MIS Quarterly

BIG DATA & ANALYTICS IN NETWORKED BUSINESS

A TREE-BASED APPROACH FOR ADDRESSING SELF-SELECTION IN IMPACT STUDIES WITH BIG DATA

Inbal Yahav
Department of Information Systems, Graduate School of Business Administration, Bar Ilan University,
Ramat Gan 52900, ISRAEL {inbal.yahav@biu.ac.il}

Galit Shmueli
Institute of Service Science, College of Technology Management, National Tsing Hua University,
Hsinchu 30013, TAIWAN {galit.shmueli@iss.nthu.edu.tw}

Deepa Mani
Indian School of Business, Gachibowli,
Hyderabad 500032, Andhra Pradesh, INDIA {deepa_mani@isb.edu}

Appendix A

Study 3: Impact of Technology Outsourcing

Outsourcing is fundamental to firm competitiveness. Firms are increasingly externalizing a variety of core business functions such as research and development, product development, and marketing to achieve diverse strategic objectives (Mani et al. 2012, 2013). Yet, despite their extended reach and impact, outsourcing initiatives involve high failure rates with adverse impacts on various critical performance metrics.[1] Prior research in information systems attributes heterogeneity in outsourcing performance to efficacy of contract design and management (Mani et al. 2013; Susarla et al. 2011). The right contract aligns incentives between the client and the vendor to engender cooperative behavior that is necessary for effective execution of the outsourced task and adaptation to disturbances. Yet, what drives design of outsourcing contracts as well as how performance varies across these contracts are issues that have only recently attracted empirical attention.

We focus on how performance of outsourcing contracts vary with two attributes of the contract: contract price and contract length. The former reflects a categorical polytomous intervention while the latter reflects a continuous intervention. Outsourcing contracts are largely categorized as either fixed price or variable price (e.g., Gopal et al. 2003, Mani et al. 2012), and are often self-selected to minimize the economic tradeoff between *ex ante* provision of incentives and *ex post* renegotiation of contractual specifications (Bajari and Tadelis 2001). As a result, strategic outsourcing initiatives that involve a higher probability that adaptations are needed, are governed by less complete variable price contracts, whereas simpler, more stable outsourcing initiatives lead to more complete fixed-price contracts that seek to primarily provide high-powered incentives to the vendor to reduce costs of task ownership.

Similarly, contract duration also reflects a tradeoff between providing *ex ante* incentives for specific or non-contractible investments and *ex post* inefficiencies of vendor lock-in and inflexibility that, in turn, result in maladaptation and underinvestment (e.g., Susarla et al. 2012). Contingent contracts facilitate adaptation but are difficult and costly to design and administer. A central goal in contract design, therefore, is

[1]In a 2005 survey by Deloitte Consulting, 70% of the respondents expressed significant dissatisfaction with their outsourcing projects. According to SAP INFO Solutions, four out of five inked outsourcing contracts will need to be renegotiated within two years. Further, 20% of all such contracts will collapse (Johnson 2006).

to choose a contract duration that "maintain[s] incentives for efficient adaptation while minimizing the need for costly adjudication and enforcement" (Crocker and Masten 1988, p. 328). In the following sections, we analyze performance outcomes in outsourcing across multiple contract types and durations after controlling for the self-selection of these contractual parameters. In brief, we consider the following two selection models:

1. Self-selection of contract pricing (multiple interventions)
2. Self-selection of contract duration (continuous intervention)

Data and Measures

Our empirical analysis is based on over 1,400 outsourcing initiatives implemented between 1996 and 2008. Information on the outsourcing initiatives is obtained from International Data Corporation's (IDC) services contracts database. This data dates back to 1996, and is the primary input to this study. We use Lexis-Nexis to verify and supplement IDC information on contract announcement and signing dates. Company data from COMPUSTAT and SDC Platinum and stock price data from the Center for Research in Security Prices (CRSP) complements the contract data.

Our initial sample comprised public, private, and government contracts signed in nearly 30 countries. Our final sample comprises outsourcing contracts that satisfy two requirements. First, the firm must be publicly traded on a major U.S. stock exchange. Second, information on the contract used to govern the outsourcing initiative must be available. Our final sample of 1,411 contracts includes 374 vendors and 710 clients. The operationalization of key variables in our analyses, including contract price and contract length, is described in Table A1. In the following analyses, we compare the tree-based method to a naïve approach only, because PS is not directly applicable to polytomous and continuous intervention variables.

