



**Impact Assessment in Observational Studies: A
Classification and Regression Tree Approach**

<http://eprints.exchange.isb.edu/85>

Working Paper

Indian School of Business

2013

Impact Assessment in Observational Studies: A Classification and Regression Tree Approach

Galit Shmueli and Deepa Mani

Indian School of Business

Hyderabad 500 032, India

Abstract

We introduce a tree-based approach for assessing the performance impact of diverse self-selected interventions in management research. Our approach, which takes advantage of “Big Data”, or observational data with large sample sizes and a large number of variables, offers important advantages over traditional propensity score matching. In particular, the tree-based approach to assessing the impact of interventions offers a data-driven methodology that applies to a wide range of intervention types (binary, polytomous, continuous), allows for examination of nascent interventions whose selection cannot be theoretically specified a priori, identifies pre-intervention variables that correlate with the self-selected intervention, and presents comparisons of ensuing performance in visuals that are easy to discern and understand. We illustrate the method and the insights that it yields in the context of two studies: analysis of the impact of an eGov service in India, and comparison of performance across different contractual pricing mechanisms and contract durations in the outsourcing of technology and technology-enabled business functions.

1 Introduction

Performance or impact assessment is an important component of empirical research in business and policy. Impact is frequently assessed at various levels - individual, organization or society - when analyzing the merits of individual behavior, organizational strategies, or government policies. An important consideration in impact assessment is that the observed interventions or strategies are largely self-selected by organizations and managers. These include diversification programs, adoption of enterprise technologies, structural and process changes, or marketing and innovation investments. In all these cases, the observations in the intervention group are not randomly chosen from the population from which the researcher would like to draw inferences. Several factors, including unobserved heterogeneity, impact organizational and individual selection of an intervention, and these factors may also impact outcomes to the intervention. The failure to correct for these observed and unobserved factors that influence selection of the intervention and ensuing performance results in biased and inconsistent estimates of performance gains from the intervention (Heckman et al., 1997).

Prior causal research in management, using observational data, largely addresses the self-selection bias using either the propensity scores matching (PSM) approach (Rosenbaum and Rubin, 1983) or the Heckman approach (Heckman, 1979). Both methods attempt to match the self-selected intervention group with a control group that has the same “propensity” to select the intervention. The difference in outcomes between the intervention group and the propensity-matched control group measures the unbiased magnitude of the impact of the intervention. Both approaches use a two-step method, where the first step consists of modeling the likelihood of self-selection (the “selection model”). The main difference between PSM and the Heckman approach is that the former assumes propensities to self-select the intervention are observable and uses the data to estimate them, while the latter treats them as unobservable¹. While there are arguments supporting the use of each method², PSM has become especially useful in the realm of ‘Big Data’, where high-dimensional data present a rich set of measurements for each record. Similarly, our proposed approach to addressing self-selection also leverages big data; therefore we contrast and benchmark our method against PSM.

¹A common approach to modeling the latent scores is by using an observed *instrumental variable* (IV) that is uncorrelated with the error but highly correlated with the causal variable. However, Guo and Fraser (2009, p. 100) note: “Heckman (1997)...concluded that when responses to treatment vary, the standard argument justifying the use of instrumental variables fails unless person-specific responses to treatment do not influence the decision to participate in the program being evaluated.”

²As Heckman et al. (1998) note, “Even if the propensity score is known, is it better, in terms of reducing the variance of the resulting matching estimator, to condition on X or P(X)? There is no unambiguous answer to this question.”

PSM is common in settings where it is impossible or even unethical to assign individuals to an intervention or a non-intervention control group (Dehejja and Wahba, 1999; Holland, 1986). Yet, researchers are interested in evaluating effect of an intervention from observational data (Stuart, 2010). Scholars have discussed the potential of this approach in epidemiological (Little and Rubin, 2000), sociological (Winship and Sobel, 2004), and econometric (Dehejja and Wahba, 2002; Heckman et al., 1997, 1998) literature, and this method has also found promising applications in management and information systems research to assess causal effects at the individual and firm levels (e.g., Rubin and Waterman, 2006; Mithas and Almirall, 2006; Mithas et al., 2009; Mithas and Krishnan, 2009; Mithas et al., 2005; Mithas and Lucas, 2010). In almost all these applications, researchers use a logistic or a probit model to compute propensity scores, with a small emergent research (Lee et al., 2010; Westreich et al., 2010) on the use of classification trees and their variants for computing propensity scores instead of logistic regression.

In management research, propensity scores are typically estimated using logistic or probit regression. Notwithstanding its widespread use in management research, the application of logistic or probit regression to estimate propensity scores is subject to several limitations especially pronounced in large datasets. First, the selection model underlying the logistic/probit specification is grounded in theoretical knowledge of the intervention and its selection. However, when interventions are determined by several firm-, industry- or task-level characteristics, or where there is little to no theoretical knowledge of the intervention effect, several pre-intervention characteristics that may be unmatched across the intervention and control groups, yet excluded from the selection model. Limited knowledge about what variables to include, their form, and the nature of their relationship with the intervention, render the use of logistic or probit regression challenging.

Second, some impact studies consider more than two interventions. In these cases, the simple logistic model is replaced with a multi-category polytomous model that divides the population into cohorts that are adjusted for differences in pre-intervention characteristics. Estimation of propensity scores in the case of more than two interventions is complicated and challenging³. As a result, the use of propensity scores to estimate performance effects of more than two interventions is limited in the management literature.

Third, the intervention of interest may be continuous rather than categorical. For example, this study evaluates the effect of contract duration on the financial performance of

³The multinomial logistic regression, also known as *discrete choice model* for polytomous or ordinal interventions, relies on more assumptions that are often violated in practice (such as independence of irrelevant alternatives), requires estimating a larger number of parameters, and involves model diagnostics that are more complicated. The alternative multinomial Probit model is challenging computationally. See Chapters 10-11 in Hilbe (2009).

outsourcing contracts, where contract duration is self-selected. In such cases, using logistic or probit regression to estimate propensity scores, requires discretizing the continuous intervention variable into bins (Heinrich et al., 2010), thereby creating multiple categories. Binning leads to information loss and especially since it is unclear how to perform the binning that best distinguishes between intervention levels while leading to maximum variance in outcomes.

Finally, in large samples, relying on statistical significance of the regression coefficients for identifying the unbalanced pre-intervention variables is ineffective. This is because in large samples, p-values quickly tend to zero and are not necessarily indicative of variable importance (Lin et al., 2013).

This study presents classification and regression trees (“trees”) as the basis for an efficient empirical alternative to address selection biases that characterize research on impact assessment in business and policy. Our proposed tree solution is data-driven, does not make distributional assumptions about the selection model. The approach can easily generalize beyond the common case of a binary intervention (treatment/control) setting to multi-interventions and to continuous interventions. Hence, this method can be applied in a variety of settings, as we illustrate in Sections 3 and 4. Most important, it provides transparent, easily understandable results for communication and use in research and practice.

Despite these important advantages, we find limited application of classification and regression trees in impact assessment research in business and policy. Our study addresses this gap in the literature to illustrate the application of trees in impact assessment research using the case of two technology decisions. In the first study, we assess the impact of eGov services on various dimensions of performance in a sample of 9500 Indian citizens, by comparing online and offline users of passport services. In the second study, we compare the performance of technology and technology-enabled business process outsourcing contracts across different contractual pricing methodologies (polytomous self-selected interventions), and across different contract durations (continuous self-selected intervention).

2 A Classification and Regression Tree Approach

Classification and regression trees (“trees”) refer to a set of data mining algorithms that generate simple IF-THEN rules to link a set of predictors to an outcome. Classification trees are used when the outcome is categorical and regression trees when the outcome is continuous. Trees partition a data set into regions so that within each region, observations are as homogeneous as possible in terms of the outcome (Breiman et al., 1984). Each region, called a leaf node, is a sub-space of the multivariate predictors’ space, so that within each

leaf node of a tree, observations will have a similar profile in terms of their predictor values, and similar outcome probabilities or values. This approach makes trees useful for predicting outcome values for individual observations as well as for selecting important predictor variables (Shmueli et al., 2010).

