Stanford Computational Antitrust

ARTICLE

Information and Transparency: Using Machine Learning to Detect Communication Between Firms

David P. Brown,* Daniel O. Cajueiro,** Andrew Eckert,***
& Douglas Silveira****

Abstract. A high degree of publicly available information in a market can enhance competition but can also facilitate coordination between firms. A recent example highlighting concerns over the use of publicly available information to communicate between firms involves the Alberta wholesale electricity market, where, until recently, anonymized bidding information was released in near real-time. Allegations were raised that firms were using unique patterns in their bids to reveal their identities to rival firms and coordinate on higher prices. This paper uses machine learning techniques to examine the extent to which firms could identify the bids of their rivals in public data, and to describe the patterns observed in these data. These techniques can be employed as possible screens to evaluate whether publicly available information is being used to identify rival behavior and facilitate coordination.

* Department of Economics, University of Alberta, Edmonton, Alberta, Canada.
** University of Brasilia, Department of Economics, Brasilia, Federal District, Brazil.
*** Department of Economics, University of Alberta, Edmonton, Alberta, Canada.
**** Department of Economics, University of Alberta, Edmonton, Alberta, Canada.

Acknowledgements. This research project received support from the Government of Canada's *Canada First Research Excellence Fund* under the Future Energy Systems Research Initiative (Brown, Eckert) and the Social Sciences and Humanities Research Council's Canada Research Chair program (Brown). Daniel O. Cajueiro thanks CNPQ (Grant 302629/2019-0) and FAPDF (Grant 00193.00001796/2022-85) for financial support.

I. Introduction

There is a longstanding debate over the impact of information and transparency on market competition. It has been shown that information can enhance competition by facilitating customers searching for alternative products or lower prices from competitors in settings where production costs are unobservable but positively correlated, and by allowing firms to respond quickly to the changing market environment.[1] Alternatively, it is well known that information can facilitate coordination and communication across firms by allowing the monitoring of rival behavior.[2]

Regulators often publish de-identified data to mitigate the potential concerns associated with information disclosure. However, sophisticated firms may undertake actions to reveal their identity by behaving in conspicuous ways. This behavior could allow firms to coordinate on high-priced outcomes by, for example, setting a high price with a unique pattern to communicate to rivals that the firm intends to hold their price at a high-level for an extended period of time. As a result, it is important to evaluate firms' abilities to identify rival behavior using publicly available information. Monitoring and detection can be difficult in market settings where firms interact repeatedly and can use a multitude of possible pathways to reveal information. This creates the need to develop flexible empirical methods that can leverage the large dimensionality of the data. We study such a case by looking at wholesale electricity markets and employing machine learning techniques.

In this paper, we use a recent example from Alberta's restructured wholesale electricity market, where firms compete by submitting bids in the form of price-quantity pairs repeatedly (every hour) in a centralized market. Like many wholesale electricity markets, Alberta's market is concentrated and has been shown to be susceptible to market power execution.[3] In this market, firms can adjust their bids up to two hours before the market clears, allowing them to respond quickly to the changing environment. Until a change in regulation in 2017, a de-identified list of all price-quantity bids used to be released publicly in near real-time, in a report called

---

[1] *See* Kai-Uwe Kühn & Xavier Vives, *Information Exchanges Among Firms and their Impact on Competition*, EUR. COMM'N WORKING PAPER 1 (1994); Xavier Vives, *Strategic Supply Function Competition with Private Information* 79(6) ECONOMETRICA 1919, 1919-1966 (2011); Pär Holmberg & Frank Wolak, *Comparing Auction Designs Where Suppliers Have Uncertain Costs and Uncertain Pivotal Status* 49(4) RAND J. ECON. 995, 995-1027 (2018).

[2] For a detailed discussion in the electricity markets context, *see* Nils-Henrik M. von der Fehr, *Transparency in Electricity Markets* 2(2) ECON. ENERGY & ENV'T POL'Y 87, 87-106 (2013).

[3] David P. Brown et. al., *Evaluating the Impact of Divestitures on Competition: Evidence from Alberta's Wholesale Electricity Market*, 89 INT'L J. INDUS. ORG. 102953, 102957-102961 (2023).

the Historical Trading Report (HTR). This provided a high degree of information and transparency.

In August 2013, Alberta's Market Surveillance Administrator (MSA) issued a report alleging that firms were using the information revealed in the HTR to coordinate on higher prices.[4] The MSA raised concerns that firms were "tagging" their bids on certain days to reveal their identities to rival firms. It was argued that these bidding patterns allowed firms to coordinate on high-priced outcomes. These concerns led to a hearing and the eventual end of the publication of the HTR in an order by the Alberta Utilities Commission.[5]

Several studies have documented the potential implications of the HTR on market outcomes. Olmstead, Ayres, and Lomas[6] and Brown and Eckert analyze the impact of the information revealed in the HTR and find that it led to higher market prices.[7] In particular, Brown and Eckert present evidence that firms were earning profits in excess of the amount that would arise from unilateral profit maximization suggesting that firms were using the HTR to coordinate.[8] Most relevant to our analysis is the work by Brown, Eckert, and Lin.[9] In this paper, the authors identify certain bidding patterns used by particular firms and present evidence that firms would have been able to identify rivals from the information revealed in the HTR. Further, the authors present evidence that firms were adjusting their bidding behavior when unique tagging patterns were employed, suggesting that firms were responding to the information in the HTR. However, a limitation of this study was the use of simple statistics and visual inspection to detect patterns in firms' bidding behavior. It is likely that firms wishing to reveal their identities or communicate with rivals could use more sophisticated techniques that visual inspections may not readily identify. More broadly, it is important to develop algorithmic screens that do not rely on visual detection and time-consuming monitoring.

---

[4] MKT. SURVEILLANCE ADMINISTRATOR, COORDINATED EFFECTS AND THE HISTORICAL TRADING REPORT: DECISION AND RECOMMENDATION 8, 8-15 (2013).

[5] ALBERTA UTIL. COMM'N, APPLICATION BY THE MARKET SURVEILLANCE ADMINISTRATOR REGARDING THE PUBLICATION OF THE HISTORICAL TRADING REPORT, Decision Proceeding 21115-D01-2017 (2017).

[6] Derek E. H. Olmstead et al., *Offer Price Information and the Exercise of Market Power: The Effect of the Publication of the Historical Trading Report on Competition in the Alberta Electricity Market*, 42 THE ENERGY J. 152, 152-161 (2020).

[7] David P. Brown & Andrew Eckert, *Pricing Patterns in Wholesale Electricity Markets: Unilateral Market Power or Coordinated Behavior?* 70(1) J. INDUS. ECON. 168, 190-204 (2022).

[8] *Id.*

[9] David P. Brown et al., *Information and Transparency in Wholesale Electricity Markets: Evidence from Alberta* 54(3) J. REGUL. ECON. 54 (3) 292, 292-330 (2018).

Machine learning techniques seem ideal for this purpose.[10] Machine learning approaches are designed to use large quantities of data to predict outcomes and detect complex patterns that may not be easily detectable by eyeball/manual inspections. Typically, machine learning models are trained on large amounts of data and use statistical techniques to identify patterns, trends, and relationships in the data. These models can then make predictions or decisions based on patterns they learned in the past. Additionally, machine learning algorithms are designed to adapt to changing data and to improve their performance (continually) over time. This makes them particularly useful for applications where the data changes or evolves – such as when agents use publicly available information to communicate or signal any particular behavior.

In this paper, we employ machine learning techniques to examine the MSA's claim that patterns in bidding behavior would have allowed firms to accurately identify the identities of the rivals associated with anonymized bids in the HTR. In particular, we consider the problem faced by an individual firm that aims to use the de-identified price-quantity offers in the HTR to predict the identities of specific rivals. Firms can observe the offers with firm identifications at a 60-day lag, allowing them to verify and update their prediction algorithm. Hence, our exercise is one of supervised multinomial classification, in which a period of time ending 60 days before the current month is the training data sample, and the current month is the test sample. We employ both decision tree and random forest algorithms.[11] While the random forest generally yields higher accuracy scores, the decision tree offers greater interpretability, allowing us to trace decision rules, and making them well-suited for illustrative purposes. The variables used to predict the identity of the firm behind a particular price-quantity offer, which is the output of the model, include different functions of the price and quantity (such as the integer and decimal portions of the price), as well as measures of the position of the price-quantity offer within the distribution. We employ data on all price-quantity offers from December 2012 to December 2013.

In general, we find that, before the MSA's report in August 2013, the identities of a firm's rivals associated with particular offers could be predicted with an overall average accuracy of 86%. The most important variables for prediction include variables whose values have economic justification (such as the MW size of an offer block), but also variables whose economic justification in the absence of

[10] David P. Brown et al., *Screening for Collusion in Wholesale Electricity Markets: A Literature Review* UTIL. POL'Y (forthcoming) (manuscript at 1-13).

[11] The motivation to use decision-tree-based algorithms is mainly due to their interpretability, versatility, and ability to capture complex, non-linear relationships in the data.

communication is unclear (such as the decimal portion of the offer price). The rules used by the algorithms reflect patterns that have been previously observed in the data through visual inspection but also capture other regularities. Importantly, we find that after the 2013 MSA report raising concerns over the use of offer patterns, the ability of the algorithms to identify firms from their offers accurately falls dramatically—with an accuracy rate lower than 50% in some cases. The reduced accuracy coincides with a distinct change in offer behavior where firms adjusted their bid strategies that appear to have been much more dispersed and randomized. It is worth emphasizing that the machine learning algorithm eventually learned how to distinguish the firms after the policy changes occurred in August 2013—when firms started to adopt more "randomized" offer price patterns. This tells us that machine learning can easily and quickly adjust to changing data environments.

Our findings show that the usage of machine learning algorithms to evaluate if firms can use publicly available information to identify rivals is promising. Our analysis has important regulatory policy implications because it demonstrates that machine learning methods may allow firms to recognize bidding patterns with high accuracy. As a result, our analysis suggests that regulators should enhance the monitoring of firms' behavior or reduce the granularity of information provided.