Analysis 1: Impact of Contractual Pricing Mechanisms

We begin with an analysis of performance differences across different contractual pricing mechanisms. Prior empirical research (e.g., Gao 2006; Mani et al. 2013) in assessing the magnitude of impact of outsourcing has largely compared mean industry- and risk-adjusted performance outcomes across fixed and variable price contracts. These studies suffer from two important limitations. First, there may be multiple contract types, and controlling for these multiple interventions is costly and difficult. Indeed, some of these studies coalesce multiple contract types (such as incentive, combination, and time and materials contracts) as variable price contracts. Second, the model of contract selection is theoretically specified *a priori*. However, given the plethora of theories used to specify contract selection—including transaction cost economics, resource based view, knowledge theory of the firm, and information processing view of the firm (for a review of this literature, see Dibbern et al. 2004)— it would be useful to use an exploratory view to identify pre-intervention characteristics that are most relevant to the model of contract choice.

We use the tree approach to address these limitations and compare performance across multiple contractual pricing mechanisms. Table A1 describes the X, T, and Y variables used in our study. We compare the six price methodologies in terms of market and financial performance for the three-year period following the implementation of the contract. In particular, we compare income efficiency gains as well as gains in short-term and long-term market value to the client following the implementation of the outsourcing contract. Prior research (Daniel and Titman 2006) suggests that financial markets may be inefficient in pricing complex events such as outsourcing, thereby necessitating a long-term perspective in the assessment of market value of such events. It is also important to assess whether the market value resulting from the outsourcing initiatives is associated with efficiency changes due to the initiative. Therefore, we complement measures of market value with the median income efficiency estimates for the three-year period following the implementation of the outsourcing contract.

Naïve Analysis: Figure A1 presents charts of performance comparisons. Each row corresponds to one performance measure. The left panels portray the baseline "naïve" comparison of mean performance across the different pricing mechanisms. Circles represent mean performance and error bars are a 90% confidence interval for the mean.

A naïve comparison of the mean long-term market value gains across different pricing mechanisms suggests the following. First, only transactional pricing contracts earn positive long-term abnormal returns following the implementation of the outsourcing contract[2] (second row, $p < 0.1$). Second, announcement period returns for all contracts are zero (top row), suggesting that the market, in general, does not impound

[2] The 90% confidence interval for mean long-term abnormal returns for transactional contracts is (0.0001, 0.0029).

Table A1. Outsourcing Contracts Data: Variable Description	
Contract.Price (Intervention)	Contract pricing mechanism, with six categories: • *Fixed Price*: Fixed payment per billing cycle • *Transactional Price*: Fixed payment per transaction per billing cycle • *Time and Materials*: Payment based on input time and materials used during the billing cycle • *Incentive*: Payment based on output improvements against key performance indicators or any combination of indicators • *Combination*: A combination of any of the above contract types, largely fixed price and time and materials • *Joint Venture*: A separately incorporated entity, jointly owned by the client and the vendor, used to govern the outsourcing relationship.
Contract.Length (Intervention)	Duration of the contract, in months
Task.Type (Engagement)	Task type is one of five categories: • Outsourcing of technology-enabled business processes and functions (includes Business Consulting and Business Process Outsourcing) • Custom technology outsourcing, including applications, networks and other systems • Information Systems Outsourcing • Outsourcing of Application, Network and Desktop Management • Training, Deploy and Support
Bid.Type (Engagement)	Prior cooperative association between the firms, indicative of trust: • *Competitive bidding*: Absence of prior association between the firms. • *Incumbent*: The vendor has an existing relationship with the client. • *Sole-sourced*: The selected vendor is the only provider of the outsourced function • *Non-competitive contracts*: Contracts that are not sole-sourced or outsourced to an incumbent vendor.
Services.Contract.Value (Engagement)	Ratio of contract value to operating expenses
Uncertainty (Firm attribute)	Uncertainty in business requirements. Variance in outsourcing firm's return on assets (RoA) in three years prior to contract year
Experience (Firm attribute)	Outsourcing experience. Cumulative number of strategic alliances across the client's life
Size (Firm attribute)	Market value of equity of the outsourcing firm. The number of shares outstanding times market price
Ann.Returns (Outcome)	Firm-specific daily abnormal returns ($\hat{\varepsilon}_{it}$ for firm *i* on day *t*), computed as $\hat{\varepsilon}_{it} = r_{it} - \hat{r}_{it}$, where r_{it} is the daily return (to the value weighted S&P) estimated from the market model: $r_{it} = \alpha_i + \beta_i r_{mt} + \varepsilon_{it}$. This model is used to predict daily returns for each firm over the announcement period [-5,+5].
Long.Term.Returns (Outcome)	Monthly abnormal returns are estimated from the Fama and French (1993) three-factor model as excess of that achieved by passive investments in systematic risk factors. Expected to be zero under the null hypothesis of market efficiency. Monthly abnormal returns are used to estimate the implied three-year abnormal return following the outsourcing contract.
Median.Income.Eff (Outcome)	Income efficiency is estimated as earnings before interest and taxes divided by number of employees. We use median income efficiency for the three-year period following the implementation of the outsourcing contract.

