Understanding Milestones in Software Development Contracts: An Empirical Study Through Principal-Agent Propositions

Extended Abstract

Yossi Lichtenstein and Gad Ariav
Tel-Aviv University

1. INTRODUCTION

The implementation schedule of software development is described as "the bedrock of successful software development" by Roditti and Bigelow, (1995; p. 7-21). However, in spite their centrality in software development, milestones have not been studied thoroughly. In this paper we use economic perspective in exploring the conditions influencing the number of milestones, the relative size of payments and damages tied to them and the extent of detail in their specification. We collectively label these characteristics as *milestone intensity*. Milestone intensity practically captures the level of control in software development processes.

Our guiding perspective is the Principal-Agent (P-A) literature. We conjecture that its explicit modeling of unobservable actions and uncertainty of outcomes is appropriate for studying software development in general, and milestone intensity in particular. We generate three hypotheses using a linear P-A model, and use three additional constructs of several P-A models. We then turn to operationalize the constructs and finally apply them in studying ten fixed price software development contracts.

2. PROPOSITIONS

We use a simple P-A **linear** model that emphasizes risk **(**Holmstrom, 1989). The main assumptions are that the principal is risk neutral, the agent's risk aversion is constant and his cost of effort is exemplified by a quadratic function ($k*Effort^2$). The proposition is that the agent's pay tends to be more performance-based as the risk, risk aversion and cost of effort are lower.

In our interpretation, when payments are given for partial outcomes (i.e., milestones) the compensation is considered less sensitive to the final performance. It maybe useful to think about milestone-tied payments as rewards for "work" and not for "performance": The agent is paid for doing the design, for programming a module and so on. Acceptance test tied payments are for performance -- if the system is not accepted, the agent would not be paid for the work done in design, programming etc. It is thus hypothesized that milestone intensity is stronger when risk is higher, agent risk aversion is higher, and cost of action is higher.

In addition, we measure three more constructs: effectiveness measurability from the linear multi-task model (Holmstrom and Milgrom, 1989); P-A relationship length and agent's reputation (Hart and Holmstrom, 1987; pp. 100-102).

The propositions and constructs are summarized in Table 1: independent variables which are hypothesized to be positively related to milestone intensity are denoted by a plus sign; a question mark says that no prediction is given for the three additional constructs.

Independent Variables	Milestone Intensity
Project risk	+
Agent's risk aversion	+
Agent's cost of effort	+
P-A relationship length	?
Agent's reputation	?
Effectiveness measurability	?

Table 1: Propositions

3. METHODOLOGY

3.1. Method

The methodology means to allow direct and detailed examination of contracts as well as their organizational contexts. We have studied two large Israeli firms as case-studies; at each, *all* suitable software development *contracts* were analyzed and contract managers were interviewed; in total, we studied 10 contracts.

Large organizations are preferred because each allow to study multiplicity of contracts, they have relatively long experience with outsourcing software development and developed relatively mature outsourcing procedures. Finally, the Principal-Agent assumption of risk neutral principals seems reasonable for large organizations.

A project is considered suitable if it satisfies the following conditions: it is developed by an external supplier, has a substantial component of custom made software and has not yet been deployed in the organization (to prevent perspective bias as noted by Barki et al., (1993)).

The number of contracts we studied, is still relatively small. However, access to actual contracts is still very limited: Ang and Beath (1993) studied 6 contracts, Lacity and Willcocks (1995) studied 61 sourcing *decisions*; other examples of related research were either organizational case studies or surveys.

3.2. Instruments

We have relied on existing instruments as much as possible; however, many had to be adapted to suit this research and some new instruments were needed. Although objective operationalizations are emphasized, quite a few subjective instruments are needed as well. The theoretical constructs are listed in Table 1 **missing**. These are given in a uniform format in order to facilitate the interview with the contract manager. The format selected is a seven point Likert scale checking the respondent attitude towards the statement. The scale labels range from complete-agreement (7), to complete-disagreement (1).