While our analysis focuses on electricity markets, our methods have applications in other settings. Pricing patterns that may be associated with coordination have been documented in a number of industries. Borenstein documented communication via fare codes by airlines.[12] Christie and Schultz examine the use of odd-eight quotes in Nasdaq stocks.[13] Abrantes-Metz, Villas-Boas, and Judge analyze the distribution of the second digit of the Libor rate and argue that the unique distribution could be associated with rate manipulation or collusion.[14] Lewis analyzes price endings in U.S. gasoline markets and finds higher and more rigid prices in locations that end with 5 and 9, suggesting that these endings may be used to establish focal prices. Our empirical methodology is particularly well suited for environments where there are numerous informational channels through which firms may be communicating.[15]

---

[12] Severin Borenstein, *Rapid Communication and Price Fixing: The Airline Tariff Publishing Company Case*, *in* THE ANTITRUST REVOLUTION: THE ROLE OF ECONOMICS 1-16 (John Kwoka & Lawrence White eds., 1998).
[13] William G. Christie & Paul H. Schultz, *The Initiation and Withdrawal of Odd-Eighth Quotes Among Nasdaq Stocks: An Empirical Analysis* 52(3) J. FIN. ECON. 409, 409-442 (1999).
[14] Rosa M Abrantes-Metz et al., *Tracking the Libor Rate* 18(10) APPLIED ECON. LETTERS 893, 893-899 (2011).
[15] Matthew S. Lewis, *Odd Prices at Retail Gasoline Stations: Focal Point Pricing and Tacit Collusion* 24(3) J. ECON. & MGMT. STRATEGY 644, 644-685 (2015).

This paper proceeds as follows. Section II provides background information on Alberta's electricity market and the Historical Trading Report. Section III describes our data. The empirical methodology is summarized in Section IV. Section V presents our results. Section VI concludes.

II. Background

A. Alberta's Wholesale Electricity Market

Alberta's wholesale electricity market operates as a single hourly uniform-priced procurement auction. In each hour, firms submit up to seven price-quantity offer blocks for each generation asset. The prices are restricted to be between $0/MWh and $999.99/MWh and reflect the price at which the generator is willing to make their specified output available. Throughout the hour, the Alberta Electric System Operator (AESO) that coordinates the market calls upon supply in order of the least cost until there is sufficient supply to meet demand. The price of the last unit called upon sets the System Marginal Price (SMP). Generation units that supply output in the hour are compensated according to the time-weighted SMP, referred to as the Pool Price.

Alberta's wholesale electricity market is an "energy-only" market design. Firms do not receive supplementary payments for constructing and maintaining generation capacity. Rather, firms must recover all of their costs of operating (both variable and fixed) by the payments they receive from producing electricity. Consequently, the MSA has indicated that firms are permitted to exercise unilateral market power via economic withholding that arises when firms bid units in excess of marginal cost to ensure they are not called upon, thus increasing the market price that they receive for the energy they continue to generate from their other generation assets.[16] Market power execution via economic withholding has been well-documented in Alberta's wholesale market.[17]

In 2013, the sample period of this study, Alberta's market was moderately concentrated, with five large generators having offer control over approximately

---

[16] This differs from many jurisdictions that provide supplementary payments for capacity to recover fixed costs. *See* James Bushnell et al., *Capacity Markets at a Crossroads* 278 Energy Inst. Hass Working Paper 3, 3-18. These market designs are often coupled with regulatory rules that mitigate generators' abilities to bid in excess of marginal cost. Christoph Graf et al., *Market Power Mitigation Mechanisms for Wholesale Electricity Markets: Status Quo and Challenges* STANF. U. WORKING PAPER 41, 41-49

[17] David P. Brown & Derek E.H. Olmstead, *Measuring Market Power and the Efficiency of Alberta's Restructured Electricity Market: An Energy-Only Market Design* 50(3) CANADIAN J. ECON. 838; David Brown et al. *supra* note 3.

65% of the market's capacity.[18] The remainder of market capacity is supplied by over 25 small firms with limited generation capacity. TransCanada had the largest offer control at 18.1%, followed by ENMAX with 13.4%, TransAlta with 12.8%, Capital Power with 10.4%, and ATCO with 9.8%.[19] In 2013, Alberta's generation output was primarily supplied by coal and natural gas; coal and natural gas provided 52% and 38% of the electricity generated, respectively. The remainder was supplied by a mix of hydro, wind, and biomass.[20]

B. Historical trading report

In the wholesale market, firms submit their initial offers the day prior to market clearing. Firms can make an unlimited number of adjustments to their price-quantity offers up to 2 hours prior to market clearing. The purpose of this rule is to allow firms to respond to changes in the market environment (*e.g.*, due to uncertainty in market demand, wind supply, and generator outages).

Until a change in 2017, generators could observe the anonymized prices and quantities for all offer blocks submitted to the wholesale market via the Historical Trading Report (HTR). The HTR was released approximately 10 minutes after the end of each hour. Consequently, this gives firms a window in which they can use the information released in the HTR to adjust their bidding behavior in future hours. More specifically, given that firms cannot adjust their bids beyond the 2 hours before market clearing and the HTR is published 10 minutes after the hour, firms can adjust their offer behavior for hour t+3 using the information revealed in the HTR 10 minutes after hour t.[21]

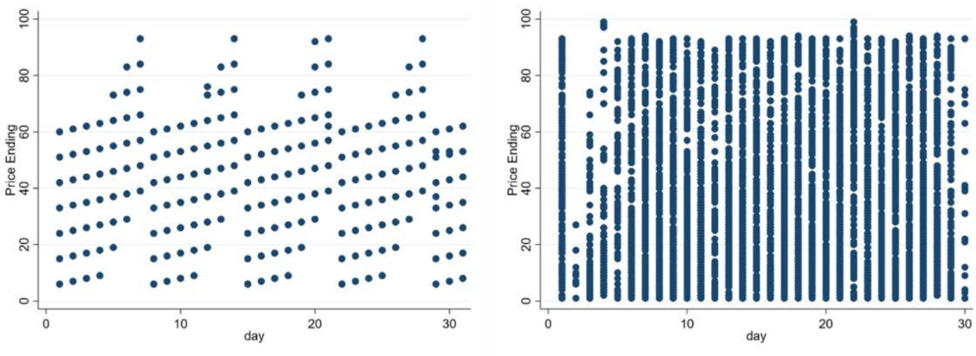
Supporters of the HTR argued that the information revealed in near real-time provided generators with important information about market conditions. However, in August 2013, Alberta's Market Surveillance Administrator (MSA) released a report alleging that certain firms were using the HTR to facilitate coordination. In particular, the MSA suggested that certain firms were "tagging" their price-quantity offers through the use of certain patterns in order to reveal the identity of the firm associated with the offers, and to communicate their intent to maintain high prices. The MSA supported these concerns by documenting a handful

[18] As part of industry restructuring in the late 1990s, existing generation capacity owned by three large utilities were virtually divested through long-term contracts giving the contract buyer offer control of these generating units (the ability to offer the output of these units into the wholesale market). These contracts had all expired by January 2021. For additional details, *see* Brown et al., *supra* note 3.
[19] MKT. SURVEILLANCE ADMINISTRATOR, MARKET SHARE OFFER CONTROL 1, 1-11 (2013).
[20] ALBERTA UTIL. COMM'N, ANNUAL ELECTRICITY DATA COLLECTION (2021).
[21] Brown et al., *supra* note 9, at 292–330.

of days where unique bid patterns were employed to effectively create a high-priced shelf of similarly priced offers that were “tagged”. In subsequent hours, firms would raise their offers up near the shelf resulting in a large quantity being offered at these high prices, which then often set the market-clearing price.

The MSA’s report led to a subsequent hearing by the Alberta Utilities Commission in December 2015, where the MSA recommended the HTR be replaced by a report with less granular information on individual price-quantity offers. In 2017, the AUC concluded that the HTR has the possibility to enhance concerns over market power execution and could possibly serve as an avenue for firms to communicate to coordinate to achieve higher prices.[22] The publication of the HTR ended on May 23, 2017.

The MSA’s 2013 report did not document the patterns that they alleged would allow firms to reveal their identities and communicate with rivals. Brown, Eckert, and Lin identify several different patterns via visual inspection in the bidding data.[23] In addition to a tendency for certain firms to employ specific quantity block sizes (which may simply reflect the sizes of different generation assets), patterns in the prices offered were also identified.[24] One of the patterns that was identified involved the behavior of the firm TransCanada. This pattern involved the price endings (*i.e.*, the digits after the decimal). On the first day of the month, TransCanada’s non-zero offer prices would end in 0.06, or in 0.06 plus a multiple of 0.09. On the second day of the month, prices would end in 0.07, plus multiples of 0.9. The starting price endings would increase to 0.8, 0.9, 0.19, 0.29, and 0.39 over the fourth to seventh days of the month, before dropping down to 0.06 again on the eighth day.



A. TransCanada, January 2013 B. TransCanada, September 2013

Figure I: Scatter Plots—TransCanada

As an illustration, Figure I.A plots all non-zero price endings for TransCanada, by day, for January 2013. Brown, Eckert, and Lin report that 86% of all of TransCanada’s

[22] Alberta Util. Comm’n, *supra* note 5.
[23] Brown et al., *supra* note 9, at 301-306.
[24] Brown & Eckert, *supra* note 7, at 191-204.

offer prices over $100/MWh from January 2011 to June 2013 conformed to this pattern.[25] Notably, this pattern in price endings disappeared following the release of the MSA's report in August 2013 detailing its concerns. As an example, Figure I.B shows TransCanada price endings for September 2013; following the MSA report, a high degree of randomness is observed in TransCanada's price endings.

Different types of patterns have also been identified in the bids of other firms. For example, Brown, Eckert, and Lin document the use by Capital Power of a sequence of prices with price endings of zero that are separated by exactly $1.[26] This pattern creates a "shelf" in the offer curve. As an example, Figure II.B shows all non-zero price-quantities offered by Capital Power for the hour from 7:00-8:00 AM on March 4, 2013. As illustrated in the market-level supply curve plotted in Figure II.A, the series of Capital Power offers with prices ranging from $974 to $980 created a large vertical portion in the supply curve.

Price	Block Size
7.75	145
7.76	132
10.75	9
10.76	9
740.00	10
741.00	10
974.00	43
975.00	70
976.00	27
977.00	70
978.00	27
979.00	123
980.00	207

Figure II.A: Capital Power Offers

[25] Brown et al., *supra* note 9, at 303-304.
[26] *Id.* at 304-306.