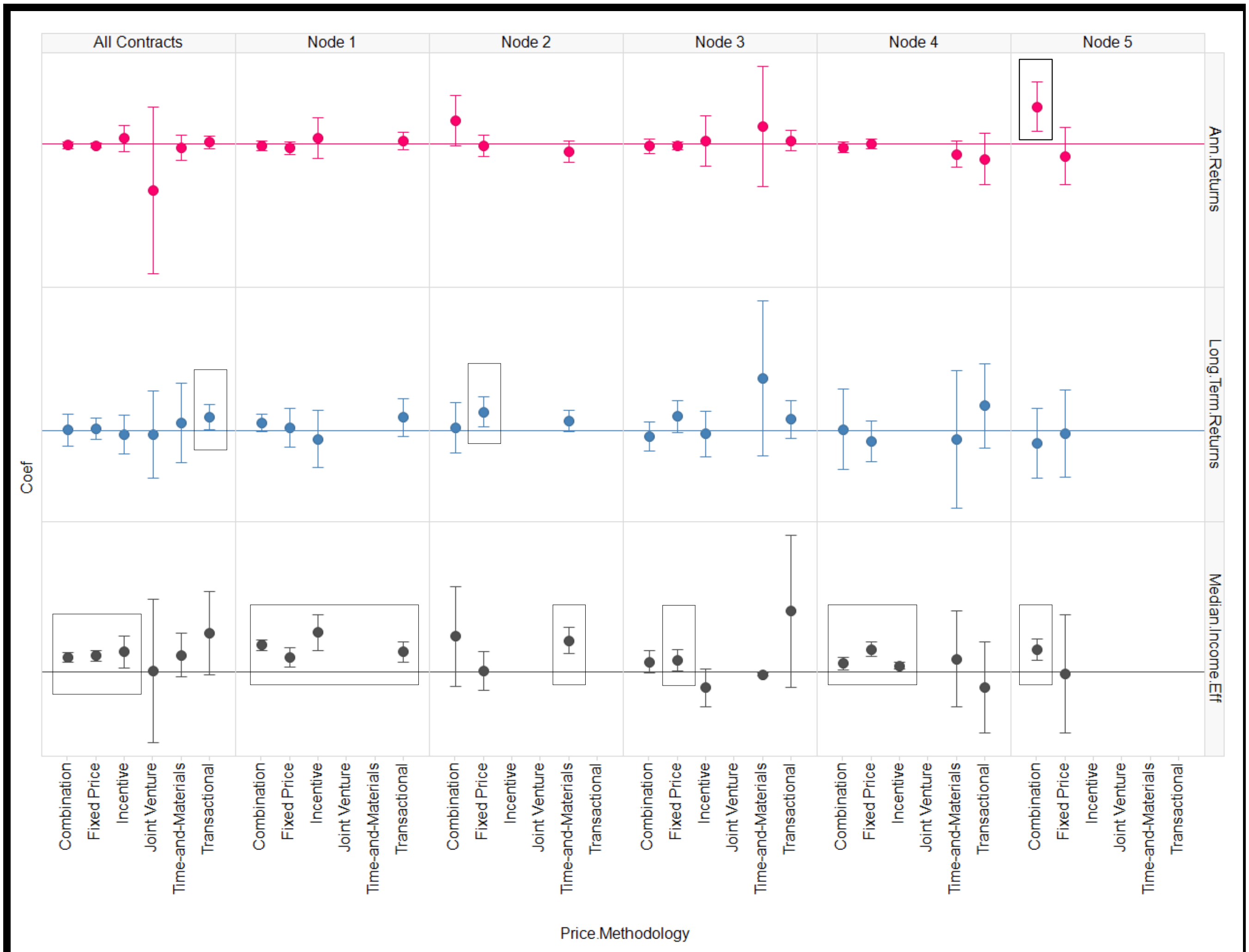

Figure A1. Comparing Performance of Six Price Methodologies. (Left column: naïve approach. Right columns: accounting for self-selection using terminal-node-level comparisons. Rectangles mark significant coefficients (p < .01).)