Trees present important advantages over other models used to link an outcome to a set of inputs, including linear and logistic regression models. First, trees do not make parametric assumptions about the relationship between the outcome and input variables. In other words, the user need only specify the input and output variables for the analysis but not an equation linking them. The tree automatically detects and models relevant outcome-input relationships in a flexible manner. The tree algorithm will automatically search for not only the ‘best’ combination of input variables but also non-linear effects such as interactions. Further, trees can handle a variety of input variable types as well as missing values. Finally, regression models require the assessment of the linearity assumption, which is typically overlooked (Westreich et al., 2010), whereas trees do not make this assumption.

In the management literature⁴, trees are much less common than regression models. Prior research studies in management that do use trees typically focus on empirical prediction of outcomes for individual observations, including forecasting and personalization. However, non-predictive uses of trees to evaluate outcomes at a population level are relatively rare.

Next, we describe our tree-based approach for assessing the impact of a self-selected intervention that controls for such bias. We use trees⁵ for the entire evaluation, assuming that a large sample is available with a rich set of pre-intervention variables. In doing so, we use the following definitions and notation:

Y = One or more performance measures of interest

X = Pre-intervention variables that are suspected to affect the choice of the intervention

T = Intervention. For a single intervention (intervention/control) T is a binary variable. For multiple interventions, T is categorical with multiple categories. For a continuous intervention T is a continuous variable.

Our tree-based approach consists of five steps for assessing the impact of an intervention:

1. Generate the selection model: Fit a tree with T as the outcome and X ’s as predictors. For a binary or categorical T , use a classification tree. For a continuous T , use a regression tree.

⁴We performed a keyword search for all papers in *Management Science* and *Information Systems Research* between 1954 to April 2013 (mansci.journal.informs.org/search), with keywords *classification tree*, *decision tree*, *CART*, *regression tree* and examined the results manually for relevance.

⁵Several tree algorithms exist. We use conditional-inference trees (Hothorn et al., 2006), where predictor choice and splits are based on statistical tests of independence.

2. Present the visual of the resulting tree, showing the splitting variables and their values. Each leaf node should display the intervention statistics for its observations.
3. Treat each leaf node as the final ‘bin’ for measuring Y (as in Westreich et al., 2010). Within each leaf node, all observations have the same pre-intervention profile (X) and the same intervention probability ($P(T = 1 | X)$ for a binary T) or distribution (for a continuous T).
4. Evaluate the effect of the intervention by computing performance measures of interest (Y) and performing analyses of interest (t-test, regression model, non-parametric tests), separately within each leaf node.
5. Display the leaf-node-level performance analysis results using leaf-node-level charts.

Note that the tree resulting from step 1 will include only pre-intervention variables (X) that are unbalanced across the intervention groups. For example, if one group includes a significantly larger proportion of women compared to the other group, GENDER will appear in the tree. If the pre-intervention variables are completely balanced across the intervention groups, as would be the case in a balanced designed experiment, the tree will have no splits. The intervention groups can then be compared with no need for matching.

The power of the tree-based approach stems from its simplicity, communicability and generalizability, yet nuanced analysis. Trees are generated automatically, not requiring the user to specify the exact form of the relationship (step 1). Further, the graphical presentation of results of the matching procedure and performance analyses (steps 2 and 5) provides improved communication with non-experts. The tree structure (step 2) highlights the pre-intervention variables that are unbalanced across the intervention groups. This information can itself be insightful, and guide future studies in terms of pre-intervention variables to balance across intervention groups.

We rely on the tree to split the sample into subsamples that have homogeneous intervention propensities, and then perform performance outcome analysis separately for each leaf node. While in PSM, matched records share the same propensity score, they might have different pre-intervention profiles. In contrast, the tree leaf nodes share not only the same propensity score, but also the same pre-intervention profile, with respect to the unbalanced pre-intervention variables. In other words, the tree-based approach avoids matching records that have similar propensity scores but different pre-intervention profiles.

Despite these merits, the use of the tree-based approach has not been used in the management literature. Our study represents an early empirical demonstration of this approach to assessment of the impact of diverse self-selected managerial interventions.

3 Study 1: Impact of eGov Services in India

e-Governance (eGov) refers to the use of electronic channels, enabled by information and communications technology (ICT), in the delivery of government information and services. An assumption underlying the emergence and growth of eGov is that the latter offers governments the potential to transform their relationships with their citizens. This assumption is consistent with research (e.g., Heeks, 2001; West, 2000; Dutton, 1996), which finds that the benefits of eGov include lower administrative costs, increased transparency and accountability, faster and more accurate response to citizen requests, improved resource management, and improved quality of feedback to policy makers through harvesting of data that is standardized and integrated across operational systems. All of these benefits serve to reinvent the organization and functioning of the government.

Yet, there is evidence that eGov implementations may not always deliver intended benefits and it remains unclear whether the benefits of electronic channels demonstrated in the private sector are applicable to the public sector as well. Prior research (Morgeson and Mithas, 2009) suggests that this research question is ultimately an empirical question that is best addressed by examining the eGov implementations of a variety of governments.

The Government of India (GoI) launched a national eGov plan in the early 2000s. The plan included creating of an income tax portal, online delivery of passport services, and online delivery of services provided by the Ministry of Company Affairs (MCA), among others. Our first study assesses the impact of online delivery of passport services provided by GoI. eGov, in this context, is defined more narrowly as the government's use of websites to help citizens complete a significant portion of the passport application online. We describe the data that informs our impact assessment. We then report the official results published by the GoI using a naive approach that does not correct for self-selection. Subsequently, we assess the impact of electronic delivery of passport services using our proposed tree-based approach. We finally perform impact assessment using classic PSM and contrast it with our tree-based approach to emphasize important advantages of the latter.

In 2006, GoI commissioned a survey of users of all their electronic services. The Indian Institute of Management, Ahmedabad designed the surveys and overall assessment framework while eleven independent market research firms were empanelled to conduct the survey across the country. The survey queried users about various dimensions of their service experience including transaction costs, quality of service, and overall satisfaction. Our study uses data from this survey with focus on users of the electronic passport services. Table 1 describes some of the key outcomes measured in the survey to assess the impact of eGov. The survey also provided information on characteristics of individual respondents and the

latter’s ex ante perceptions of various service parameters such as clarity of rules and procedures or convenience of service facilities. A complete list of such variables that were used in this study and allied questions from the survey is provided in Table 3.

Responses to the survey of passport services constitute a representative sample of 13 passport offices selected from different regions of the country. A sample of 9500 users was drawn from cities/towns where these offices were located. The study was designed as a quasi-experiment, where a large group of offline users was selected to match a group of online users in terms of their geography and demographics.

3.1 Naive Analysis

Given the geographic and demographic match between the online and offline users, the impact of electronic delivery of passport services was measured and reported as the difference in average outcomes between the online and offline groups after testing the statistical significance of these differences. Model I in Table 2 presents the differences in mean performance between the two groups. The difference is statistically significant ($p < 0.05$) for all measures except % Paying bribes to passport office and % Using intermediary agent⁶. The results of this naive approach to impact assessment are presented in the official report by GoI.

3.2 Assessment using the tree-based approach

A drawback of the naive approach to impact assessment stems from that the study lacks the important component of randomization and does not qualify as a true experiment. Although the two samples of users are matched, individuals self-select the mode of service delivery that is likely to be influenced by multiple factors aside from geography and demographics. These factors that impact the individual’s decision to adopt electronic services may also impact ensuing performance outcomes.