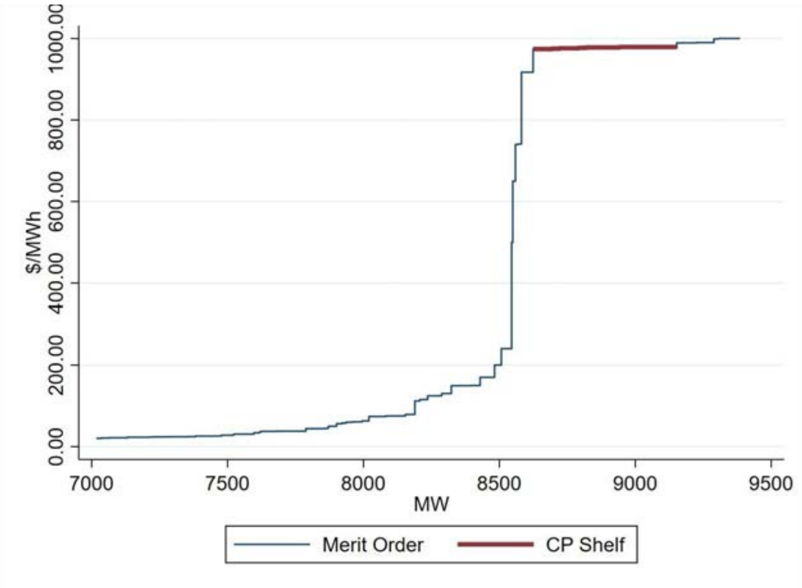


Figure II.B: March 4, 2013 offer curve

The bidding patterns reported in Brown, Eckert, and Lin were identified through visual inspection and simple statistics, including the frequencies with which the firms employ certain prices, price endings, or block sizes, and the first differences between prices in a firm's offer stack.[27] More complicated bidding patterns are likely to go undetected using these approaches; as a result, more sophisticated statistical techniques that utilize the large dimensionality of the data to evaluate how well firms can identify their rivals are desired. More broadly, this setting provides us with an opportunity to illustrate the potential ability of machine learning to evaluate if firms were able to identify their rivals with a high degree of accuracy.

III. Data

Our analysis considers the period December 3, 2012 – December 31, 2013, and relies on publicly available hourly data from the Alberta Electric System Operator's Merit Order Snapshot dataset. These data include the observed (final) price and quantity bids for all generation units, the generation company that had offer control of each unit, import supply from neighboring jurisdictions, and market demand. It is important to note that the data are released with a 60-day lag and include the identity (or *label*) of the firms associated with each asset, information that would not be available in the HTR released immediately after each hour.

In our analysis, we ask whether firms could identify the rival firms associated with each price-quantity bid in the merit order data for a particular hour by using

[27] Brown et al., *supra* note 9, at 301-304.

features of the prices and quantities in the different bids.[28] Because the focus of the MSA's 2013 report was on the bidding behavior of three specific large firms (ATCO, Capital Power, and TransCanada), we label each offer as coming from ATCO, Capital Power, TransCanada, or 'Other;'[29] we then consider the ability of each of the three firms to identify the label associated with each rival offer correctly.

Our sample period was chosen for two key reasons. First, our data covers part of the period in which the MSA alleged that firms were using the information revealed in real-time to communicate with other generators. Second, prior to December 3, 2012, the AESO's data does not include the offer control of every price-quantity block. This limits our ability to evaluate *ex-post* if the de-identified data released in the HTR could have been used to identify the firm submitting the bid correctly.

A. Input variables

The purpose of our empirical exercise is to determine, using machine learning techniques, the accuracy with which firms could identify the rival firms associated with different price-quantity offer blocks observed in the HTR. This requires presenting the machine-learning algorithms with the information contained in the HTR in a sufficiently general way to allow the detection of patterns that may be present. To do this, we construct variables based on the prices and quantities of each offer in an hour, and the relationships of those offers to others in the same hour. Table A.1 in the appendix provides detailed definitions of each of the input variables.

We include basic properties of the price-quantity offers submitted by firms, including the quantity offered (*q block size*), the integer portion of the price (*price int*), and the decimal portion of the offer price (*price decimal*). We break up the offer price into both the integer component and the decimal to capture the fact that firms may

[28] In addition to reporting the final price and quantity associated with each offer for a particular hour, the HTR reported the initial price-quantity submission for that offer block for that hour, submitted the previous day. Unfortunately, these data are not included in the Merit Order Snapshot dataset and are no longer available historically. It is possible that firms may have been able to send messages through their final offers, but also through the initial offers, or changes between initial and final offers. Hence, the ability of firms to communicate through the HTR may have been stronger than our analysis will indicate.

[29] As Table A.2 in Appendix A reveals, we have the following frequency for each label in our database: 11.40% for ATCO (99,277 obs); 13.40% for Capital Power (116,465 obs); 15.9% for TransCanada (138,340 obs) and 59.40% for the firms grouped and labeled as Other (518,144 obs). There are two large firms in the Other category, ENMAX and TransAlta. The remaining bids are made by more than 25 small firms. ENMAX is vertically integrated in the retail market, while TransAlta's capacity was primarily cogeneration which is must-run and hydro generation that was subject to long-term contracts during our sample period. Both aspects led these firms to bid competitively. For more details on Alberta's market structure, *see* Brown et al., *supra* note 3, at 102957-102961.

be using both elements to signal to their rivals, as was alleged in reference to the price decimals. It is possible that firms may be placing their offers in the "stack" of other offers to communicate with their rivals. For example, firms may be bidding a price slightly above their rivals to signal that they intend to coordinate on high-priced outcomes and not subsequently undercut their rivals' high-priced offers. To capture the relationship of an offer to other nearby offers, we sort all offers in ascending order of price, and construct the difference between the price, price ending, and quantity of a block, and those of the next lowest and highest blocks (*Dprice up*, *Dprice down*, *Dpdec up*, *Dpdec down*, *Dquant up* and *Dquant down*). For example, for the middle offer of the sequence of price-quantity offers (200.10, 45), (230.25, 60), and (240.95, 55), *Dprice up* = 10.70, *Dprice down* = 30.25, *Dquant down* = 15, *Dquant up* = —5, *Dpdec down* = 15, and *Dpdec up* = —5.

As the portion of the price ending following the decimal is less likely to be driven by economic considerations, there may be more scope for signaling through price endings. Because of this, we construct additional variables based on the decimal price ending of an offer. In particular, we sort the decimal price endings used in each hour, and compute the *rank* of the price ending from highest to lowest (*rank*), the frequency of offers in the hour employing that price ending (*dec freq*), and the distance between that price ending and the next lowest price ending (*decimal diff*).[30] Table A.2 presents summary statistics of the input variables for each of the firms.

IV. Empirical Methodology

Advances in data analytics naturally pave the way for regulators to integrate new (and easy-to-implement) detection tools and methodologies, such as machine learning, into economic analysis to improve enforcement and policy.[31] Algorithmic screening may overcome the drawbacks related to traditional inspection methods, such as eyeball/manual examination, structural changes, and causal inference analysis. Thus, machine learning is a promising way to identify communication strategies firms may employ through public data.

The intuition of our classification exercise is as follows. Consider a large firm that observes the offers of rival firms without identifiers in the HTR. The machine learning algorithm will be presented with information on an offer, such as its price,

[30] The machine learning algorithm can handle a large number of additional variables. As a robustness check, the set of variables was extended to include similar additional statistics based on block size and price integer. These additional variables led to marginal improvements in the predictions, but the overall qualitative conclusions are robust.
[31] Susan Athey & Guido W. Imbens, *Machine Learning Methods that Economists Should Know About* 11 ANN. REV. ECON. 685, 686-689.

quantity, and price ending, and must classify the offer as coming from one of the firm's rivals. We use explanatory variables (attributes) based on firms' price-quantity offers. The data we use to train the model has labels to identify the firms attached to these bids. We use decision-tree-based classifier algorithms (Decision Tree and Random Forest) based on these attributes to predict the identity of the firms (outputs) associated with a particular price-quantity offer. We then use these algorithms on the de-identified HTR data to evaluate how well we can correctly label the identities of rival firms.

In Section IV.A, we introduce the supervised machine learning model we use to predict the identities of the firms and to inspect how firms may have used the available public information to coordinate their market behavior. Sections IV.B and IV.C explain how we train/test and evaluate the machine-learning algorithms, respectively. Section IV.D presents the Permutation Importance (PI) technique we use to assess the relationship between the explanatory variables and the target variable (the identities of the firms). This step identifies the input variables that are most useful in predicting the identity of the firms associated with individual offers.

A. Decision Tree and Random Forest

A decision tree classifier uses a tree-like structure to classify (label) input data. The general approach to constructing a decision tree involves recursively partitioning the attributes into regions, with each region corresponding to a node in the tree. The goal is to create decision rules that accurately classify new input data instances.[32] Figure III illustrates how the decision tree algorithm classifies the data using the following simplified example with only four attributes.

The decision rules are guided by nodes and branches. Each node represents a decision based on an attribute of the data, and each branch represents the outcome of that decision. Therefore, nodes play a crucial role in the classification process by allowing the model to make decisions about how to classify the input data based on the information contained in each attribute. At the top of the tree, we have the root node, *i.e.,* the first decision made. The decision tree algorithm calculates the "impurity" of the tree nodes to select the best attribute to split the data and separate the classes.[33] Essentially, this procedure seeks to find the input variable (attribute) that results in the most "homogeneous" child nodes, where each child node contains

---

[32] GARETH JAMES ET AL., AN INTRODUCTION TO STATISTICAL LEARNING 327-361 (2013).

[33] Impurity is a measure of how mixed the classes are at a given node. Two common ones are Gini impurity and entropy. *See id.*, at 312-321.

mostly one class. By doing so, it can create a tree that effectively separates the classes in the data. From the root node, the tree branches out into multiple paths. These paths represent distinct possible outcomes of the decision. Each subsequent node tests different attributes of the data, and the tree continues to branch out until a final decision is reached at the bottom of the tree. This final decision is called a leaf node and represents the classification or prediction made by the model.

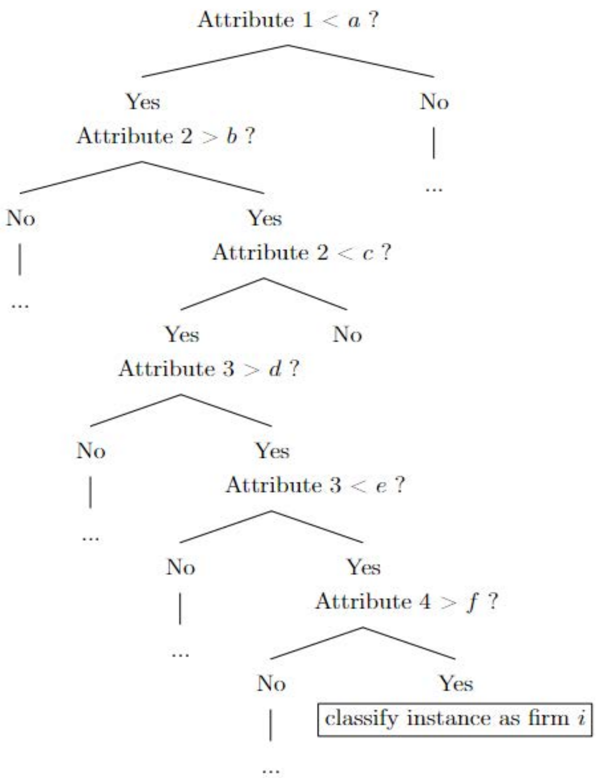


Figure III: Decision Tree with Four Attributes Example

Random forest is an ensemble learning method largely used for classification problems in machine learning. It combines the outputs of multiple decision trees to improve the accuracy of the classification model. The basic idea is to build a large number of decision trees, each on a randomly sampled subset of the training data and using a random subset of the attributes. Each decision tree is trained independently, and its prediction is combined with those of the other trees in a majority vote.[34] Random forests have a greater ability than decision trees to learn the actual structure of the data. Unlike decision trees, random forests are less likely to overfit the data, as they use only a portion of it and a subset of the original input variables. This allows

[34] In majority voting, we first group the outputs (*i.e.*, the labels predicted) by the individual trees. Then, the most frequent output is voted as the final predicted label.

them to learn only the essential aspects of the data. Another advantage of random forest algorithms is their robustness to noise and outliers in the data.[35]

B. Training and Test Samples

When training a machine learning model, it is common to use more data for training than for testing, as we are ultimately interested in how well our model can generalize to new – and unseen – data. The data used for testing serves as a proxy for this new data. While there is no fixed rule for how much data to use for training and testing, a common industry practice is to use a split of 75% for training and 25% for testing, or 70% for training and 30% for testing. Ultimately, the split choice should depend on the specific application and the available data.[36] In our application, we use a three-month training sample for each one-month test sample. Our training set contains approximately 75% of the total observations, and our test set computes out-of-sample predictions of approximately 25% of the sample. The classification task is carried out on a rolling window scheme because the dataset with the identity (or label) of the firms is released with a 60-day lag.[37] In other words, to predict the identity of firms in May 2013 (test sample), we used data for December 2012, January 2013, and February 2013 as the training sample. We apply this same procedure to predict the identity of firms in subsequent months (June 2013 to December 2013).