Milestone intensity

Four instruments capture the level of control concerning partial software development products and services:

- Number of milestones.
- Length (in pages) of all development milestone specifications, divided by contract length.
- Accumulated milestone-tied payments divided by custom software price.
- Milestone-tied damages (in thousand dollars per week).

Project risk

The risk questionnaire of Barki, Rivard and Talbot (1993) is the basis for our instrument -- alas shorter and adapted to outsourcing: Variables with relatively low loading (Barki et al., 1993; table 8, p. 214) and variables that are difficult to answer with confidence by the principal in the context of outsourcing are either removed or abbreviated. The risk is computed by multiplying uncertainty scores by the degree of potential loss.

According to Schmidt et al. (1996) three important items are not covered by Barki et al. (1993): misunderstanding the requirements, lack of frozen requirements and changing scope and objectives. We thus added the uncertainty variables used by Ang and Beath (1993): the interviewee is asked to what extent the principal was able to specify ex-ante the system requirements, development time and cost. In addition, a single criterion variable asking for the interviewee's assessment of project *uncertainty* is also included (following Barki et al., 1993; p. 213).

The risk instrument is divided into five parts: loss, objective uncertainty (e.g., number of users), subjective uncertainty (e.g., evaluation of the agent's expertise), requirements stability, and interviewee assessment of uncertainty.

Agent's risk aversion

Agent's size is typically used as a proxy for risk aversion. We use the most accessible measure of vendor size, namely, its number of employees. It should be noted that as we do not measure production cost, the agent's size may represent economies of scale in addition to risk aversion.

Agent's cost of effort

The construct represents the cost for the agent to apply effort that results in outcome. In our context, it amounts to the cost to achieve development timeliness for the specified functionality. We measure the cost of effort rather crudely through the principal's impression. In addition, we ask for the number of similar applications already developed by the same agent, following the assumption that developing many systems reduces the cost of effort.

Length of P-A relationship

Represents the *anticipated* length of the principal's relationship with the agent in developing software. We use subjective direct assessments of the contract managers.

Agent's reputation

We ask for the contract manager's assessment of the developer reputation for each of the four software development outcomes.

Effectiveness measurability

Represents the extent of the possible error in measuring the long term effectiveness of the system. The contract manager is asked to agree or disagree with the statement that it is easy to measure the long-term effectiveness of the system.

4. RESULTS

We have studied ten contracts in two large Israeli service providers. They are named here **A** and **B**; the contracts are denoted by **A1** to **B5** and given in the appendices (A and B).

Firm **A** is a service organization, has about 3 million customers and more than 35,000 employees. The IS department is in a "turnaround" situation -- following years of under-budgeting, IT is currently considered essential for productivity and service in a market turning more competitive. Most IS development is outsourced and most of the projects are fixed price contracts which were subject to competitive bidding. Currently there are 5 on-going development projects.

Firm **B** is a financial services organization with more then one million customers and about 9,000 employees. The information systems department employs about 800 professionals. About a third of the current 35 projects is outsourced, and 5 contracts meet the requirements of our study. Information systems are managed by a strong central department and have been considered a strategic asset for many years.

Table 2 lists the directly observable characteristics of the projects:

	A1	**A2**	**A3**	**A4**	**A5**	**B1**	**B2**	**B3**	**B4**	**B5**
Price ($K)	90	100	110	555	580	16	92	95	131	290
Development time (months)	4	6	14	9	21	4	6	5	4	6
Contract length (pages)	4	25	100	10	220	22	38	15	55	200

Table 2: Contracts

The two sites differ: **A**'s average project costs $287K, is developed for 11 months with a contract of 72 pages; **B**'s average project is much smaller -- it costs $125K, is developed for 5 months and its contract is 66 pages long.