To evaluate performance outcomes, while controlling for self-selection, we use the following variables in the analysis:

Y = Performance indicators (Table 1)

T = Passport application mode: online ($T = 1$) or offline ($T = 0$)

X = demographic and geographic variables as well as survey questions that are likely to affect choice of passport application mode (Table 3).

⁶Statistical significance is easily achieved due to the large sample size (see Lin et al., 2013). From a practical point of view, some of the differences might be considered practically insignificant, such as the difference in the average number of trips (2.49 vs. 2.2 trips)

Following the five-step approach, we first fit a classification tree of T on X (step 1). As step 2 prescribes, we present the tree graphically, showing the splitting variables, their values, and the interventions statistics at each leaf node. The resultant tree, shown in Figure 1, has a single split on the *Awareness* variable from the survey. The tree emphasizes two important issues. First, although we use the entire set of demographic, geographic and ex ante perceptions of service parameters measured in the survey in our analyses, the only variable picked up by the tree is the individual’s awareness of the government’s eGov initiatives. This result indicates that the online and offline samples are indeed balanced in terms of demographic and geographic information.

Second, the tree also emphasizes that while the online and offline samples are balanced along all other pre-intervention variables measured in the survey, they are not balanced in terms of *Awareness*. In other words, awareness of the e-Gov initiatives of GoI is most likely to influence citizens’ likelihood of using the online systems. Indeed, the tree shows that those individuals who responded that they were aware of the eGov initiatives were more likely to choose the online system (probability 0.72) compared to those unaware of these systems (probability 0.11). Hence, to assess the impact of the online system in terms of the various performance metrics, we must compare the online and offline groups conditional on their awareness of such systems.

Following steps 3 and 4, performance analyses are done separately at each leaf node, i.e., online and offline performance is compared separately for the aware and unaware groups. The separate analyses avoid selection bias due to the *Awareness* variable. The results are shown in Figure 2 (right panels) and Table 2 (Model III), in line with step 5.

We find that the direction of the effect (online versus offline) is identical in both the aware and unaware groups across all but one performance measure. In particular, in the ‘aware’ group, users of the online system were 2% less likely to employ an intermediary agent ($p < 0.05$) while in the ‘unaware’ group, users were 9% more likely to use an agent ($p < 0.001$). The results offer an interesting contrast to the naive analysis (Model I), which suggested that the impact of the online system on the use of an intermediary agent is statistically insignificant. However, as note above, this is an important outcome in the self-selection corrected tree-based approach.

Another interesting difference between Model I and Model III is the impact of the online system on the proportion of users paying bribes to the police. For the latter, we see a classic Simpson’s Paradox effect (Simpson, 1951; Alin, 2010; Pearl, 2011): the aggregate data show a higher rate of bribes in the online group, whereas in each of the aware and unaware groups, the offline group has a higher rate of bribe payers.

3.3 Comparison with classic PSM

We repeated the analysis using PSM with logistic regression⁷ using the same variables. Given the lack of theory guiding the exact specification of the relationship between X and T in our context, we do not include higher order variables or interactions in the selection model.

The results of the logistic regression of T on X (Table 4) highlight important limitations of using PSM to control for self-selection. First, using statistical significance of the pre-intervention coefficients to determine unbalanced variables is inappropriate due to the large sample size (Lin et al., 2013). While *awareness* is statistically significant, so are most of the other pre-intervention variables.

Next, the estimated logistic model is used to compute propensity scores that are used to create a matched sample. Each record in the intervention (online) group is matched with a record in the control (offline) group that has the closest propensity score or likelihood of selecting the intervention. Comparing the performance of the intervention group with the matched control group yields performance estimates corrected for self-selection.

Model II in Table 2 presents the performance analysis for the PSM approach. Overall, the outcome values estimated using PSM are very close to those from the naive approach (Model I). In other words, there is not much correction for self-selection. In particular, despite the statistical significance and relatively large coefficient for *awareness*, its effect on outcome adjustments is minor.⁸

Compared to our tree-based approach, which discovered the effect of the online system on usage of an intermediary agent, with effects in opposite directions for the aware and unaware groups, in the PSM approach the imbalancing effect of *awareness* is lost, and we do not observe the effect at all.

4 Study 2: Performance of Outsourcing Contracts

Outsourcing is fundamental to firm competitiveness. Firms are increasingly externalizing a variety of core business functions such as R&D, product development and marketing across a wide range of industries to achieve strategic objectives such as faster time to market, revenue increases and business transformation, all of which shape a firm’s competitive position. Yet, despite their extended reach and impact, outsourcing initiatives involve high failure rates with adverse impacts on various critical performance metrics⁹. Prior research in information

⁷We implemented PSM using R’s *matchit* function. Similar results were obtained using Stata’s *psmatch2*.

⁸Simply replacing the logit with a tree function in the PSM procedure yields nearly identical numbers.

⁹70% of the respondents in a 2005 survey by Deloitte Consulting 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, and 20% of such contracts will collapse (Johnson, 2006).

systems attributes heterogeneity in outsourcing performance to efficacy of contract design and management (Susarla and Barua, 2011). The right contract aligns incentives between the client and the vendor to engender cooperative behavior that is necessary for execution of interdependencies in the outsourced task and effective 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, complex outsourcing initiatives that involve a higher probability that adaptations are needed, should be governed by less complete variable price contracts, whereas simpler, more stable outsourcing initiatives should lead to more complete fixed price contracts that seek to primarily provide high powered incentives to the vendor to reduce costs of ownership of the outsourced function.

Similarly, contract duration too reflects a tradeoff between providing ex ante incentives for specific investments 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 such adaptation but are difficult and costly to design and administer. A central goal in contract design, therefore, is to choose a contract duration that “maintains incentives for efficient adaptation while minimizing the need for costly adjudication and enforcement” (Crocker and Masten, 1988). 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.

Our empirical analysis is based on over 1400 outsourcing initiatives implemented between 1996 and 2008. Information on the outsourcing initiatives and their governing contracts is obtained from International Data Corporation’s (IDC) services contracts database. IDC data on the top 100 outsourcing contracts signed each year includes contract value, length, announcement and signing date, geography, industry, outsourcing type, and a detailed description of the service provided. We use Lexis-Nexis and the Dow Jones News Retrieval Service to verify and supplement IDC information on 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 final sample of 1411 contracts includes 374 vendors and 710 clients. The opera-

tionalization of key variables in our analyses, including contract price and contract length, is described in Table 5.

4.1 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., 2010) 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 as discussed earlier, 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 apriori and the data are used to confirm these analyses. 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 - it would be useful to use an exploratory view to identify pre-intervention characteristics 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 5 describes the X , T and Y variables used in our study. Figure 3 portrays the tree of price mechanism on pre-intervention variables. We find that relational variables indicative of mutual trust and task variables or total contract value indicative of complexity of the outsourcing engagement, explain heterogeneity in contract choice.

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. The persistence of systematically nonzero abnormal returns for a long horizon following an event suggests that financial markets are inefficient in pricing the event when it occurs.

Figure 4 presents the performance comparison through charts. Each row corresponds to one performance measure. Left panels portray the baseline ‘naive’ comparison of mean performance values across the different contractual pricing mechanisms. Right panels (columns 2-6) portray the performance for subsets of the data, which correct for the mechanism of self-selection. Each of the five right columns corresponds to a leaf node in the classification tree in Figure 3.

4.2 Naive Analysis

A naive 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 (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 the value of outsourcing contracts into stock prices. Given 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., 2012). Yet, they provide an incomplete picture of outsourcing value. Important questions remain: do all simple outsourcing 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?