the value of outsourcing contracts into stock prices. Given the results for long-term abnormal returns, we conclude that the market, in particular, underestimates the value created by transactional pricing contracts. Third, income efficiency gains are positive for fixed-price and combination contracts, although these gains are not impounded in market value (third row). These findings are somewhat consistent with prior research in IS (e.g., Mani et al. 2013), which suggests that markets underestimate the value of simple outsourcing engagements, often governed by fixed or transactional pricing mechanisms, where the ownership of the outsourced process is transferred to the vendor. Yet, they provide an incomplete picture of outsourcing value. Important questions remain: Do all simple engagements, governed by fixed or transactional price contracts, create value? What types of complex outsourcing engagements create value for the client? How do participant firms mitigate risks inherent to these engagements?

Tree-Based Selection Model: Figure A2 portrays the tree of contract price methodology. We find that relational variables (such as prior association between the firms), indicative of mutual trust, and task variables (such as type of outsourcing initiative) or total contract value that are indicative of complexity of the outsourcing engagement, explain heterogeneity in contract price. Trust serves to reduce appropriation concerns inherent to incomplete contracts while higher levels of adaptation inherent to complex tasks increase the likelihood of choosing more incomplete contracts. Fixed-price contracts involve high renegotiation costs, while time and materials contracts require difficult and costly monitoring of input resources. Indeed, incentive contracts are observed either when there exists a prior association between the firms or when

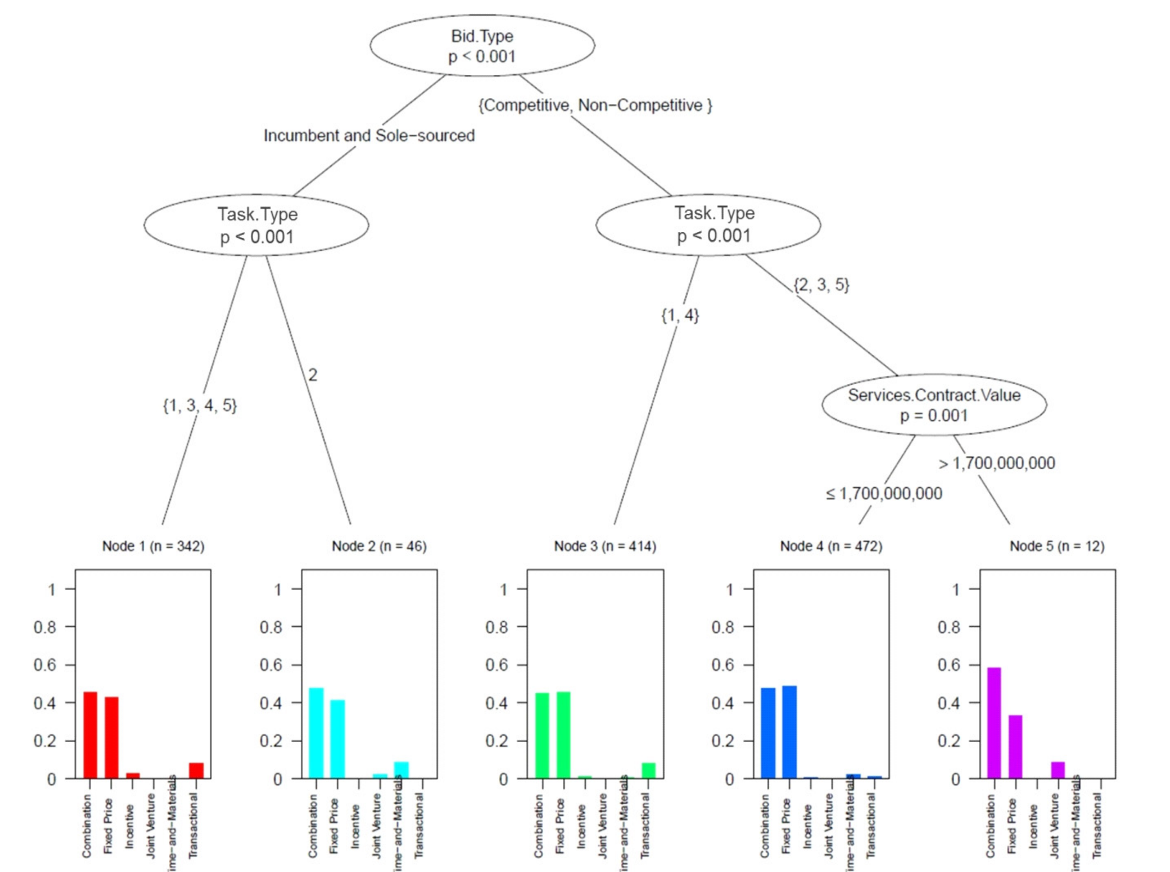

Figure A2. Classification Tree for Price Methodology as a Function of Pre-intervention Variables **(Bar charts in terminal nodes represent the proportion of contracts for each price methodology.)**

the task is complex and characterized by high levels of adaptation (nodes 1 and 3). Similarly, all joint ventures are observed either for development of complex enterprise systems with an existing vendor (node 2) or for high value contracts greater than \$170 million (node 5). Node 5 also has a significantly greater proportion of combination contracts than fixed price.