Table 3 presents milestone intensity data. Two of the measures manifest considerable variance -- number of milestones and milestone tied payments. The two other measures -- damages and milestone-length -- manifest small variability: 8 out of 10 contracts do not specify damages for not achieving timeliness goals; most milestones are specified briefly, typically on a single page. The two sites do not differ much with respect to milestone number (2 and 2.2 in average), and milestone tied payments (57% and 70% in average). It should also be noted that the payments in **A5** are more then 100% because some of the cost of the packaged software is tied to the milestones, in addition to the cost of custom software.

	A1	**A2**	**A3**	**A4**	**A5**	**B1**	**B2**	**B3**	**B4**	**B5**
Number of milestones	0	0	1	4	6	6	1	1	1	4
Milestone tied payments (%)	0	0	40	63	180	100	100	33	53	66
Damages ($K/week)	0	0	1	0	5	0	0	0	0	0
Milestone length (pages)	0	0	0	1	50	1	1	1	1	10

Table 3: Milestone Intensity

The tables given in appendix C present the number of milestones against the independent variables and those in appendix D present milestone tied payments versus the independent

variables. Risk and cost of effort are given in several forms because we found significant differences among their measures in various items. As mentioned above, project risk is computed by multiplying uncertainty scores by the degree of potential loss. The uncertainty is expressed through four instruments, namely: objective uncertainty measures (e.g., number of users), subjective uncertainty measures (e.g., evaluation of the agent's expertise), requirements stability, and the interviewee's uncertainty assessment. We have found that the objective and subjective measures are correlated (r^2 is 0.66) and that the requirements stability and the assessment are weakly correlated (r^2 is 0.22); the other combinations are not correlated. Similarly, the two measures for agent's cost of effort -- subjective assessments and previous number of similar systems have been found to be negatively correlated.

The tables of appendix C *do not* support the hypothesized relations between the number of milestones and the independent variables. In order to ascertain this qualitative conclusion, we used linear regression between number of milestones and each of the independent variables and have not found significant correlation.

As can be seen in the tables, the **B1** contract has 6 milestones, but its various risk measures are low. As we interviewed the project coordinator we found that company **B** as well as the competing vendors, estimated that the development would cost about $500K. However, the winning vendor had developed a similar system for a previous customer and was ready to customize it for $16K. The contract has not been changed accordingly and the 6 milestones set in the original RFP remained in effect. Removing **B1** does not change the basic conclusion -- the number of milestones does not support the hypothesized relations. After removing **B1** it seems, though, that the number of milestones is related to the number of previous similar systems developed by the same agent, but in the *opposite* direction to our prediction.

Most of the tables of appendix D *do not* support the hypothesized relations between milestone-tied payments and the independent variables. One exception is again the measure for the agent's cost of effort -- this time the subjective assessment given by the principal. Again, in order to study further this qualitative conclusion, we applied linear regression between milestone-tied payments and each of the independent variables and have found that only the subjective cost of effort manifests a significant correlation (r^2 is 0.64).

5. DISCUSSION

The research addressed the conditions influencing milestone intensity in software development contracts. We have not found support to the theoretical predictions that milestone intensity is dependent on either project risk, agent's cost of effort, risk aversion, or effectiveness measurability. The conclusion concerning risk seems relatively robust -- we checked it with respect to multiple, objective and subjective measures. The other conclusions are weaker; in particular, two measures for the cost of effort gave contradictory results. We also have not found a significant correlation between milestone intensity and the length of P-A relationship and agent's reputation.

There could be a number of explanations behind these observations. Specifically, we could attribute these somewhat surprising observations to the crude assumption about the ability of the principals to behave "economically correct" in the first place, the inspection of a single influencing condition instead of the combined influence, an erroneous interpretation of the economic models, the use of faulty instruments and measures, and finally that there are alternative explanations, outside the realm of this particular inspection. We will address each of the above in turn.