4.3 Tree-based Analysis

An analysis that accounts for self-selection into the various contract types tells a slightly different, more nuanced story compared to the naive analysis, thereby, addressing the questions raised thereof. Examining the leaf-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 across 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 also reflected in differences in contract value between these 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 IT outsourcing engagements involve high levels of specificity, uncertainty and incompleteness in task specification. As a result, node 2, which also comprises contracts that are indicative of high levels of trust between the client and vendor, 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 revealed in node 2 is that fixed price contracts that are characterized by high levels of mutual 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. Indeed, prior research finds that service delivery often presents tradeoffs between client profitability and satisfaction (e.g., Anderson et al., 1997) or profitability and innovation (e.g., Mani et al., 2010). Yet, prior empirical research on the value of outsourcing has largely focused on profitability or income gains from outsourcing to conclude that fixed price contracts create value when the likelihood of ex post renegotiation is low, either because of reduced relational uncertainty arising from mutual trust or because of reduced task uncertainty. 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, while the market does not price the value of these efficiency gains in the announcement period (-5, +5), 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. It is interesting that these contracts create value for the client irrespective of the nature of the outsourced task.

Lastly, 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.

4.4 Analysis 2: Impact of Contract Durations

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 underin-

vestment (e.g., Crocker and Masten, 1988; Susarla et al., 2012). The limited theoretical development of 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.

4.5 Naive Analysis

The top row of Figure 6 presents a naive 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.

4.6 Tree-based Analysis

Figure 5 presents the regression tree for duration of the outsourcing contract as a function of pre-intervention variables. We find that task variables and prior experience of the client in managing similar inter-firm alliances, explain heterogeneity in contract duration in our data. The shortest durations (nodes 1, 6, 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, 5). These results suggest that for complex engagements, where the likelihood of specific or non-contractible investments may be higher, firms implement contracts of longer duration to provide ex ante incentives for the vendor to undertake these investments.

In addition, in contracts of low to moderate complexity, where fewer specific investments are required, prior experience of the clients leads them to reduce ex post inefficiencies of lock-in and maladaptation through implementing shorter duration contracts.

For each of the leaf nodes, we compare market and financial performance gains to the outsourcing client across different contractual lengths. The results are displayed in Figure 6.

Compared to the naive 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: (a) high value contracts (*Services.Contract.Value* > 140 million) in outsourcing of technology-enabled business processes and functions, IS outsourcing, and outsourcing of application, network and desktop management, where the clients have prior experience in managing similar alliances (node 5), and (b) low value, custom IT outsourcing contracts (node 8).

Prior cumulative experience of the clients helps in two ways. First, it helps clients find potentially useful solutions to inefficiencies in long-term contracts such as price or performance lock-in. Second, the greater the prior experience of the client, the easier for the firm to interpret and respond to unforeseen contingencies that are common in long-term outsourcing contracts.

Result (b) suggests that in outsourcing engagements that require specific investments, the benefits of long-term contracts are limited to those where the scope is minimal. In engagements with larger scope, the costs may outweigh the 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 stock prices during the announcement period.

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 outsourcing contract. 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.

5 Conclusions

We introduced a methodology for evaluating causal performance effects of diverse self-selected managerial interventions given large samples of observational data. Studies in management (e.g., Agarwal et al., 2009; Overby and Jap, 2009; Gopal et al., 2003) have traditionally corrected for the selection bias characterizing these interventions using the Heckman model or PSM. While both methods have been particularly useful in producing improved estimates of average treatment effects, they are subject to important limitations that we seek to address in this study using the tree-based approach. Similar to the PSM approach, our tree-based approach assumes that the propensity to self-select an intervention is observable. However, in contrast to the Heckman model and PSM that assume that the selection processes are known and are correctly specified in the selection model, our tree-based approach is a data-oriented method that makes no assumptions about the form of the selection model. Given the growing availability and interest in large samples of data with large numbers of measurements (Big Data), our data-oriented approach can generate more nuanced results, both in terms of identifying pre-intervention variables that are unbalanced across intervention and control groups, and outcome analyses that control for more complex relationships between the outcomes and the intervention.

References

- Agarwal, R., Animesh, A., and Prasad, K. (2009). Research notesocial interactions and the digital divide: Explaining variations in internet use. *Information Systems Research*, 20(2):277–294.
- Alin, A. (2010). Simpson’s paradox. *Wiley Interdisciplinary Reviews: Computational Statistics*, 2(2):247–250.
- Anderson, E. W., Fornell, C., and Rust, R. T. (1997). Customer satisfaction, productivity, and profitability: Differences between goods and services. *Marketing Science*, pages 129–145.
- Bajari, P. and Tadelis, S. (2001). Incentives versus transaction costs: A theory of procurement contracts. *RAND Journal of Economics*, pages 387–407.
- Brieman, L., Friedman, J., Olshen, R., and Stone, C. (1984). *Classification and Regression Trees*. Chapman & Hall, London.
- Crocker, K. J. and Masten, S. E. (1988). Mitigating contractual hazards: Unilateral options and contract length. *The RAND journal of economics*, pages 327–343.
- Dehejja, R. and Wahba, S. (1999). Causal effects in nonexperimental studies: Reevaluating the evaluation of training programs. *Journal of American Statistical Association* (94:448), pages 1063–1062.
- Dehejja, R. and Wahba, S. (2002). Propensity score-matching methods for nonexperimental causal studies. *Review of Economics and Statistics* (84:1), pages 151–161.
- Dutton, W. (1996). *Information and Communication Technologies: Visions and Realities*. Oxford: Oxford University Press.
- Fama, E. and French, K. R. (1993). Common risk factors in the returns on stocks and bonds. *Journal of Financial Economics*, 33(1):356.
- Gao, N. (2006). What does stock and accounting performance tell us about outsourcing? Working paper, SSRN: <http://ssrn.com/abstract=681321>.
- Gopal, A., Sivaramakrishnan, K., Krishnan, M. S., and Mukhopadhyay, T. (2003). Contracts in offshore software development: An empirical analysis. *Management Science*, 49(12):1671–1683.

- Guo, S. and Fraser, M. W. (2009). *Propensity Score Analysis: Statistical Methods and Applications*. Sage Publications.
- Heckman, J., Ichimura, H., and Todd, P. (1997). Matching as an econometric evaluation estimator: Evidence from evaluating a job training programme. *Review of Economic Studies*, pages 605–654.
- Heckman, J., Ichimura, H., and Todd, P. (1998). Matching as an econometric evaluation estimator. *Review of Economic Studies*, pages 261–294.
- Heckman, J. J. (1979). Sample selection bias as a specification error. *Econometrica: Journal of the econometric society*, pages 153–161.
- Heeks, R. (2001). Understanding e-governance for development. i-Government Working Paper Series, No. 11, Manchester: Institute for Development Policy and Management, University of Manchester.
- Heinrich, C., Maffioli, A., and Vázquez, G. (2010). A primer for applying propensity-score matching. Technical report, Inter-American Development Bank, Office of Strategic Planning and Development Effectiveness (SPD).
- Hilbe, J. (2009). *Logistic Regression Models*. Chapman & Hall/CRC Press.
- Holland, P. (1986). Statistics and causal inference (with discussion). *Journal of American Statistical Association (81:396)*, pages 261–294.
- Hothorn, T., Hornik, K., and Zeileis, A. (2006). Unbiased recursive-partitioning: A conditional inference framework. *Journal of Computational and Graphical Statistics*.
- Johnson, L. K. (2006). Successful business process outsourcing. *MIT Sloan management review*, 47(2):5–6.
- Lee, B., Lessler, J., and Stuart, E. (2010). Improving propensity score weighting using machine learning. *Statistics in Medicine*, pages 337–346.
- Lin, M., Lucas, H. C., and Shmueli, G. (2013). Too big to fail: Large samples and the p-value problem. *Information Systems Research*.
- Little, R. and Rubin, D. (2000). Causal effects in clinical and epidemiological studies via potential outcomes: concepts and analytical approaches. *Annals Review of Public Health*, pages 121–145.