Tree-Based Performance Analysis: The right panels (columns 2–6) in Figure A1 portray the performance outcomes for subsets of the data, which correct for the self-selection into the underlying contract type. Each of the five right columns corresponds to a terminal node in the classification tree in Figure A2. An analysis that accounts for self-selection into the various contract types tells a slightly different, more nuanced story compared to the naïve analysis. Examining the terminal-node-level analyses, we find the following: First, when there are relatively higher levels of trust between the client and vendor, the ensuing reduction in relational uncertainty positively impacts all types of outsourcing contracts. This positive impact is reflected in the positive income efficiency gains across all contractual pricing mechanisms in node 1.

However, that the impact of mutual trust on efficiency gains is moderated by the complexity of the underlying task is reflected in the difference in income efficiency gains between nodes 1 and 2. While the tasks underlying a bulk of the contracts in node 1 are likely simple and relatively stable, node 2 comprises complex IT outsourcing engagements that are idiosyncratic to the business needs of the client and offer little scale. The differences in task complexity are reflected in differences in contract value between the two nodes; while the average contract value in node 1 is nearly \$58 million, the equivalent value for node 2 is nearly \$266 million. These custom engagements involve high levels of uncertainty and incompleteness in task specification. As a result, node 2, which also comprises contracts that are indicative of high levels of trust, displays positive income efficiency gains only for time and materials contracts. The result for node 2 suggests that in these cases, time and materials contracts that allow for risk sharing of cost overruns incentivize the vendor to create value.

Second, an interesting result in node 2 is that fixed price contracts characterized by high levels of trust between the client and vendor do not yield income efficiency gains but have a positive impact on long-term market value. A possible explanation for this outcome is that these contracts create value along performance dimensions other than income efficiency gains. Future research could explore further how clients leverage mutual trust in these contracts and the nature of performance gains that ensue.

Third, incentive contracts too are observed for custom IT outsourcing contracts and outsourcing of business processes and functions when the levels of trust between participant firms are relatively high. These contracts are characterized by positive income efficiency gains in the three-year period following the implementation of the contract (nodes 1 and 4). Further, in the case of node 1, we find positive returns for the announcement period (-10, +10). Therefore, the market recognizes the importance of trust in incomplete incentive contracts with some leakage about this valued information in the days preceding the implementation of the outsourcing contract. Interestingly, these contracts create value for the client irrespective of the nature of the task.

Finally, results for nodes 4 and 5 offer important contrasts in terms of total contract value. To the extent that total contract value is indicative of the complexity of outsourcing, combination contracts create value for complex engagements, as reflected in the positive income efficiency gains and announcement period returns following the implementation of the contract (node 5). In contrast, fixed price contracts are best suited for simpler engagements, as reflected in the positive income efficiency gains to these contracts in node 4.

In short, the tree-based method uncovers potential self-selection factors and, when conditioning on those, reveals more nuanced insights compared to the naïve approach.

Analysis 2: Impact of Contract Duration

The limited theoretical work on the selection of contract duration focuses on the tradeoffs between providing *ex ante* incentives for specific or non-contractible investments and *ex post* inefficiencies of vendor lock-in and inflexibility that result in maladaptation and underinvestment (e.g., Crocker and Masten 1988; Susarla et al. 2012). The limited theory in this space has in turn, resulted in little empirical research on models of selection of contract duration or the impact of the latter on performance. Further, controlling for self-selection of continuous interventions such as contract duration requires effective instruments that influence duration but not performance. Such instruments are difficult to find. Finally, as we noted in our comparative assessment of performance across different contract types, it would be difficult to *a priori* hypothesize and test for the performance impact of interactions between contract length and different firm, task, and relational variables.

Naïve Analysis: The top row of Figure A3 presents a naïve comparison for each of the performance measures as a function of contract duration. The insignificant slope in all cases suggests that contract duration does not have any impact on performance gains from outsourcing.