The assumption about the principals ability to respond to risk is fundamental in this study. We did not have a reason to doubt it a priori since all interviewees are experienced project managers, as noted by Ropponen and Lyytinen (1997; Table 2). In retrospect, though, it raises the question whether risk management capabilities can be acquired without explicit attention - none of our sites applied a well defined methodology in that domain. Is risk management intuitive? We find this question intriguing and a rich ground for further research.

The context-dependent interpretation of the economic models is the essence of this study and one of its contributions. Inappropriate interpretation of the economic models is an inherent risk of any study that attempts to contrast abstract models with reality. The main remedies are the care in, and the explicitness of, the deliberations through which we transformed aspects of the formal models into empirically testable constructs. Furthermore, we have relied on fundamental models and derived essential,
broad and basic propositions.

We have basically 7 instruments (Table 1) with varying levels of robustness. Nevertheless, Milestone intensity - our dependent variable, and Project risk -- the major independent variable are comprehensive and generally internally consistent. In light of that critical stance we can argue that the main observation of this study, namely the lack of a clear milestone response to risk stands. The remaining conclusions are admittedly weaker and call for further investigation and a finer definition of related instruments.

An alternative view is demonstrated in Table 4 -- it presents the number of milestones with respect to custom software *price*. When project **B1** is removed, it seems that the number of milestones is positively related to price. A linear regression gives us an estimate that a milestone represents roughly about $90K of software development price; the correlation (r^2) is 0.66. According to this alternative explanation, number of milestones is set according to price; risk, risk aversion, cost of effort and so on are not directly addressed. However, price is correlated with risk (r^2 is 0.40); so, we conjecture that price is used as a simple surrogate for risk assessment.

	B1	**A1**	**B2**	**B3**	**A2**	**A3**	**B4**	**B5**	**A4**	**A5**
Price ($K)	16	90	92	95	100	110	131	290	555	580
Number of MS	6	0	1	1	0	1	4	4	4	6

Table 4: Price

Another alternative explanation is that additional mechanisms are used when risk or cost of effort are high. User participation in a software development team is considered a useful management technique to handle low structure projects (Applegate et al.; 1996; pp. 272 - 279). Indeed, only in the two projects with the highest risk (**A2** and **B4**) users take part in the development teams. Another example is prototyping: the vendor of project **B4** supplied a pilot system as part of the bidding; this may explain why a single milestone is considered satisfactory for this high risk project.

Yet another alternative explanation is suggested by the incomplete contracting literature. For example, Hart (1996; p. 3) describes milestones in house construction as a necessary protection needed because of the impossibility to draft complete contracts. In the context of IS development, Roditti and Bigelow (1995) explain that milestones protect the vendor from client requests that are made after a particular phase of work has been completed. So, milestones can be seen as a response to one component of risk, i.e. completeness of

requirements. However, it seems that this explanation is less relevant for fixed price contracts and indeed the ten projects in our study are considered stable.

The main limitation of this work is the limited number of contracts we have studied. More participating companies with substantial number of software development contracts are called for. This study is part of a larger and broader research (Lichtenstein, 1998) which we expect to shed more light on the issues raised here. The intuitions that govern software development efforts indeed deserve further understanding.

6. BIBLIOGRAPHY

Ang S. and Beath C. M., "Hierarchical Elements in Software Contracts", Journal of Organizational Computing (3:3), 1993, pp. 329-361.

Applegate L. M., McFarlan F. W. and McKenney J. L., Corporate Information Systems Management, The Issues Facing Senior Executives, Fourth Edition, Irwin, 1996.

Bakos J. Y. and Kemerer C. F., "Recent Applications of Economic Theory in Information Technology Research", Decision Support Systems (8:5), 1992, pp. 365-386.

Barki H., Rivard S., Talbot J., "Toward an Assessment of Software Development Risk", Journal of Management Information Systems (10:2), 1993, pp. 203-225.

Hart O., Firms, Contracts, and Financial Structure, Clarendon Press, 1996.