- Mani, D., Barua, A., and Whinston, A. (2010). Does strategic outsourcing create financial value? In Pinedo, M., editor, *Operational control in asset management: Processes and Costs*, pages 190–208. SimCorp Strategy Lab, 1 edition.
- 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):618–634.
- Mithas, S. and Almirall, D. (2006). Do crm systems cause one-to-one marketing effectiveness? *Statistical Science*, pages 223–233.
- Mithas, S., Almirall, D., and Krishnan, M. (2009). A potential outcomes approach to assess causality in information systems research in: Economics, information systems and electronic commerce research ii: Advanced empirical methodologies, r.j. kauffman and p.p. tallon (eds). Technical report, Armonk, New York.
- Mithas, S. and Krishnan, M. (2009). From association to causation via a potential outcomes approach. *Information Systems Research*, pages 295–313.
- Mithas, S. and Lucas, H. (2010). Are foreign it workers cheaper? u.s. visa policies and compensation of information technology professionals. *Management Science* (56:5), pages 745–765.
- Mithas, S., Lucas, H., and Fornell, C. (2005). Why do customer relationship management applications affect customer satisfaction? *Journal of Marketing*, pages 201–209.
- Morgeson, I. F. and Mithas, S. (2009). Does e-government measure up to e-business? comparing end-user perceptions of u.s. federal government and e-business websites. *Public Administration Review*, 69:740752.
- Overby, E. and Jap, S. (2009). Electronic and physical market channels: A multiyear investigation in a market for products of uncertain quality. *Management Science*, 55(6):940–957.
- Pearl, J. (2011). Simpson’s paradox: An anatomy.
- Rosenbaum, P. and Rubin, D. (1983). The central role of the propensity score in observational studies for causal effects. *Biometrika* (70:1), pages 41–55.
- Rubin, D. and Waterman, R. (2006). Estimating the causal effects of marketing interventions using propensity score methodology. *Statistical Science*, pages 206–222.

- Shmueli, G., Patel, N., and Bruce, P. (2010). *Data Mining for Business intelligence: Concepts, Techniques and Applications in MS Excel with XLMiner*. Wiley, 2 edition.
- Simpson, E. H. (1951). The interpretation of interaction in contingency tables. *Journal of the Royal Statistical Society. Series B (Methodological)*, pages 238–241.
- Stuart, E. (2010). Matching methods for causal inference: A review and a look forward. *Statistical Science*, pages 1–21.
- Susarla, A. and Barua, A. (2011). Contracting efficiency and new firm survival in markets enabled by information technology. *Information Systems Research*, 22(2):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):23–41.
- West, D. (2000). State and federal e-government in the united states 2000. Available at www.insidepolitics.org/policyreports.html.
- Westreich, D., Lessler, J., and Funk, M. (2010). Propensity score estimation: machine learning and classification methods as alternatives to logistic regression. *Journal of Clinical Epidemiology (63:8)*, pages 826–833.
- Winship, C. and Sobel, M. (2004). *Causal Inference in Sociological studies*, pages 481–503. Sage, London.

Table 1: Performance outcomes, as defined by the government

Dimension of Impact	Indicator
Cost of Availing Service (Measured Directly)	Number of trips made for the service
	Average travel cost of making each trip
	Average waiting time in each trip
	Estimate of wage loss due to time spent in availing the service
	Total time elapsed in availing service
	Amount paid as bribe to functionaries
	Amount paid to agents to facilitate service
Overall Assessment	Preference for manual versus computerized system
	Composite Score measured on 5-point scale factoring in the key attributes of delivery system seen to be important by users
Quality of Service	Interaction with staff, complaint handling, privacy, accuracy measured on 5-point scale
Quality of Governance	Transparency, participation, accountability, corruption measured on a 5-point scale

Table 2: Comparing average outcomes of online/offline passport systems: Naive, PSM, and tree-based approaches. Numbers in parentheses are standard errors.

	Model I: Naive Approach (raw data)		Model III: PSM Logistic		Model II: Tree Approach			
					Unaware of Computerization		Aware of Computerization	
	Offline <small>n = 4551</small>	Online <small>n = 4575</small>	Offline <small>n = 4209</small>	Online <small>n = 4209</small>	Offline <small>n = 2968</small>	Online <small>n = 387</small>	Offline <small>n = 1583</small>	Online <small>n = 4188</small>
Paying bribes to police (%)	42.43 (0.73)	48.09 (0.74)	42.20 (0.76)	47.73 (0.77)	38.48 (0.89)	34.63 (2.42)	49.84 (1.26)	49.33 (0.77)
Paying bribes to passport office (%)	4.33 (0.30)	4.42 (0.30)	4.25 (0.31)	4.63 (0.32)	2.43 (0.28)	1.29 (0.57)	7.90 (0.68)	4.70 (0.33)
Preference for online system (%)	59.37 (0.73)	90.43 (0.44)	60.04 (0.76)	90.64 (0.45)	61.49 (0.89)	86.56 (1.74)	55.40 (1.25)	90.78 (0.45)
Using intermediary agent (%)	34.83 (0.71)	33.64 (0.70)	34.00 (0.73)	32.40 (0.72)	34.74 (0.87)	44.19 (2.53)	35.00 (1.20)	32.66 (0.72)
Number of trips	2.49 (0.02)	2.20 (0.02)	2.48 (0.02)	2.22 (0.02)	2.42 (0.02)	2.21 (0.06)	2.63 (0.03)	2.20 (0.02)
Waiting time (min)	97.95 (0.94)	88.18 (0.89)	98.51 (0.88)	88.77 (0.93)	94.75 1.13	69.67 (2.38)	103.90 (1.67)	89.91 (0.94)
Total elapsed time (days)	43.04 (0.40)	41.12 (0.39)	42.73 (0.41)	40.91 (0.40)	42.58 (0.49)	41.55 (1.08)	43.90 (0.68)	41.08 (0.41)
Overall quality of service	3.51 (0.01)	3.63 (0.01)	3.53 (0.01)	3.61 (0.01)	3.52 (0.02)	3.72 (0.04)	3.49 (0.02)	3.62 (0.01)
Overall governance	3.30 (0.01)	3.48 (0.01)	3.32 (0.01)	3.48 (0.01)	3.30 (0.02)	3.45 (0.04)	3.31 (0.02)	3.48 (0.01)
Overall trust in the government	3.42 (0.01)	3.48 (0.01)	3.44 (0.02)	3.47 (0.02)	3.42 (0.02)	3.56 (0.05)	3.40 (0.02)	3.48 (0.02)

Table 3: Questionnaire questions used as pre-intervention variables in the selection model

Variable	Variable Description	Scale
Awareness	Awareness of electronic services provided by the Government of India	Binary, 1='aware'
Availability	Ease of availability and accessibility of information pertaining to the service	5-point Likert scale
Experience	Indicator of prior usage of any other e-Gov application	Binary, 1='prior usage'
Clarity of Processes	Extent to which the processes and procedures characterizing the e-Gov application are clear and simple	5-point Likert scale
Clarity of Rules and Procedures	Extent to which the rules and procedures characterizing the e-Gov application are stated clearly without ambiguity and mistakes	5-point Likert scale
Convenience of Hours	Extent to which the working hours of the passport center or office are perceived as convenient	5-point Likert scale
Convenience of Location	Extent of satisfaction with the present location of the passport center or office	5-point Likert scale
Form Design	Extent of satisfaction with the design and layout of the application forms	5-point Likert scale
Service Area Facilities	Extent of satisfaction with the service area facilities	5-point Likert scale

Table 4: Estimated logistic regression for online/offline on pre-intervention variables