Tree-Based Selection Model: Figure A4 presents the regression tree for duration of the outsourcing contract as a function of pre-intervention variables. We find that task variables indicative of complexity of the engagement, and prior experience of the client in managing similar inter-firm alliances, explain heterogeneity in contract duration. The shortest durations (nodes 1, 6, and 8) have the lowest value contracts, while longer-term contracts are observed for the highest end of contracts in terms of total contract value (nodes 4 and 5). These results suggest that for complex engagements, where the likelihood of non-contractible investments may be higher, firms implement contracts of longer duration to provide *ex ante* incentives for the vendor to undertake these investments.

Tree-Based Performance Analysis: For each of the terminal nodes, we compare market and financial performance gains to the outsourcing client across different contractual lengths. The results are displayed in Figure A4.

Compared to the naïve analysis, the node-level charts suggest the following results: First, results for announcement period returns suggest that markets reward long-term contracts ($\beta > 0$, $p < 0.1$) in two cases: (1) high value contracts (Services.Contract. Value > 140 million) in outsourcing of technology-enabled business processes and functions, IS outsourcing, and application, network, and desktop management, where the clients have prior experience in managing similar alliances (node 5), and (2) low value, custom IT outsourcing contracts (node 8).

The positive effect of prior cumulative experience identified in result (a) may be an outcome of two factors. First, experience helps clients find potentially useful solutions to inefficiencies in long-term contracts such as price or performance lock-in. Second, the greater the experience of the client, the easier for the firm to interpret and respond to unforeseen contingencies that are common in long-term contracts. Result (b) suggests that in engagements that require specific investments, the benefits of long-term contracts are limited to those where the scope is