Hart O. and Holmstrom B., "The theory of contracts", Advances in Economic Theory, Fifth World Congress, editor: T. Bewley, Cambridge University Press, 1987, pp. 71-155.

Holmstrom, B., "Agency Costs and Innovation", Journal of Economic Behaviour and Organization (12:3), 1989, pp. 305-327.

Holmstrom B., and Milgron P., "Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design", Journal of Law, Economics and Organization (7:1), 1991, pp. 24-52.

Lacity M. C. and Willcocks L. P., "Interpreting Information Technology Sourcing Decisions From a Transaction Cost Perspective: Findings and Critque", Accounting, Management and Information Technology (5:3/4), 1995, pp. 203-244.

Roditti E. C. and Bigelow R. W., Computer Contracts: Negotiating, Drafting, Mathew Bender, 1995.

Ropponen J. and Lyytinen K., "Can Software Risk Management Improve System Development: An Exploratory Study", European Journal of Information Systems (6), 1997, pp. 41-50.

Schmidt R., Lyytinen K., Keil M. and Cule P., "Identifying Software Project Risks: An International Delphi Study", Proceedings of The Seventeenth Annual International Conference on Information Systems, Cleveland, 1996.

Appendix A: Firm A

A1 is a customization project of a laboratory management system. The system is adapted to one of its units, and is meant to replace an old system. All $90K are to be paid *after* acceptance and no milestones are specified. Potential loss is considered to be small, because the existing system is operational. The vendor is small with a large part of its business with **A**.

A2 is a decentralized budgeting system. The contract includes the licensing a software package ($120K) and its customization ($100K). Additional customization is done internally. Potential loss is considered high as the existing centralized budgeting system fails to answer the needs of the regional offices. The package is supplied by a large international firm and the customization is done by a small Israeli vendor.

A3 is a communication system with 1,500 independent service providers who are contracted by the organization. Loss will be large if the system is deployed but does not function correctly. If it is not deployed loss is not severe, because the current system is functional. The vendor is small with a large part of its business with **A**.

A4 is a payroll system, replacing an outdated system. The contract coordinator estimates that the vendor losses on the development and expects to make a profit on maintenance.

A5 is a logistic management system which is going to replace outdated and loosely linked systems. Uncertainty is mostly due to **A**'s organizational difficulties. The project coordinator estimates that a successful deployment of the system will save about $1M a year in logistics expenses. The suppliers are two large vendors.

Appendix B: Firm B

B1 automates a small frequent transaction. The contract is for the development and pilot deployment of the system. The agent is a small vendor that developed a similar system for another customer and is ready to customize it for $16K. However, it hopes to eventually do the company-wide deployment which includes large scale document conversion with estimated cost of $1.2M. The project coordinator, as well as the competing vendors, estimated that the development would cost about $500K. With this in mind, the RFP which became part of the contract set 6 milestones.

B2 is a customization of a Workflow package and its pilot deployment. Risk is considered low; however, if the pilot is successful, company-wide deployment of the system will have large implications on operations and customers.

B3 is a management support tool for internal IS development and it should replace an outdated system. Uncertainty is mainly due to the system's usage of new OS and communication infrastructure. The vendor is a new small software house; its founder has done several successful projects for **B**.

B4 is a payroll simulator that supplies both decision-support and operational functionalities. The two competing bidders have supplied pilot systems costing $40K each. Uncertainty is due to the system's usage of new OS and DB infrastructure. Potential loss is considered high because the system's decision-support is critical in wage negotiations.

B5 is an informational web-site. It is considered preliminary as an infrastructure for future commerce and the risk is considered low.