It is worth mentioning that while we use the identity of the firms to train our supervised machine learning algorithm based on the AESO's Merit Order data, we remove their identities in the test set to evaluate the predictive performance. Consequently, the (test) data used in our analysis provide the de-identified final bids that would have been published immediately after the hour. These data provide us with a rich set of data that can be used to evaluate if firms were able to identify their rivals using this information with a high degree of accuracy. We use the cross-validation approach to compare and select the best machine-learning model based on its ability to predict new data.[38]

C. Evaluation Metrics

---

[35] Trevor Hastie et al., *Boosting and Additive Trees*, *in* THE ELEMENTS OF STATISTICAL LEARNING: DATA MINING, INFERENCE, AND PREDICTION 337, 351 (Trevor Hastie, Robert Tibshirani, & Jerome Friedman eds., 2009); Leo Breiman et al., *Introduction to Tree Classification*, *in* CLASSIFICATION AND REGRESSION TREES 55-58 (1984).

[36] *Id.*; James et al., *supra* note 32.

[37] In machine learning classification tasks, a labeled dataset refers to a set of data instances having an associated label (or class) representing the true output or target for that input, these labels are also referred to as input examples.

[38] Cross-validation may flag overfitting. Intuitively, overfitting describes the case in which the model performs well for the training dataset but generates poor predictions for new data.

As we propose a classification task using a labeled dataset, we evaluate our findings by comparing our out-of-sample predictions with the firms' actual identities by assessing the *Precision* and *f1-score* metrics. *Precision* quantifies the number of correct predictions. More precisely, *Precision* is the fraction of times that the algorithm is correct when it labels an offer as coming from a specific firm. Thus, high *Precision* is associated with a low incidence of Type I errors (false positives). As an example, a false positive in our analysis would reflect a situation where ATCO does label a firm as being "TransCanada" when it is not "TransCanada".[39]

To understand the *f1-score*, we first must introduce the *Recall* metric. A high *Recall* is associated with a low incidence of Type II errors (false negatives). For example, a false negative in our analysis would reflect a situation where ATCO does not label a firm as being "TransCanada" when it is "TransCanada". This reasoning is analogous to predictions related to "Capital Power" and "Others". To achieve maximal *Precision* (no false positives) and *Recall* (no false negatives), there must be an absence of type I and II errors, respectively. The *f1-score* provides a single score that balances (via the harmonic mean) the concerns of both *Precision* and *Recall* in the same measure.[40]

To analyze further the model's predictive power and how that power changed following the release of the MSA's 2013 report in August, we consider Receiver Operating Characteristics (ROC) and Area Under the Curve (AUC) metrics.[41] ROC is a widely used performance metric in machine learning classification tasks and measures the ability of a classifier to distinguish between positive and negative classes by plotting the True Positive Rate (TPR) against the False Positive Rate (FPR). The ROC curve relies on a probabilistic model that outputs the probability of an observation belonging to the positive class—popularly named as "class 1", which is the class we want to predict. So, the model outputs probability values (between 0 and 1). Thus, to build the ROC curve, we use these model outputs and establish a threshold, *i.e.*, the point at which a classifier outputs positive (class 1) or negative (class 0) predictions. The most common value for the threshold is 0.5, but to plot the ROC curve, we use the interval [0,1]. Therefore, to draw the ROC curve, we run the classification model on a test set of labeled data and use the predicted class probabilities to generate a set of predictions for different probability thresholds.

---

[39] In our applications, false positives are more costly or harmful than false negatives. For example, a false positive for ATCO (incorrectly identifying Capital Power or TransCanada) can lead to "wrong" communication/coordination strategies. Hence, ATCO would prioritize precision over recall to reduce false positives.

[40] *f1-score* = 2 x [(*Precision* x *Recall*) / (*Precision* + *Recall*)].

[41] Andrew P. Bradley, *The Use of the Area under the ROC Curve in the Evaluation of Machine Learning Algorithms*, 30 PATTERN RECOGNITION 1145 (1997).

In our analysis, positive classes represent the label of the specific firm we aim to identify. For example, if ATCO wants to predict TransCanada, then TransCanada represents the positive class—and Capital Power and Others the negative class. While TPR measures the classifier's ability to identify TransCanada correctly, FPR measures the classifier's tendency to misclassify TransCanada.[42] We can plot different points on the ROC curve by varying the threshold for classifying instances as positive or negative. The closer the curve is to the top-left corner of the plot, the better the classifier is at distinguishing between the two classes. The area under the ROC curve (AUC) measures the classifier's performance. An AUC of 1 indicates a perfect classifier and an AUC of 0.5 points to a random classifier. In summary, the ROC measure provides an intuitive way to evaluate the trade-off between TPR and FPR. This helps us to assess the overall performance of our classification model.

D. Permutation Importance (PI)

We employ Permutation Importance (PI) to determine which variables are most important for predicting the identities of rival firms. Permutation Importance is a technique used in machine learning that helps to identify the variables that contribute the most to the predictive power of a model.[43] The idea behind the technique is relatively simple: by randomly permuting (*i.e.*, shuffling) the values of a given variable in the dataset and observing how much the model's performance decreases, we can estimate the importance of that variable.[44] Intuitively, PI measures the decrease in a model accuracy score when a single feature value is randomly shuffled, and can be summarized as follows: (i) train a model on a dataset including all input variables to calculate its performance on the test sample; (ii) randomly shuffle the values of a single input variable in the test sample while all other variables remain unchanged to calculate the model's performance again on the shuffled dataset; (iii) compare the original performance metric to the shuffled performance metric to observe how much the model's performance decreased; and (iv) *rank* the variables by the decrease in the model's performance. More important variables will lead to larger decreases.

[42] Typically, ROC curves apply to binary classification tasks, where the True Positive and False Positive rates are unambiguous. In the case of our multiclass classification, we must binarize the outputs to provide a notion of both True Positive and False Positive. To reach that aim, we use the One-vs-Rest scheme to compare each class ("TransCanada" for instance) against all the others ("Capital Power" and "Others" assumed as one).

[43] André Altmann et al., *Permutation Importance: A Corrected Feature Importance Measure*, 26 BIOINFORMATICS 1340 (2010).

[44] Broadly, using the PI technique, we can standardize the base rule to compare the predictive importance of the explanatory variables of various algorithms - including those that do not have such attributes or use subtly different mechanisms for this aim.

Suppose we have a variable with a PI of 0.15. In practical terms, this variable leads to a 0.15 decrease in the model accuracy score when replaced by its shuffled version. We can also interpret the PI score in relation to other attributes within the same model—and use it as a tool for variable selection or understanding the relative importance of different input variables in the model's predictions. Thus, following these steps for each input variable, PI allows us to identify which input variables are the most important in making accurate prediction.

V. Results

In this section, we present the results from the usage of the decision-tree-based algorithms to predict the identities of rival firms associated with particular price-quantity offers. To illustrate and provide intuition for our main results, we first consider the perspective of a single firm (ATCO); specifically, we consider to what extent ATCO could correctly predict which rival offer blocks come from Capital Power, TransCanada, or Other, based on the most recent three months of training data for which identifying information is available. We categorize in the label "Other" all the remaining firms in the data set. In section V.A we first present the results for ATCO using a random-forest algorithm; we then discuss results using a simple decision tree algorithm in order to gain intuition on the rules ATCO could use to identify its rivals. Finally, Section V.B provides an overview of the results from the perspective of Capital Power and TransCanada.

A. ATCO

To better understand and evaluate the random forest classifier's performance, we first use the illustration of Figure 4 to assess the average Precision and f1-score – from May 2013 to December 2013—for each individual class. ATCO can identify Capital Power with an overall average Precision of 0.78 and 0.64 f1-score and recognizes TransCanada with an average Precision of 0.74 and 0.67 f1-score. It is worth noting, however, that these scores are distinctly higher between May 2013 and July 2013, especially for TransCanada, where we find an average Precision and f1-score given by 0.88 and 0.81, respectively. We observed a noticeable fall in the score metrics between August 2013 and October 2013, which led to a decrease in the average performance—after this period, it started to increase again between November 2013 and December 2013. Although it is possible to notice a downward trend for the firms labeled as "Other," it is not as striking as those we found for Capital Power and TransCanada.

Figures IV.A and IV.B show that following the MSA's August 2013 report alleging that firms were tagging offers, the ability of ATCO to identify the offers of TransCanada declined sharply, with Precision and f1 scores falling from above 0.8 in

July to below 0.5 and 0.4 in September. This result suggests that the dramatic change in TransCanada's price ending pattern in August may have had an important negative effect on the ability of rival firms to recognize TransCanada's offers. Notably, however, we see an increase in this ability by November, once TransCanada's new price-ending behavior has started appearing in the training data. Hence, our results suggest that even with the switch to random and noisy price endings, TransCanada's offers can be identified.