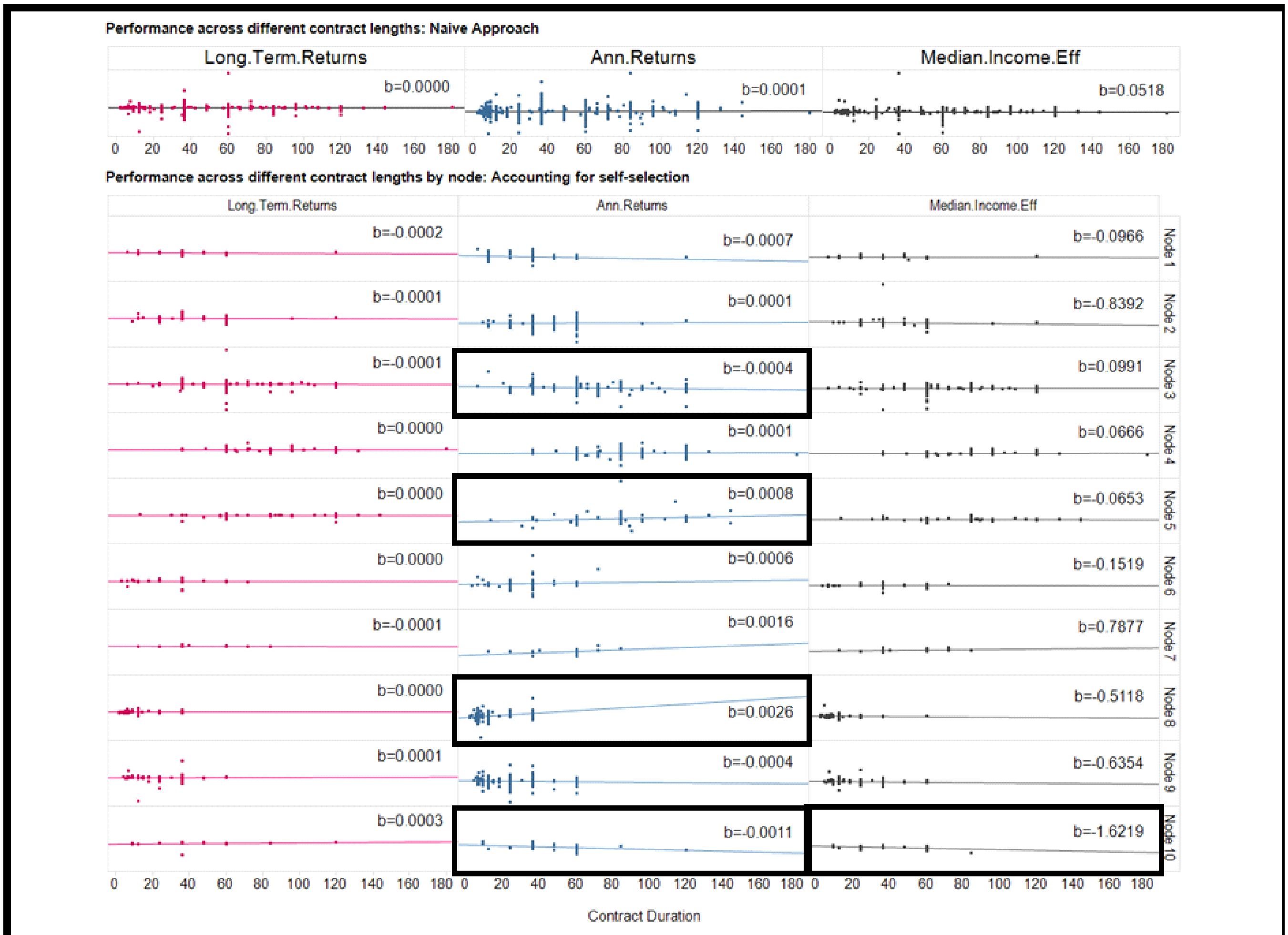

Figure A3. Comparing Effects of Contract Length on Performance **(Top row: naïve approach. Rows 2–11: accounting for self-selection using terminal-node-level models. Rectangles mark significant effects (*p* < .05).)**

minimal. In engagements with larger scope, the costs may outweigh benefits of *ex ante* incentives for specific investments. Insignificant long-term income efficiency gains in both cases suggest that the market efficiently anticipated these gains and impounded them in announcement stock prices.

Second, nodes 3 and 10 are characterized by negative announcement period returns ($p < 0.1$). The magnitude of value destruction is greater for node 10 as evidenced by the significant income efficiency losses ($p < 0.01$) over the three-year period following the implementation of the contract. In other words, the negative announcement period returns do not completely reflect the extent of value destroyed by these long-term contracts. The results reaffirm that, in the case of long-term IT outsourcing contracts that require specific or non-contractible investments, as the scope of the engagement increases, the costs of long-term contracts outweigh the benefits. For node 3, representative of IS outsourcing, application development, and network and desktop management contracts with contract values between 10 and 140 million, announcement period returns are negative and significant ($p < 0.1$). The result could be an outcome of the outsourced technology services being out of step with changing business objectives. For instance, the client may be locked into high prices that do not reflect the lower prices of inputs such as hardware and data storage or into performance benchmarks that do not reflect the changing value of underlying technologies.

Here, as in Analysis 1, the tree-based approach uncovers potential selection-bias variables that affect choice of contract duration. Conditioning on these variables leads to more insightful and accurate conclusions.