Appendix C: Number of Milestones

	A1	B5	B1	A3	B3	B2	B4	A4	A5	A2
Risk	.03	.08	.08	.11	.13	.14	.22	.23	.27	.28
Number of MS	0	4	6	1	1	4	1	4	6	0

Table C1: Risk

	A1	B5	A4	B3	B1	A5	A3	B4	B2	A2
Risk assessment	0	.04	.14	.28	.34	.37	.41	.44	.51	.76
Number of MS	0	4	4	1	6	6	1	1	4	0

Table C2: Risk assessment

	A5	B4	A4	B5	A1	A3	B2	A2	B1	B3
Risk aversion	500	500	300	150	30	30	20	10	10	10
Number of MS	6	1	4	4	0	1	4	0	6	1

Table C3: Risk aversion -- number of employees

	B3	B4	A1	A2	B1	B2	A3	B5	A4	A5
Cost of effort	0	0	1	1	2	2	5	5	10	10
Number of MS	1	1	0	0	6	4	1	4	4	6

Table C4: Cost of effort -- previous similar systems

	A1	A3	A2	B2	B3	B5	B4	A4	B1	A5
Cost of effort	0	.29	.33	.38	.42	.42	.46	.50	.50	.71
Number of MS	0	1	0	4	1	4	1	4	6	6

Table C5: Cost of effort -- assessment

	B5	A3	A5	B2	B3	A1	A2	A4	B1	B4
P-A length	.33	.50	.50	.83	.83	1	1	1	1	1
Number of MS	4	1	6	4	1	0	0	4	6	4

Table C6: Length of P-A relationships

	B4	A2	A5	B3	B2	B5	B1	A3	A1	A4
Reputation	0	.50	.50	.54	.67	.67	.71	.92	1	1
Number of MS	4	0	6	1	4	4	6	1	0	4

Table C7: Agent's reputation

	A1	A4	B1	B2	B3	B4	A2	A3	A5	B5
Effectiveness measurability	.17	.50	.50	.83	.83	.83	1	1	1	1
Number of MS	0	4	6	4	1	4	0	1	6	4

Table C8: Effectiveness measurability

Appendddix D: Milestone Tied Payments

	A1	B5	B1	A3	B3	B2	B4	A4	A5	A2
Risk	.03	.08	.08	.11	.13	.14	.22	.23	.27	.28
MS pay (%)	0	66	100	40	33	100	53	63	180	0

Table D1: Risk

	A1	B5	A4	B3	B1	A5	A3	B4	B2	A2
Risk assessment	0	.04	.14	.28	.34	.37	.41	.44	.51	.76
MS pay (%)	0	66	63	33	100	180	40	53	100	0

Table D2: Risk assessment

	A5	B4	A4	B5	A1	A3	B2	A2	B1	B3
Risk aversion	500	500	300	150	30	30	20	10	10	10
MS pay (%)	180	53	63	66	0	40	100	0	100	33

Table D3: Risk aversion -- number of employees

	B3	B4	A1	A2	B1	B2	A3	B5	A4	A5
Cost of effort	0	0	1	1	2	2	5	5	10	10
MS pay (%)	33	53	0	0	100	100	40	66	63	180

Table D4: Cost of effort -- previous similar systems

	A1	A3	A2	B2	B3	B5	B4	A4	B1	A5
Cost of effort	0	.29	.33	.38	.42	.42	.46	.50	.50	.71
MS pay (%)	0	40	0	100	33	66	53	63	100	180

Table D5: Cost of effort -- assessment

	B5	A3	A5	B2	B3	A1	A2	A4	B1	B4
P-A length	.33	.50	.50	.83	.83	1	1	1	1	1
MS pay (%)	66	40	180	100	33	0	0	63	100	53

Table D6: Length of P-A relationships

	B4	A2	A5	B3	B2	B5	B1	A3	A1	A4
Reputation	0	.50	.50	.54	.67	.67	.71	.92	1	1
MS pay (%)	53	0	180	33	100	66	100	40	0	63

Table D7: Agent's reputation

	A1	A4	B1	B2	B3	B4	A2	A3	A5	B5
Effectiveness measurability	.17	.50	.50	.83	.83	.83	1	1	1	1
MS pay (%)	0	63	100	100	33	53	0	40	180	66

Table D8: Effectiveness measurability