Pre-intervention Variable	Estimate	Std. Error	z value	p-value
(Intercept)	-2.41	0.37	-6.52	0.00
HH_roleHead of HH	-0.09	0.08	-1.16	0.25
HH_roleOthers	0.06	0.24	0.25	0.80
HH_roleServent	1.13	0.54	2.09	0.04
HH_roleSpouse	0.29	0.10	2.81	0.01
educationGraduate & Above	-0.11	0.09	-1.24	0.22
educationHigher Secondary	-0.11	0.11	-1.01	0.31
educationIlliterate	1.78	0.33	5.41	0.00
educationLiterate without Education/Below Primary	0.10	0.24	0.44	0.66
educationMatric	-0.02	0.13	-0.12	0.91
educationMiddle	0.16	0.18	0.87	0.39
educationOthers	0.72	0.27	2.66	0.01
educationPrimary	0.36	0.21	1.76	0.08
houseSemi-Permanent	0.44	0.13	3.41	0.00
houseTemporary	-0.57	0.12	-4.80	0.00
houseUnclassified	0.32	0.19	1.66	0.10
cityBareilly	-0.21	0.24	-0.90	0.37
cityCochin	-0.55	0.17	-3.26	0.00
cityDelhi	-0.68	0.15	-4.53	0.00
cityHyderabad	-0.12	0.16	-0.74	0.46
cityJalandhar	-0.23	0.18	-1.31	0.19
cityJammu	0.15	0.34	0.45	0.65
cityKolkata	-0.67	0.16	-4.07	0.00
cityMallapuram	-0.27	0.22	-1.22	0.22
cityMumbai	-0.94	0.17	-5.60	0.00
cityPanaji	0.40	0.28	1.42	0.16
cityRanchi	-0.00	0.26	-0.02	0.99
citySurat	0.16	0.24	0.68	0.49
genderMale	-0.07	0.08	-0.82	0.41
age_group 30-40	0.17	0.08	2.09	0.04
age_group 40-60	0.02	0.09	0.24	0.81
age_group 60 Above	0.03	0.19	0.17	0.86
age_group Below 18	-0.09	0.42	-0.21	0.84
occupation Businessman with 10+ Employees	0.75	0.16	4.70	0.00
occupation Clerical Person	0.37	0.13	2.83	0.00
occupation Cultivators	1.00	0.35	2.87	0.00
occupation Dependent	0.09	0.23	0.37	0.71
occupation Executive/Managerial Level	0.09	0.11	0.81	0.42
occupation Household Duties	0.26	0.14	1.88	0.06
occupation Household Industry Worker	1.31	0.29	4.51	0.00
occupation Others	0.10	0.17	0.59	0.55
occupation Pensioner	0.19	0.22	0.87	0.39
occupation Self Employed	0.27	0.11	2.47	0.01
occupation Student	0.29	0.13	2.22	0.03

Table 4 Continuation

Pre-intervention Variable	Estimate	Std. Error	z value	p-value
occupation Supervisory level	0.28	0.13	2.07	0.04
income Below 500	0.47	0.34	1.38	0.17
income 500-1000	0.23	0.50	0.46	0.64
income 3000-5000	0.43	0.28	1.55	0.12
income 5000-7000	0.05	0.25	0.19	0.85
income 7000-10000	0.11	0.24	0.48	0.63
income More than 10000	0.16	0.23	0.70	0.49
experience	0.49	0.07	7.41	0.00
awareness	3.40	0.08	44.96	0.00
clarity_process	-0.16	0.04	-4.09	0.00
clarity_rules	0.17	0.03	5.00	0.00
service_area_facilities	0.11	0.04	2.72	0.01
form_design	-0.08	0.04	-2.29	0.02
convenience_location	-0.06	0.03	-1.83	0.07
availability	-0.02	0.03	-0.57	0.57

Table 5: Outsourcing contracts data: Variable descriptions

Variable Name	Description
Contract.Price (Intervention)	<p>Contract pricing mechanism, with six categories:</p> <p>Fixed Price Fixed payment per billing cycle</p> <p>Transactional Price Fixed payment per transaction per billing cycle</p> <p>Time and Materials Payment based on input time and materials used during the billing cycle</p> <p>Incentive Payment based on output improvements against key performance indicators or any combination of indicators</p> <p>Combination A combination of any of the above contract types, largely fixed price and time and materials</p> <p>Joint Venture A separately incorporated entity, jointly owned by the client and the vendor, used to govern the outsourcing relationship.</p>
Contract.Length (Intervention)	Duration of the contract, in months

Table 5 Continuation

Variable Name	Description
Task.Type (Engagement)	Task type is one of five categories: <ol style="list-style-type: none"> 1. Outsourcing of technology-enabled business processes and functions (includes Business Consulting and Business Process Outsourcing) 2. Custom technology outsourcing, including applications, networks and other systems 3. Information Systems Outsourcing 4. Outsourcing of Application, Network and Desktop Management 5. Training, Deploy and Support
Bid.Type (Engagement)	Prior cooperative association between the firms, indicative of trust: <p>Competitive bidding Absence of prior association between the firms.</p> <p>Incumbent The vendor has an existing relationship with the client.</p> <p>Sole-sourced The selected vendor is the only provider of the outsourced function</p> <p>Non-competitive contracts that are not sole-sourced or outsourced to an incumbent vendor.</p>
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 ($\hat{\epsilon}_{it}$, for firm i on day t), computed as $\hat{\epsilon}_{it} = r_{it} - \hat{r}_{it}$, where r_{it} is the daily return (to the value weighted S&P) estimated from the market model: $r_{it} = \alpha_i + \beta_i r_{mt} + \epsilon_{it}$. 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.

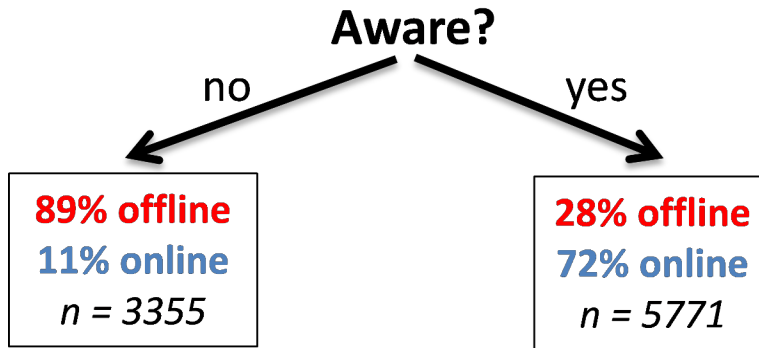


Figure 1: Classification tree for T=online/offline passport service as a function of pre-intervention variables.