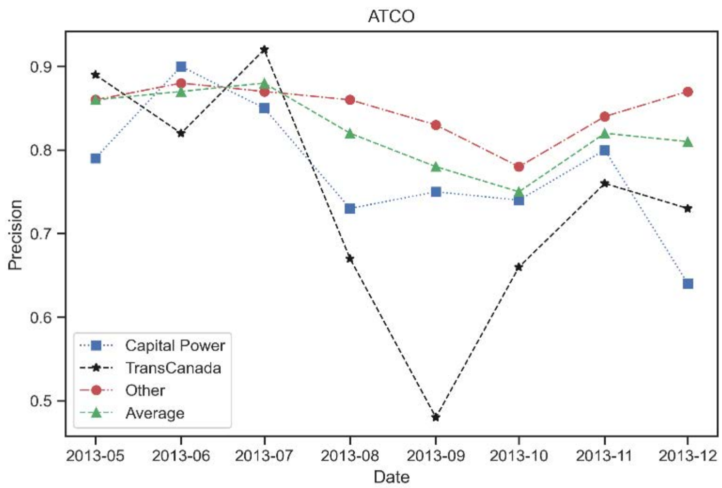


Figure IV.A: Precision

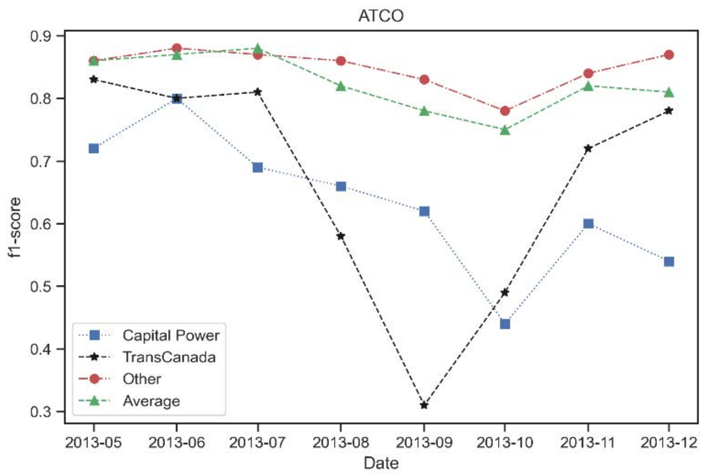


Figure IV.B: f1-score—ATCO

Figures V.A and V.B illustrate the rates of correct predictions (on the vertical axis) versus the fraction of errors on the horizontal axis. In the best scenario, the rates of correct predictions and errors would be null, respectively. Then, a perfect prediction would occupy the coordinate (0,1) in the upper-left corner of the graph. Furthermore, the ROC curve captures the true and false positive rates for different classification threshold probabilities. Poor quality predictions form a diagonal (dashed in black) line from coordinate (0,0) to coordinate (1,1). Along this dashed line, the algorithms predict all firms as Capital Power or TransCanada. Predictions lying below this dashed line have little or no ability to distinguish Capital Power from TransCanada.

From Figures VA and VB, we can see, in relative terms, a decrease in the ability of the random forest algorithm to identify Capital Power and TransCanada in the period ranging from June 2013 to September 2013. Generally, an AUC score of 0.8 indicates that the model has a high level of discriminatory power and is able to distinguish between positive and negative instances with reasonable accuracy.[45] Thus, the AUC value we achieve (0.82) still points to a good performance of the model and may give us some clues about the broader concerns of having a large set of information revealed in near real-time to get the following conclusion: as long as there are enough randomized-looking patterns, a machine learning algorithm could learn these patterns to predict identities relatively well in our setting.

Recall that, until August 2013, firms were signaling via price decimals, mainly TransCanada and Capital Power—and their pattern was duly captured by the attributes of the model, in particular the "*dec freq*" input variable, as shown in Panel (A) of Table 1. Even when firms stopped signaling via the (frequency of) price decimals, it was still possible to reach a satisfactory level for AUC in September 2013. Therefore, the main takeaway is that – at a relatively moderate cost in terms of forecasting error – our model manages to predict the identity of companies even when they do not send signals to rivals with well-defined (and frequent) price decimals patterns.

[45] DAVID W. HOSMER JR. ET AL., APPLIED LOGISTIC REGRESSION 173-182 (3rd ed. 2013).

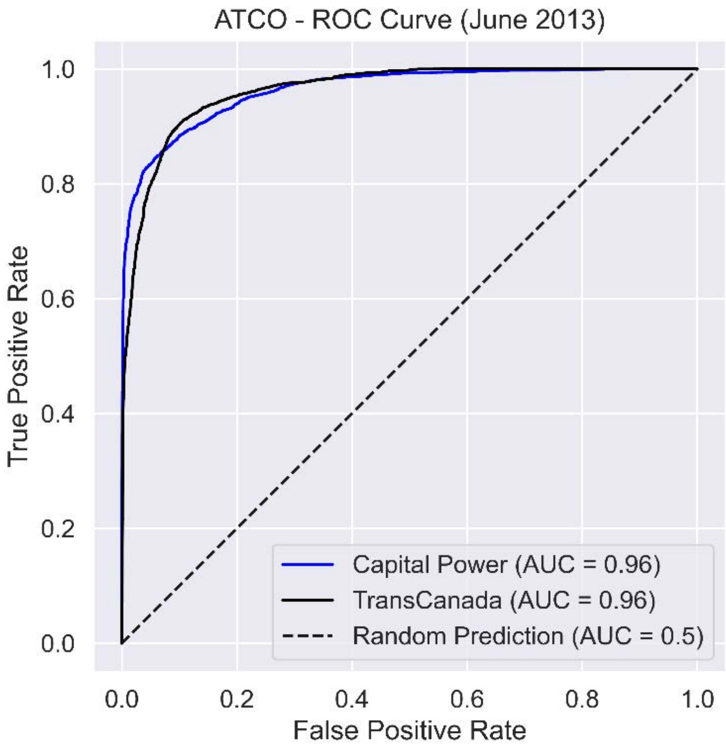


Figure V.A: ROC Curve—ATCO June 2013

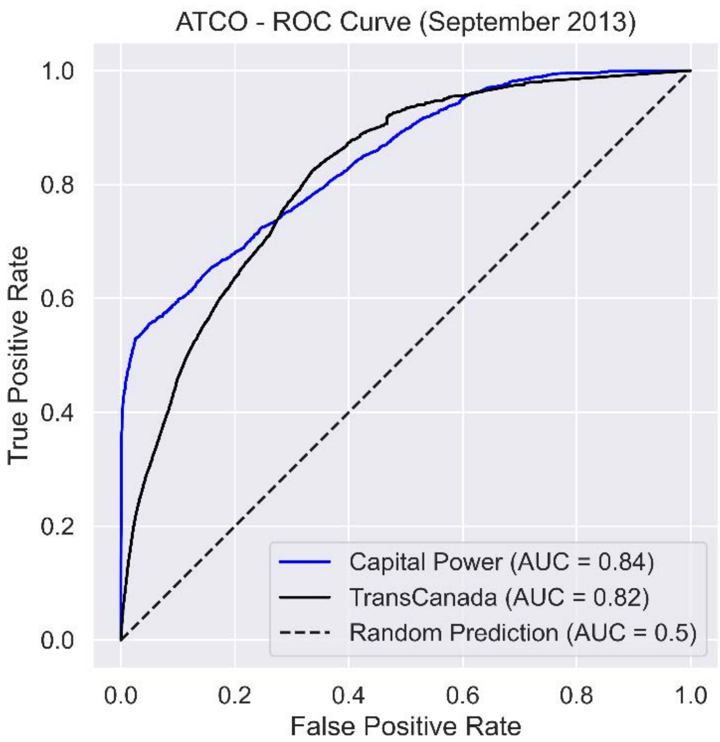


Figure V.B: ROC Curve—ATCO September 2013

Table I presents the results of the PI technique when applied to ATCO's classification exercise using the random forest algorithm. In the first column of each panel, we show the estimated increase in prediction error when we replace each input variable with its random shuffling counterpart. The drop in the model accuracy captures how much the identification of the firms depends on each input variable. In Panel (A), *q block size* has a Permutation Importance of 0.118. This tells us that the drop in the model accuracy when we replace *q block size* by its random shuffling counterpart is given by 0.118. In addition, we see that the "top-four" input variables to predict TransCanada and Capital Power identities (from May 2013 until July 2013) are *q block size*, *price int*, *price decimal*, and *dec freq*. This is not observed in Panel (B), where the input variable *dec freq* is the less relevant variable. In fact, this suggests that, from August 2013 until December 2013, *dec freq* decreases the ability of ATCO to predict TransCanada and Capital Power identities. We find similar outcomes in the cases where we evaluate Capital Power's and TransCanada's ability to predict their rivals' identities (see Tables B.1 and B.2, respectively).

Panel (A) Average (May 2013 -Jul 2013)			Panel (B) Average (Aug 2013 -Dec 2013)		
	Permutation Importance	Std. Dev.		Permutation Importance	Std. Dev.
q block size	0.118	0.001	*price int*	0.119	0.002
price int	0.106	0.002	*q block size*	0.095	0.002
price decimal	0.071	0.001	*price decimal*	0.060	0.002
dec freq	0.035	0.001	*Dprice down*	0.010	0.001
Dprice up	0.021	0.001	*Dprice up*	0.010	0.001
Dprice down	0.011	0.001	*rank*	0.003	0.001
rank	0.009	0.001	*Dquant down*	0.003	0.001
Dquant down	0.008	0.001	*Dpdec down*	0.003	0.001
Dpdec up	0.007	0.001	*Dquant up*	0.003	0.001
Dpdec down	0.006	0.001	*Dpdec up*	0.002	0.001
decimal diff	0.003	0.001	*decimal diff*	-0.001	0.0004
Dquant up	0.002	0.001	*dec freq*	-0.002	0.001

Table I: Permutation Importance—ATCO

A. 1. Paths and Thresholds to Identify TransCanada

Figure VI presents the decision tree model estimates to predict which blocks are TransCanada.[46] In this illustration, we are taking the decision rule developed on the training sample and applying it to the test sample (May 2013). For reference, there are a total of 51,344 blocks. Of these, 63% are Other, 15% are Capital Power, and 22% are TransCanada.

[46] It is important to emphasize that the results of the decision tree classifier presented here have a didactic purpose since its decision rule may be more easily interpreted. We performed cross-validation analysis to determine the decision tree's minimum split node size and maximum depth. In addition, we used the balanced class weight mode to adjust weights proportionally to class frequencies in the input data. We also summarize the results for this particular decision tree in Appendix B.2.