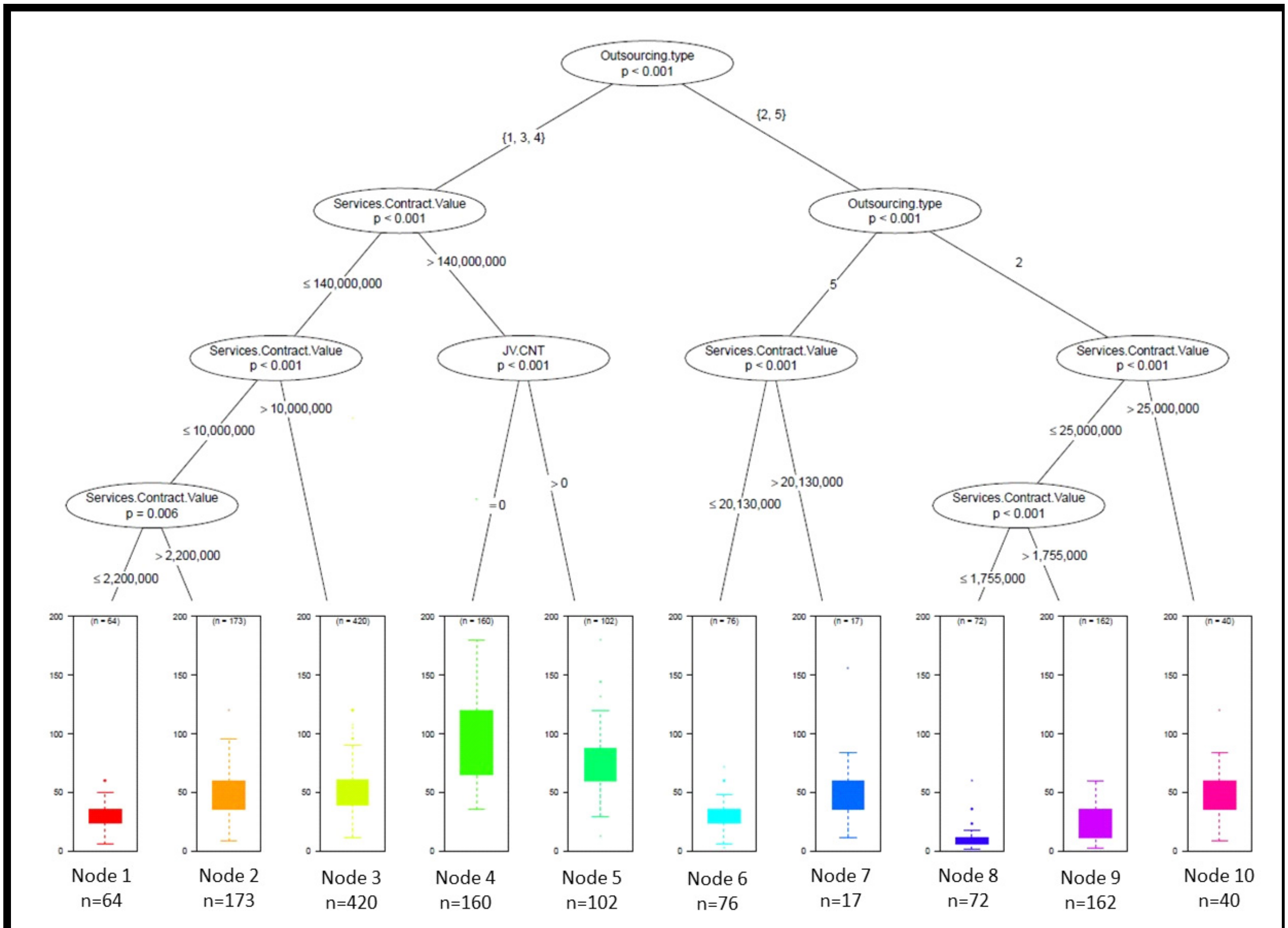

Figure A4. Regression Tree for Contract Duration as a Function of Pre-intervention Variables (Box plots in terminal nodes represent contract duration distribution in that node.)

References

Bajari, P., and Tadelis, S. 2001. "Incentives Versus Transaction Costs: A Theory of Procurement Contracts," *RAND Journal of Economics* (32:3), pp. 387-407.

Crocker, K. J., and Masten, S. E. 1988. "Mitigating Contractual Hazards: Unilateral Options and Contract Length," *The RAND Journal of Economics* (19:3), pp. 327-343.

Daniel, K., and Titman, S. 2006. "Market Reactions to Tangible and Intangible Information," *The Journal of Finance* (61:4), pp. 1605-1643.

Dibbern, J., Goles, T., Hirschheim, R., and Jayatilaka, B. 2004. "Information Systems Outsourcing: A Survey and Analysis of the Literature," *Communications of the ACM* (35:4), pp. 6-102.

Fama, E., and French, K. 1993. "Common Risk Factors in the Returns on Stocks and Bonds," *Journal of Financial Economics* (33), pp. 3-56.

Gao, N. 2006. "What Does Stock and Accounting Performance Tell Us about Outsourcing?," Working Paper, Finance Group, University of Pittsburgh (SSRN: http://ssrn.com/abstract=681321).

Gopal, A., Sivaramakrishnan, K., Krishnan, M. S., and Mukhopadhyay, T. 2003. "Con tracts in Offshore Software Development: An Empirical Analysis," *Management Science* (49:12), pp. 1671-1683.

Johnson, L. K. 2006. "Successful Business Process Outsourcing," *MIT Sloan Management Review* (47:2), pp. 5-6.

Mani, D., Barua, A., and Whinston, A. 2010. "Does Strategic Outsourcing Create Financial Value?," in *Operational Control in Asset Management: Processes and Costs*, M. Pinedo (ed.), Basingstoke, UK: Palgrave Macmillan, pp. 190-208.

Mani, D., Barua, A., and Whinston, A. B. 2012. "An Empirical Analysis of the Contractual and Information Structures of Business Process Outsourcing Relationships," *Information Systems Research* (23:3, Part 1), pp. 618-634.

Mani, D., Barua, A., and Whinston, A. B. 2013. "Outsourcing Contracts and Equity Prices," *Information Systems Research* (24:4), pp. 1028-1049.

Susarla, A., and Barua, A. 2011. "Contracting Efficiency and New Firm Survival in Markets Enabled by Information Technology," *Information Systems Research* (22:2), pp. 306-324.

Susarla, A., Oh, J.-H., and Tan, Y. 2012. "Social Networks and the Diffusion of User-Generated Content: Evidence from YouTube," *Information Systems Research* (23:1), pp. 23-41.