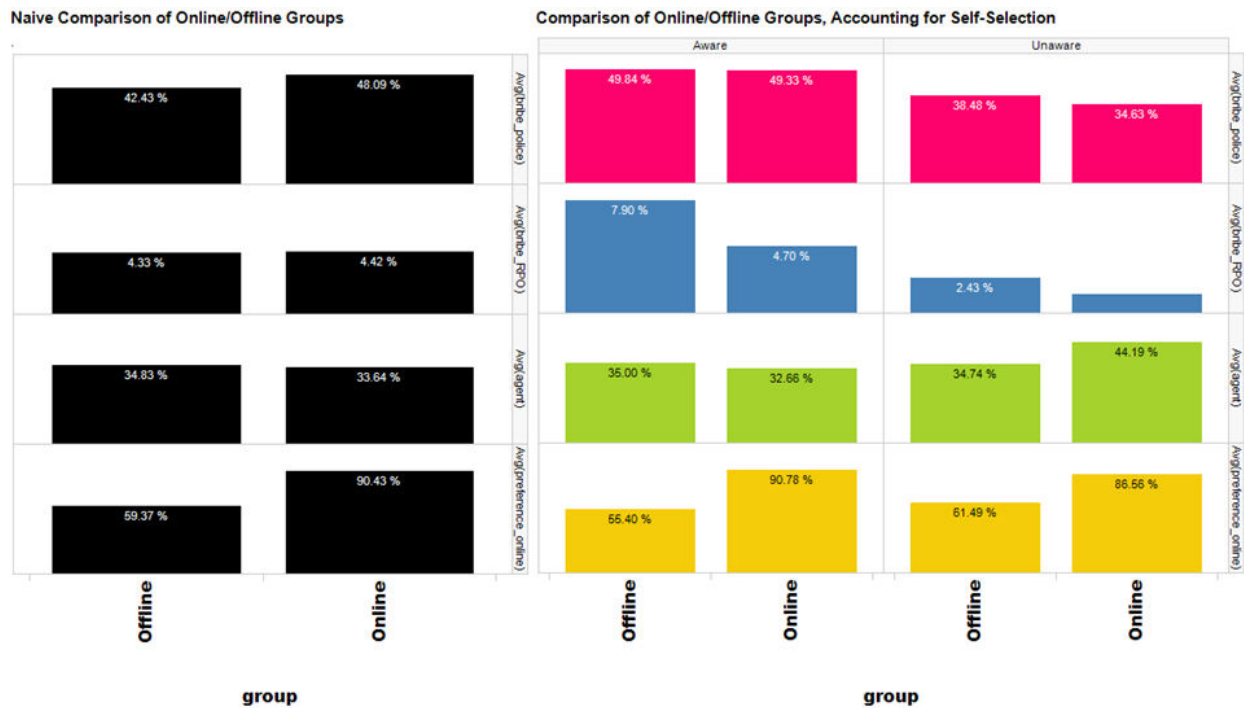


Figure 2: Distribution of outcomes for each of the two leaf nodes (right panel), compared to naive comparison (left panel). The effect of the online system on *usage of an intermediary agent* (third row) is reversed for the aware and unaware groups.

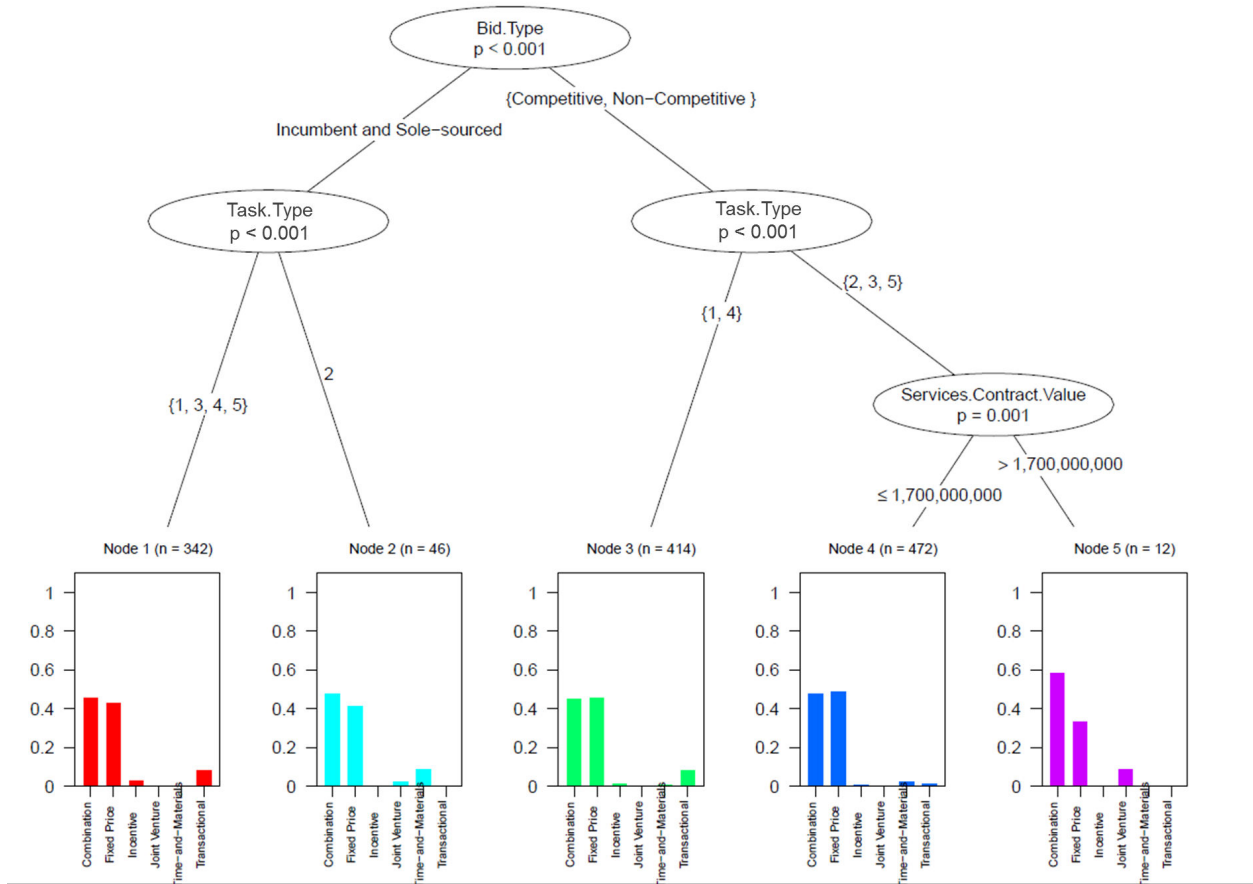


Figure 3: Classification tree for price methodology as a function of pre-intervention variables. Bar charts in leaf nodes represent the proportion of contracts for each price methodology.