The first cut separates blocks based on whether the block size is less than 66.5 or greater than 66.5. There are 46,244 blocks with a size less than 66.5, and 5,100 blocks with a size greater than 66.5. TransCanada represents 25% of the blocks with a size less than 66.5, but only 2% of blocks greater than 66.5. Hence, as our focus here is to describe the path allowing us to predict TransCanada, we must follow the left side of the tree. Note that the right side of the tree would be interesting to follow the path that allows us to identify Capital Power – since the percentage of blocks larger than 66.5 provided by Capital Power increases to 35%.[47]

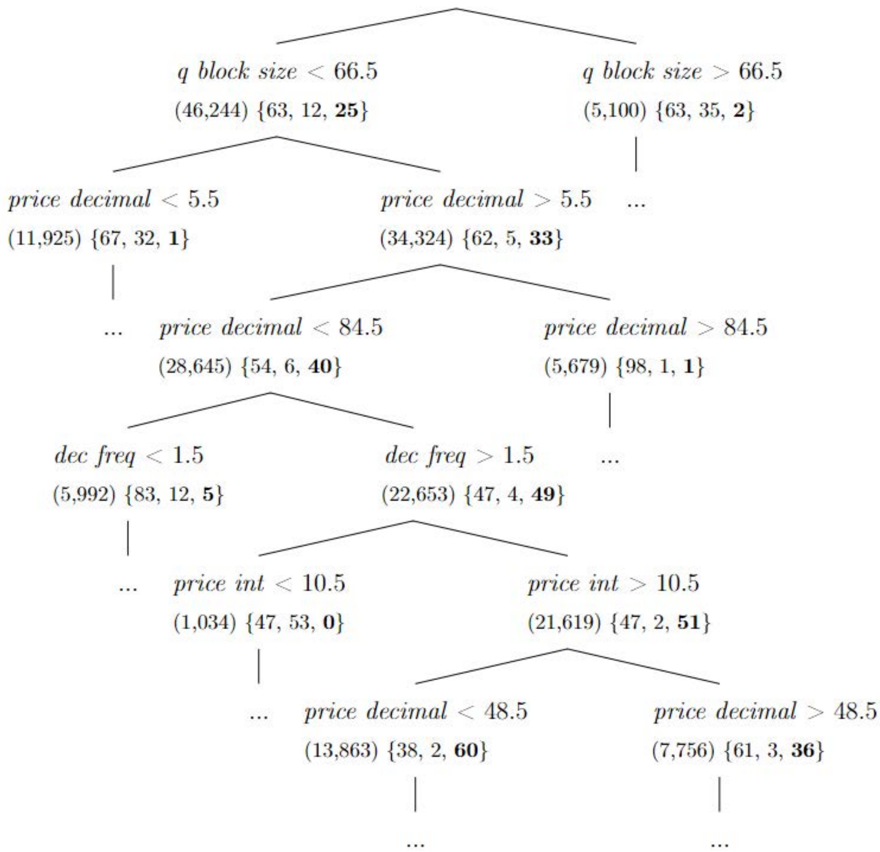


Figure VI: Decision Nodes Used by ATCO to Identify TransCanada
The parentheses contain the number of blocks at each node. The braces provide the percentage of observations that are Other, Capital Power, and TransCanada, respectively. TransCanada block percentages at each node are in bold within the braces.

Within the set of blocks smaller than 66.5, the algorithm's next cut is to divide blocks according to whether the price ending (*price decimal*) is less than or greater than 5.5 (note that a price of 20.05 has a price decimal ending of 5, for example). Of the

[47] It is worth mentioning that although most instances labeled as TransCanada are in the leaves on the left branches of the tree, the right branches can also contain leaves associated with the TransCanada pattern. Likewise, although the right side of the tree would make it possible to map Capital Power, the tree's left side may also allow us to identify it correctly.

11,925 blocks in this subset that have a price ending less than 5.5, only 1% are TransCanada. In contrast, 33% of the 34,324 blocks in this subset that have a price ending greater than 5.5 are TransCanada.

Essentially, the algorithm has decided that to find TransCanada blocks, it should focus on blocks smaller than 66.5 and have price endings greater than 5.5. The latter is interesting since we know from our previous (eyeball) examination of price endings that TransCanada's price endings are almost always 6 and higher.

Along this path, the next step is to separate the blocks with price endings less than 84.5 from those with endings greater than 84.5. Again, this fits neatly with our pattern; in May 2013, greater than 99% of all of TransCanada's blocks had price endings of 84 or smaller, and 84 fits within TransCanada's pattern. Of the blocks along this path that have price endings greater than 84.5, only 1% are TransCanada, so we will not follow that path further in this discussion.

Focusing on blocks that have less than 66.5 MW, and price endings from 6 to 84, the algorithm then looks at *dec freq*, which is the percentage of blocks in the hour that share the same price ending as the block in question. Note that within the entire sample, the only price ending that has large frequencies is 0. Looking at blocks less than 66.5 MW and with price endings from 6 to 85, *dec freq* ranges from 1 to 14 with a mean of 3.5. The algorithm divides the blocks in our current subset according to whether *dec freq* is less than or greater than 1.5. Of the 5,992 blocks in the subset that have *dec freq*<1.5, only 5% are TransCanada; in contrast, of the 22,653 blocks in the subset with *dec freq*>1.5, 49% are TransCanada. This seems to capture that on a given day, TransCanada uses a small number of price endings, so that the same ending is used on multiple blocks.

The tree we present here shows a couple of more steps. It separates the subsample by price integer (threshold 10.5) and price ending again (threshold 48.5). The latter is interesting since it captures a breakpoint in TransCanada's price-ending pattern – in that, it uses endings in the ranges [42-48] and [51-57], but the pattern never lands on price endings of 49 or 50. This sort of threshold shows up in other decision rules as well. Likewise, there are paths that identify blocks as TransCanada if they have price endings between 82.5 and 84.5.

B. Capital Power and TransCanada

Qualitatively, as illustrated by Figure B.1 in Appendix B, the results obtained for Capital Power and TransCanada are similar to those found for ATCO. In general, we observe that the forecasts made between May and July 2013 have substantially higher Precision and f1-score rates. There is a significant drop in the ability of firms to predict the identity of rivals between August and October 2013.[48]

Figure B.2 illustrates the AUC for Capital Power and TransCanada. As observed for ATCO, we notice a decrease in the AUC metrics when we compare June 2013 with September 2013. However, their values still point to a good performance of the model. This provides additional evidence that machine learning algorithms could learn patterns to predict identities relatively well, even after the policy change observed in August 2013.

Table B.1 summarizes the permutation importance outcome for Capital Power. In Panel (A), we see that the top-four input variables to predict TransCanada and ATCO identities (from May 2013 until July 2013) are *price int*, *q block size*, *price decimal*, and *dec freq*. In Panel (B) we observe the following top-four input variables: *price int*, *q block size*, *Dprice up*, *Dprice down* – and *dec freq* is the less relevant variable between August 2013 and December 2013. B.2 shows the outcomes of the permutation importance technique for TransCanada. The top-four variables in Panel (A) are *price int*, *q block size*, *price decimal*, and *Dprice up*. On the other hand, the top-four variables to identify Capital Power and ATCO are *price int*, *q block size*, *Dprice up*, and *price decimal*. Finally, it is worth noticing that when TransCanada wants to identify ATCO and Capital Power between August 2013 and December 2013, the inclusion of the variable *dec freq* does not necessarily decrease its accuracy. This outcome presented in Panel (B) is different (opposite) from when ATCO (Capital Power) is trying to predict Capital Power (ATCO) and TransCanada offer prices.

VI. Future Steps

In this paper, we illustrate the potential benefits of machine learning tools to competition policy enforcement by applying them to a case of possible coordinated behavior in Alberta's wholesale electricity market. In particular, in 2013, the Alberta Market Surveillance Administrator alleged that firms were employing patterns in their bids, allowing rivals to identify them in anonymized public data and potentially sending signals about bidding intentions. While previous studies of this case have used visual data inspection and simple summary statistics to search for bidding patterns, we employ a more rigorous approach by using machine learning algorithms

[48] Much of the downward trend in both the f1-score and precision metrics observed for TransCanada and Capital Power is driven by the drop in August 2013. Before the drop, the downward trend was not as apparent. For ATCO, the bids more randomly and less predictably in June and July 2013 may have driven this downward trend.

to examine whether firms could identify their rivals' offers in public data. In addition, we consider which features and characteristics of price-quantity offers are most important in revealing firms' identities.

We find, using random forest algorithms, that before the release of the MSA's report, the ability of firms to identify large rivals through public bidding data was high, but this ability declined when firms adjusted their bidding behavior after the report was released. We find that the bid characteristics that contribute most to identifying the firms associated with particular bids include features that likely have physical or simple economic justification (such as the quantity of the offer or the integer portion of the price), but also characteristics such as the decimal price ending and the frequency of price endings whose justification is less obvious in the absence of signaling and communication. Therefore, our results suggest that the recognizability of firms in public offers is at least partly the result of deliberate behavior.

Our analysis has important regulatory applications in settings where firms interact repeatedly and have access to high-frequency publicly available data. Our results highlight the potential use of emerging machine learning and data mining techniques to evaluate if firms can use de-identified data to identify rival behavior. In particular, these empirical methods could be used to develop *ex-ante* screens to be employed on an ongoing basis by regulators to monitor firm behavior.[49] These screens can identify which firms should be subject to further scrutiny. It is important to note that a key goal of such screens is to minimize the number of errors that either wrongly flag behavior as reflecting communication and/or collusion (a Type I error) or fail to identify such behavior (a Type II error). A key challenge facing regulators is to strike a balance between enforcement and the allocation of costly resources to monitor behavior. If regulators conclude that firms are using publicly available information to communication, they may need to change policies to reduce the amount of granular information that is made available to firms.

Our paper suggests several directions for future research. First, our analysis considers the case of Alberta's wholesale electricity market. Future research should consider alternative jurisdictions and industries to demonstrate the potential applicability of machine learning techniques in informing regulators about the appropriate level of information disclosure. Second, our random forest approach is well suited to determining whether firms were revealing their identities in public

[49] For additional details on developing screens for collusion in the electricity sector and the use of machine learning and data mining techniques, *see* Brown et al., *supra* note 10.

data as alleged, but less well suited to identifying the price patterns used for this purpose. The development of machine learning tools that monitoring or antitrust agencies could use to find suspicious patterns in pricing and bidding data is a crucial avenue for further research. A potentially fruitful approach may incorporate data mining algorithms used for pattern recognition; see for example Mooney and Roddick[50] and Fournier-Viger *et al.*[51] for recent surveys. Third, our analysis relies on standard decision trees and random forest algorithms. Future research could apply and compare different ML techniques, such as Gradient Boosting classifiers, which are ensemble methods that combine multiple weak learners (usually decision trees) to make accurate and robust predictions. Another possibility would be to use deep learning models such as Convolutional Neural Networks (CNNs) and Recurrent Neural Networks (RNNs), which have achieved excellent performance in multiclass classification tasks, and can learn more complex patterns and relationships from the data.

---

[50] Carl H. Mooney & John F. Roddick, *Sequential Pattern Mining–Approaches and Algorithms*, 45 COMPUTING SURVEYS, 1-19 (2013).

[51] Philippe Fournier-Viger et al., *A Survey of Sequential Pattern Mining*, 1 DATA SCIENCE AND PATTERN RECOGNITION, 54-77 (2017).

