insure the firm. Old firms could be less prone to moral hazard than young firms, and therefore bear less risk than young firms. The evidence we have presented on mean and variance of the overrun cannot be used to answer to this possible criticism, since they are endogenous: young firms could be generating the same level of overruns as old firms do precisely because they face higher punishments. Note however that this would be a real alternative explanation to the facts only if the levels of moral hazard were common knowledge (in the opposite case, it is just a restatement of what we assume in our reputation model). In this case, it is not clear why it should make any difference that firms are working with a new or a repeated client. Their level of moral hazard would not have the time to change between the first and the second contract realized for a client. The fact that firms are treated differently the second time around must therefore indicate that there is learning going on about the characteristics of the firm.

5.2.3 Varying levels of honesty

Young firms could also differ from old firms in their propensity to report costs honestly. They could be more prone to try report inflated costs, or to pretend that changes due to their own incompetence are due to the client changing his mind. If the client could not tell the cheaters apart, then the analysis of such a model would be similar to the analysis of the model we propose, and lead to the same conclusion (the reputation of old firms would be a reputation for honesty instead of a reputation for reliability). As we mentioned earlier, our modeling choice was to model a reputation for reliability, but it is clear that the analysis could be carried out with a different reason for the importance of reputation.

Note however that if clients could tell apart cheaters and honest firms, and punish cheaters by imposing them to pay more of the overrun, then we would also observe that young firms would pay on average more of the overrun (but this would not result in any social cost, unlike in our model or a version of the model with a reputation for honesty). Assuming that firms report in the questionnaire what they have reported to the client, then the evidence that young firms
pay more often than old firms overruns reportedly caused by the client would simply reflect the fact that they are lying more often then old firms. Because this argument rests on the fact that firms are lying in what they report to us as well as in what they say to their client, it is not easily verified or invalidated in the data. Note however that this argument implies that the clients never make any mistake in telling apart cheaters and honest firms. It is therefore self-defeating: why would firms cheat in the first place if they know that they are going to be found out?

Moreover three facts are difficult to reconcile with this explanation:

First firm pay on average 50% of the overrun when they report that the client is fully responsible for it. The suggested explanation would therefore implies an implausibly high fraction of cheaters among Indian software companies (young and old).

Moreover, if the client has perfect information and can enforce any sharing ex-post, there should be no variation in the contractual form, or at least it should not be related to the final outcome. However, firms pay more of the overrun when they have fixed costs contracts then when they have time and material contracts. Furthermore, young firms have more often fixed cost contracts. Therefore the ex-ante contracts seems both to be relevant and to be used by the clients, which is not consistent with the world we just described.

Finally, note that such a model would not explain the difference between contracts with repeated and new clients, or the difference between internal and external contracts: if the client has perfect information, then it is not easy to explain why firms would behave differently when dealing with different types of clients.

6 Conclusion

We set out in this paper to look for evidence that reputation plays an important role in determining contractual outcomes. We find that the evidence seems to strongly support this view, though given that the evidence is indirect (we do not actually observe people looking at reputation when deciding on contracts) and there are important firm characteristics that are potentially correlated with our measures of reputation, some doubts clearly remain.

The conclusion that reputation matters is of course important in itself: it gives support to a range of theories that are based on limitations of contracting. Moreover it might suggest an explanation of why the Indian software industry is not much larger (Indian Software export
were only worth 3.4% of the 1995 worldwide outsourcing business) given its obvious labor-cost advantage and the fact that this is a very labor-intensive industry. Or, to state the same point differently, why is it an equilibrium for software professionals in India to get paid so much less than their US counterparts? Reputation at the firm level is one possible explanation: most Indian firms are simply not trusted enough to be given important contracts. While our evidence cannot directly substantiate this view, the fact that reputation is important within the Indian industry suggests that it also ought to be important when an American client is deciding whether to go to a firm in India or to one in the US.

To add support to this view, our results also suggest that the process of reputation formation is rather inefficient. This is reflected in the fact that after controlling for age, whether or not a firm is dealing with a repeat buyer still makes a substantial difference to the contract. In other words, repeat buyers clearly know much more about the firm than the market does. In other words, the fact that a firm performed well in the past vis a vis one firm takes time to become public information. This is of course consistent with rational behavior on the part of the client but it clearly hurts the firm.

The policy implication of this view is that a credible system for rating firms modeled on credit rating systems may play an important role in the evolution of industries like the software industry where contracting is inherently problematic, by making it possible for the market to efficiently aggregate all that is known about each firm.

References


The U.S. imports a very large number of Indian software professionals for short-term assignments at a cost of more than twice what they would earn in India.