Performance Comparison

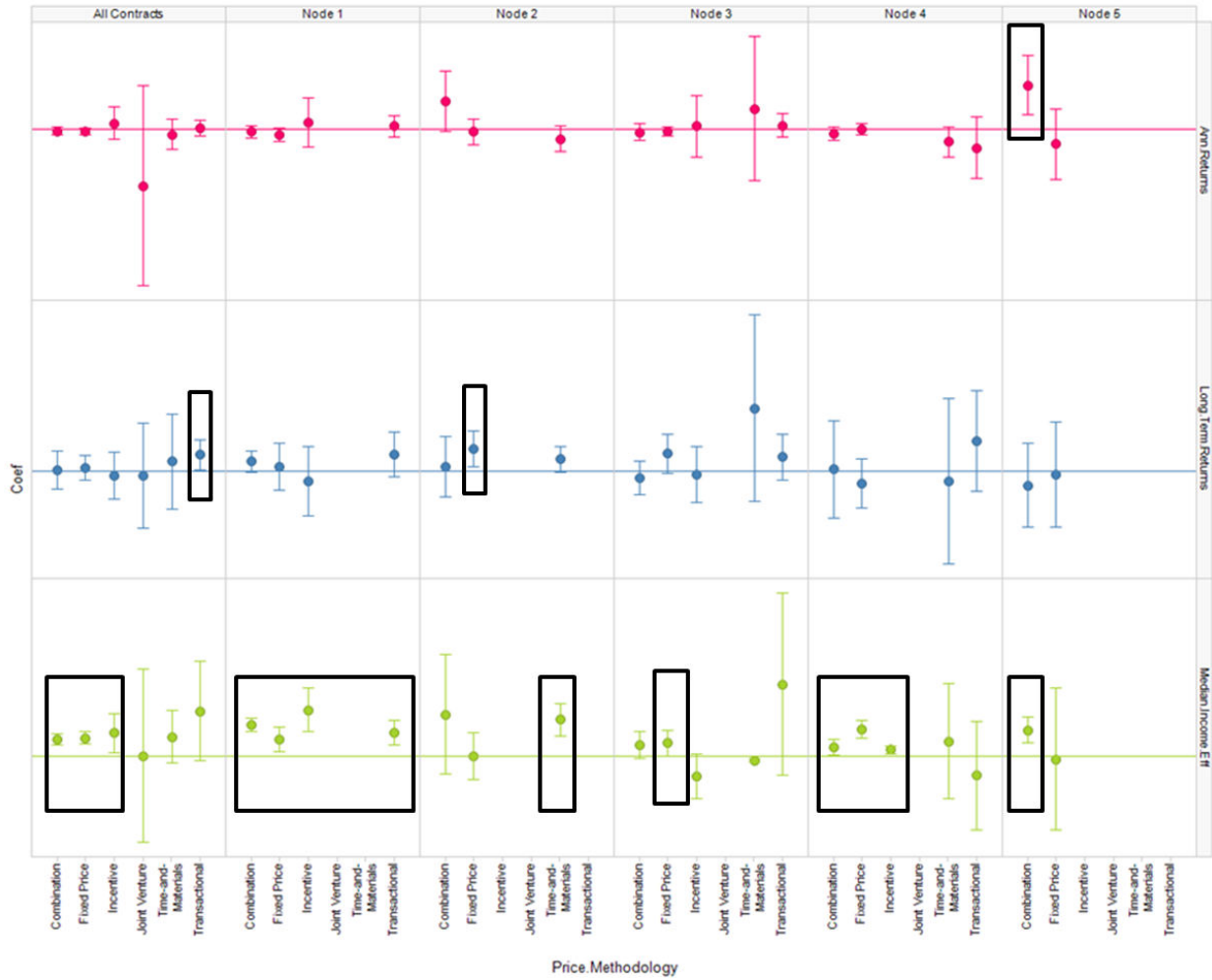


Figure 4: Comparing performance of six price methodologies. Left column: naive approach. Right columns: accounting for self selection using node-level comparisons. Circles represent mean performance and error bars are 90% confidence intervals for the mean. Rectangles mark significant coefficients.

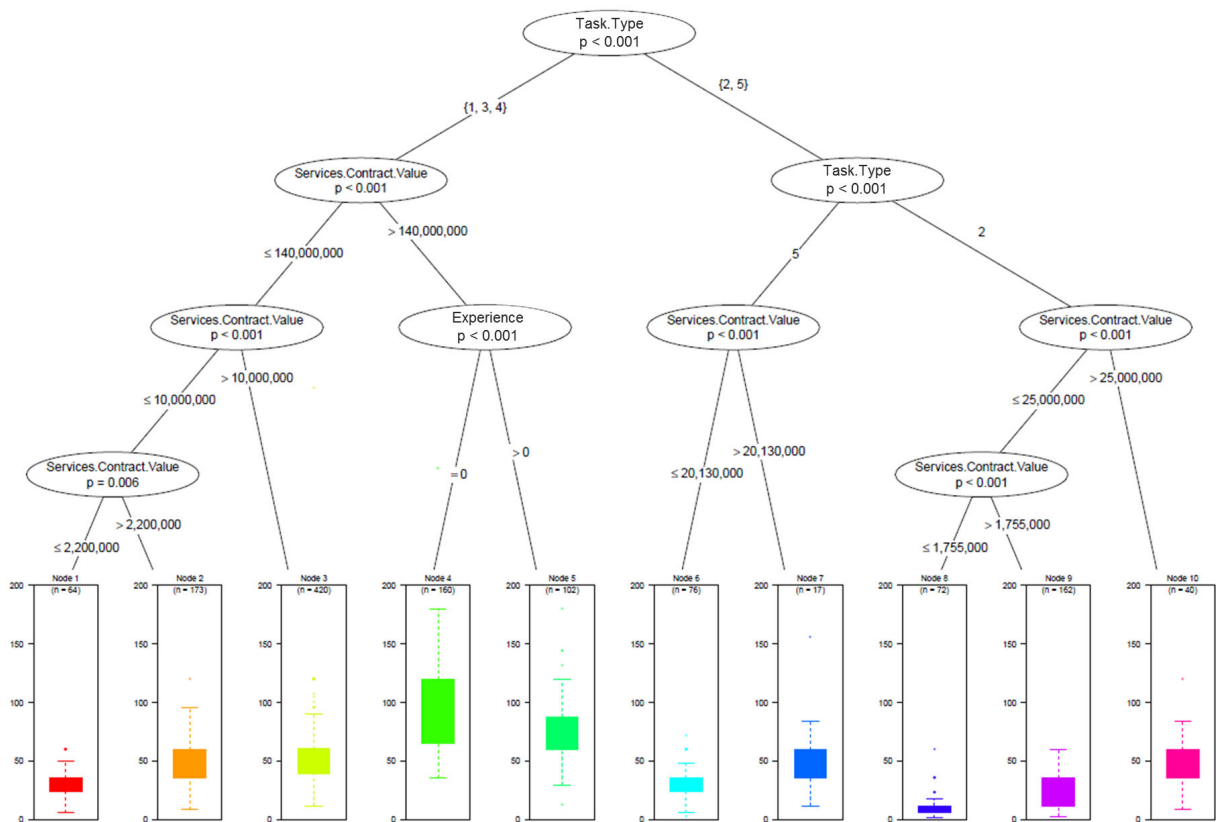


Figure 5: Regression tree for contract duration as a function of pre-intervention variables. Box plots in leaf nodes represent contract duration distribution in that node.

Performance across different contract lengths: Naive Approach



Performance across different contract lengths by node: Accounting for self-selection

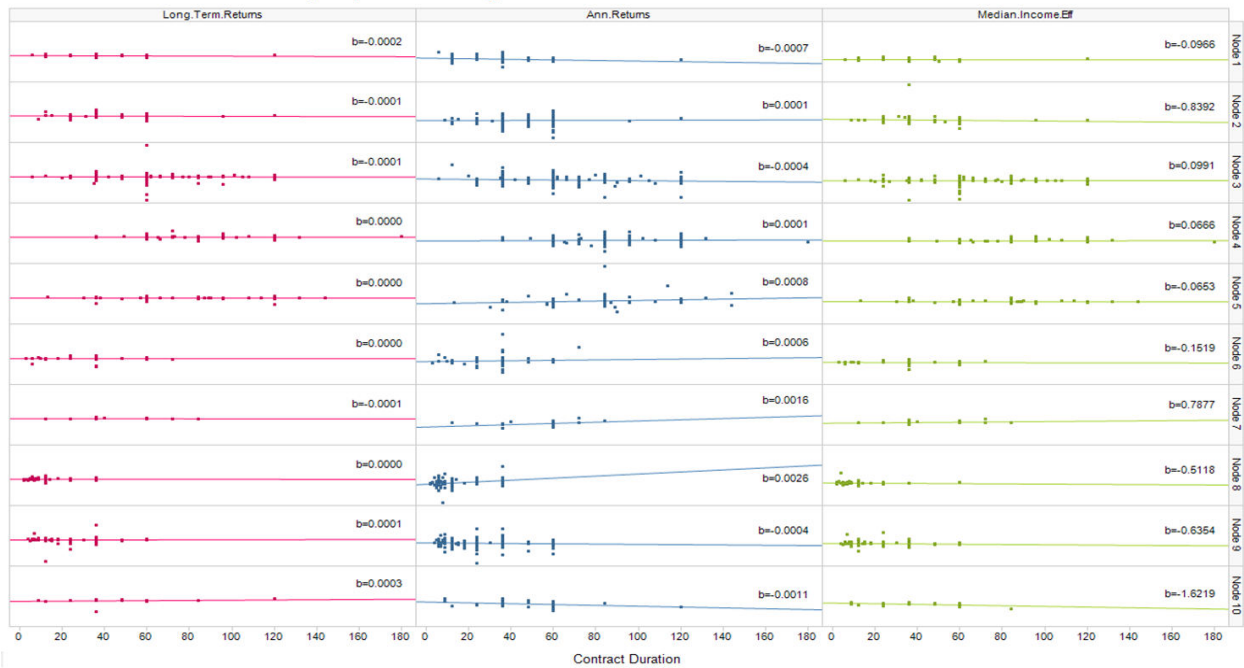


Figure 6: Comparing effect of contract length on performance. Top row: naive approach. Rows 2-11: accounting for self selection using node-level comparisons.