VII. Appendices

A. Input Variables and Summary Statistics

Input Variable	Definition
q block size	the quantity offered in a particular price-quantity offer
price int	integer portion of the offer price
price decimal	decimal portion of the offer price
dec freq	frequency with which a particular price decimal is used
rank	*rank* of the decimal ending from highest to lowest
decimal diff	distance of a price decimal to the next lowest distinct price decimal
Dpdec up	price decimal minus price decimal of the next highest price-quantity offer
Dpdec down	price decimal minus price decimal of the next lowest price-quantity offer
Dprice up	price integer minus price integer of the next highest price
Dprice down	price integer minus price integer of the next lowest price
Dquant up	quantity minus quantity of the next highest price-quantity offer
Dquant down	quantity minus quantity of the next lowest price-quantity offer

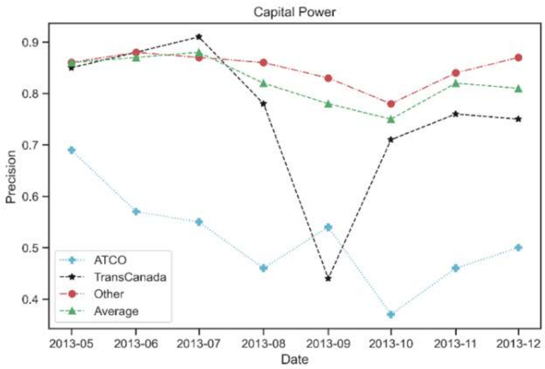
Table A.1: Variables Used to Recognize Patterns Firms May Have Used to Tag Themselves

	Obs	Mean	Std. Dev.	Min	Max		Obs	Mean	Std. Dev.	Min	Max
Panel (A) ATCO						Panel (B) Capital Power					
price int	99,277	271.83	353.69	17	999	*price int*	116,465	382.33	412.25	6	999
price decimal	99,277	37.23	35.20	0	99	*price decimal*	116,465	35.11	33.88	0	99
dec freq	99,277	7.68	9.37	1	46	*dec freq*	116,465	7.63	9.53	1	46
rank	99,277	17.76	16.00	1	66	*rank*	116,465	17.40	16.00	1	68
decimal diff	99,277	1.63	1.88	0	25	*decimal diff*	116,465	1.37	1.65	0	28
Dpdec up	99,277	-3.72	46.07	-99	99	*Dpdec up*	116,465	1.76	34.46	-99	99
Dpdec down	99,277	1.39	43.79	-99	99	*Dpdec down*	116,465	-1.89	34.94	-99	99
Dprice up	99,277	7.94	30.26	0	666.27	*Dprice up*	116,465	10.47	31.34	0	718
Dprice down	99,277	11.91	43.86	0	643.17	*Dprice down*	116,465	16.04	48.30	0	718
q block size	99,277	29.77	17.58	1	140	*q block size*	116,465	44.81	35.16	1	266
Dquant up	99,277	-2.73	23.06	-135	324	*Dquant up*	116,465	-0.14	35.43	-222	327
quant down	99,277	0.54	23.25	-331	119	*Dquant down*	116,465	-2.39	36.78	-307	231
Panel (C) Other						Panel (D) TransCanada					
price int	518,144	258.59	375.08	0	999	*price int*	138,340	252.43	368.72	2	999
price decimal	518,144	46.36	36.92	0	99	*price decimal*	138,340	37.69	23.52	0	99
dec freq	518,144	6.49	8.57	1	46	*dec freq*	138,340	3.32	3.37	1	43
rank	518,144	21.96	16.92	1	68	*rank*	138,340	18.28	10.93	1	61
decimal diff	518,144	1.67	1.78	0	28	*decimal diff*	138,340	2.62	2.12	0	18
Dpdec up	518,144	0.53	40.28	-99	99	*Dpdec up*	138,340	2.02	33.10	-99	99
Dpdec down	518,144	0.76	40.75	-99	99	*Dpdec down*	138,340	0.81	33.67	-99	97
Dprice up	518,144	12.53	39.30	0	703	*Dprice up*	138,340	6.35	22.76	0	675
Dprice down	518,144	10.13	31.01	0	646.35	*Dprice down*	138,340	7.89	29.01	0	651
q block size	518,144	33.43	34.03	1	350	*q block size*	138,340	39.13	18.79	1	177
Dquant up	518,144	1.01	42.96	-348	348	*Dquant up*	138,340	-6.29	25.19	-149	310
Dquant down	518,144	-1.61	39.15	-348	348	*Dquant down*	138,340	3.25	34.64	-333	168

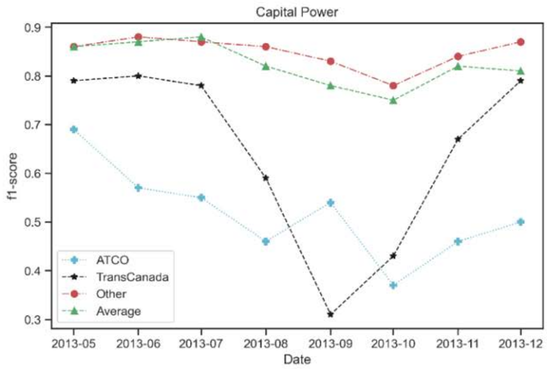
Table A.2: Summary Statistics of the Input Variables, December 3, 2012—December 31, 2013

B. Supplemental Results

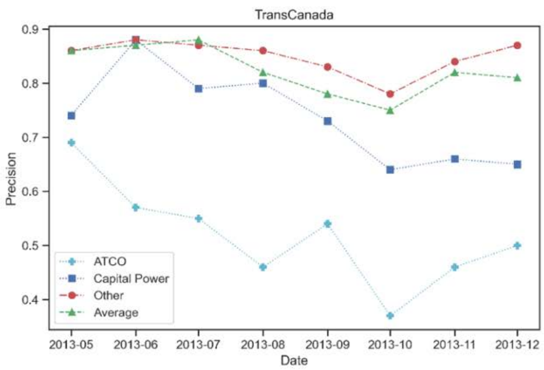
B.1. Random Forest



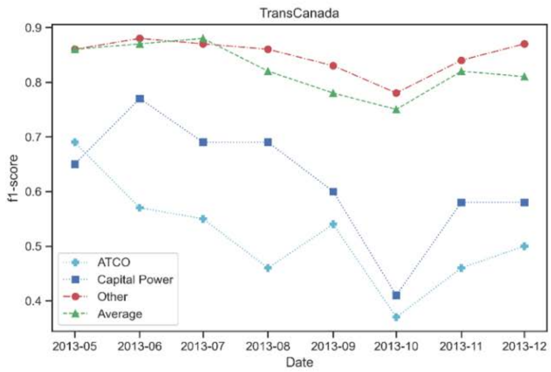
(a)



(b)

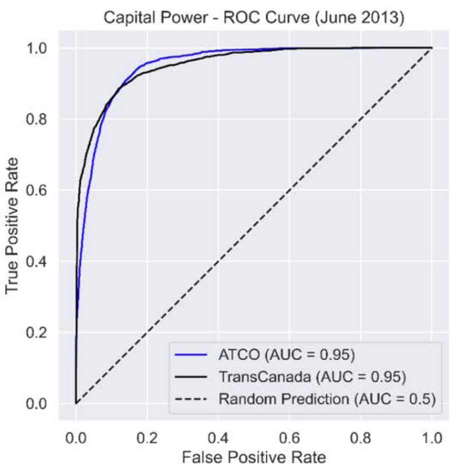


(c)

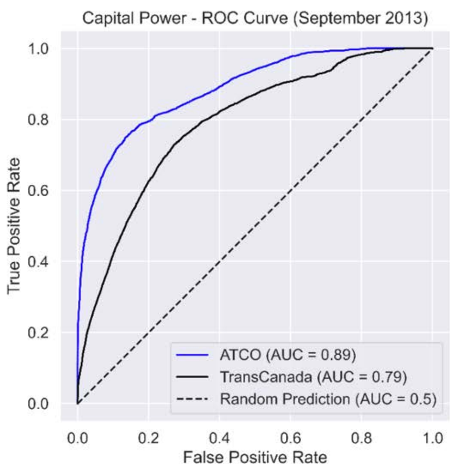


(d)

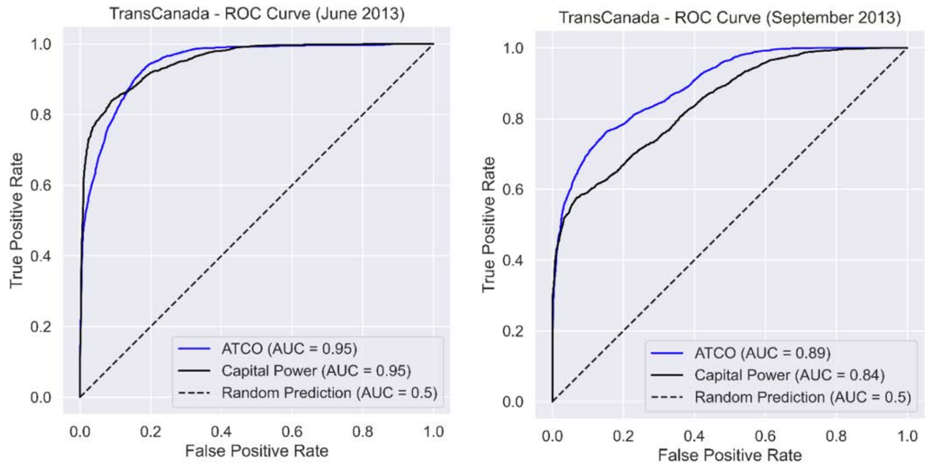
Figure B.1: Precision and f1-score—Capital Power and TransCanada



(a)



(b)



(c) (d)

Figure B.2: ROC curves—Capital Power and TransCanada

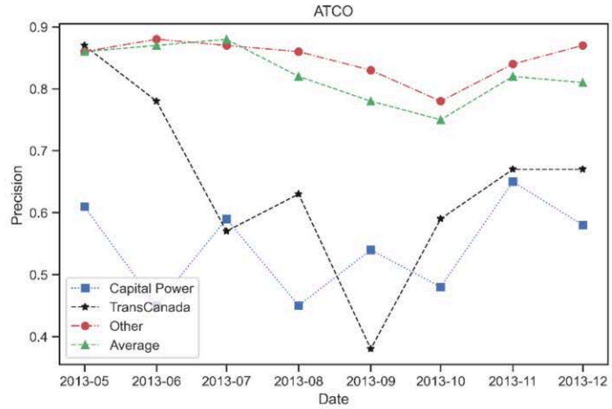
Panel (A) Average (May 2013 -Jul 2013)			Panel (B) Average (Aug 2013 -Dec 2013)		
	Permutation Importance	Std. Dev.		Permutation Importance	Stc Dev.
price int	0.155	0.003	*price int*	0.165	0.003
q block size	0.090	0.002	*q block size*	0.096	0.002
price decimal	0.043	0.001	*Dprice up*	0.030	0.001
dec freq	0.028	0.001	*Dprice down*	0.022	0.001
Dprice down	0.028	0.002	*price decimal*	0.021	0.001
Dprice up	0.028	0.002	*rank*	0.005	0.001
rank	0.006	0.001	*Dquant down*	0.002	0.001
Dpdec up	0.005	0.001	*Dpdec down*	0.002	0.001
Dpdec down	0.005	0.001	*Dpdec up*	0.002	0.001
Dquant down	0.002	0.001	*Dquant up*	0.002	0.001
decimal diff	0.001	0.001	*decimal diff*	-0.001	0.001
Dquant up	-0.0004	0.001	*dec freq*	-0.004	0.001

Table B.1: Permutation Importance—Capital Power

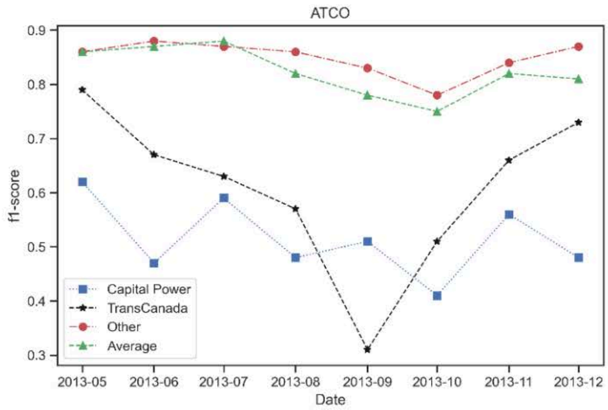
Panel (A) Average (May 2013 -Jul 2013)	Permutation Importance	Std. Dev.	Panel (B) Average (Aug 2013 -Dec 2013)	Permutation Importance	Std. Dev.
price int	0.221	0.003	*price int*	0.190	0.003
q block size	0.100	0.002	*q block size*	0.067	0.002
price decimal	0.034	0.001	*price decimal*	0.038	0.001
Dprice up	0.029	0.001	*Dprice up*	0.033	0.001
Dprice down	0.027	0.002	*Dprice down*	0.025	0.001
Dquant down	0.007	0.001	*Dquant down*	0.002	0.001
Dquant up	0.004	0.001	*Dpdec up*	0.002	0.001
dec freq	0.004	0.001	*Dpdec down*	0.001	0.001
rank	0.004	0.001	*Dquant up*	0.001	0.001
Dpdec up	0.003	0.001	*dec freq*	0.0004	0.001
decimal diff	0.003	0.001	*decimal diff*	0.0001	0.001
Dpdec down	0.002	0.001	*rank*	-0.002	0.001

Table B.2: Permutation Importance—TransCanada

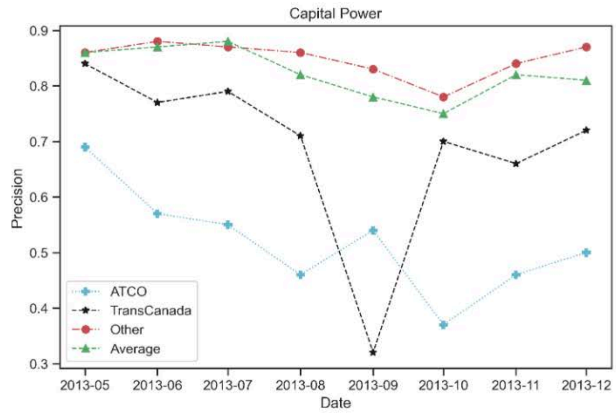
B.2. Decision Tree



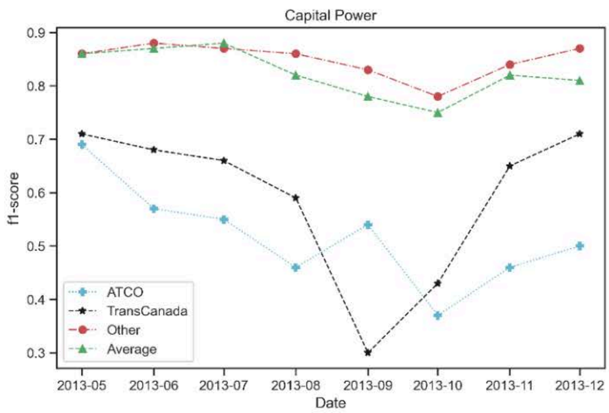
(a)



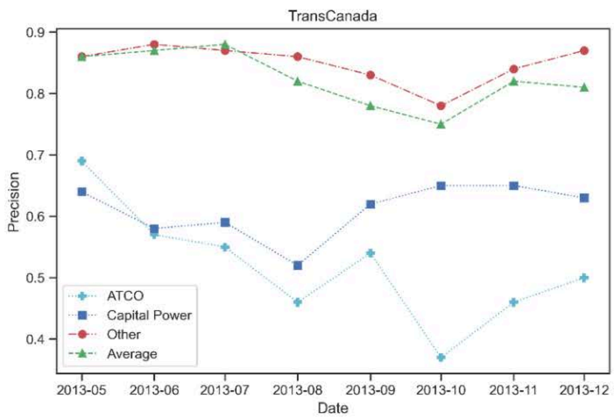
(b)



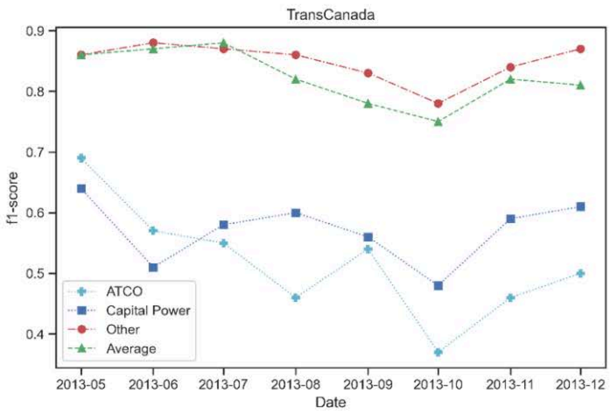
(c)



(d)

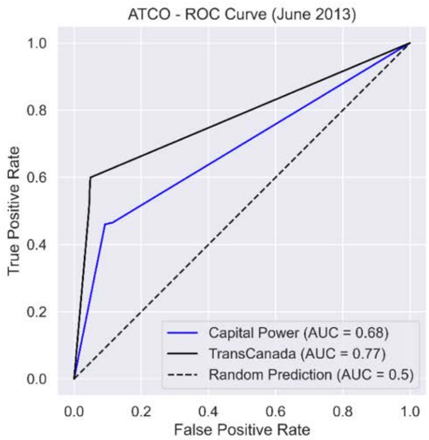


(e)

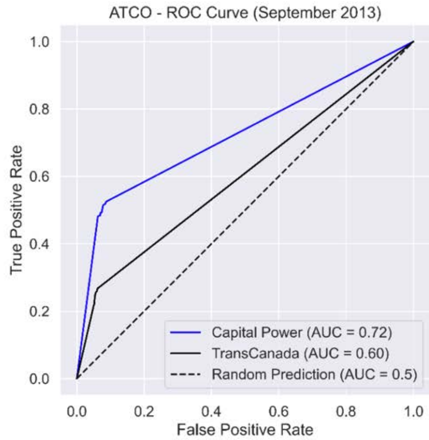


(f)

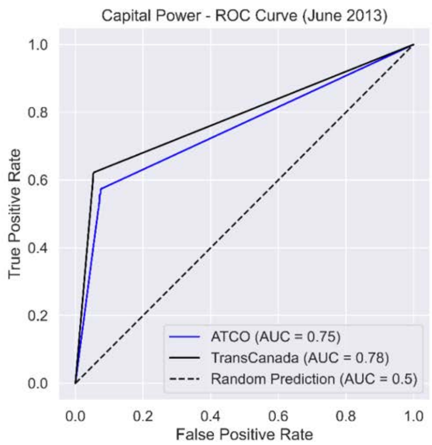
Figure B.3: Precision and f1-score



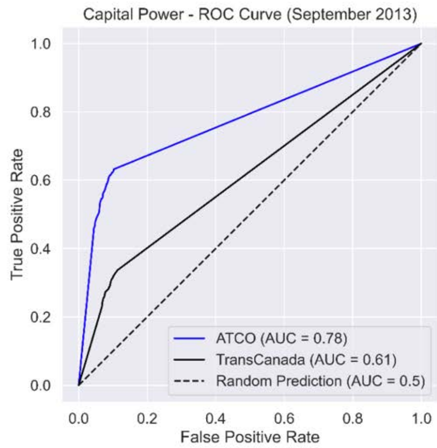
(a)



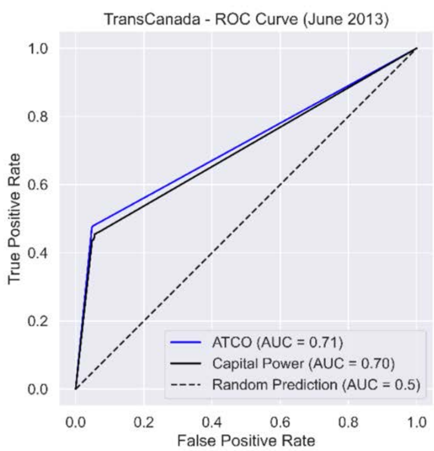
(b)



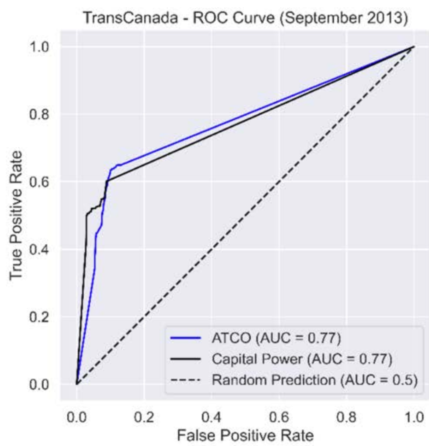
(c)



(d)



(e)



(f)

Figure B.4: ROC Curves

	Permutation Importance	Std. Dev.
q block size	0.196	0.003
price decimal	0.158	0.003
price int	0.132	0.003
dec freq	0.047	0.001
Dprice down	0.027	0.001
Dpdec down	0.026	0.001
Dprice up	0.021	0.001
Dquant down	0.020	0.001
Dquant up	0.013	0.001
Dpdec up	0.011	0.001
rank	0.004	0.001
decimal diff	-0.001	0.001

Table B.3: Permutation Importance—ATCO (May 2013) Using the Decision Tree Classifier