Investigating a Case of Alleged Collusion in Michigan Public Oil and Gas Lease Auctions

Lucas Do[1]

Professor James W. Roberts, Faculty Advisor
Professor Michelle P. Connolly, Honors Seminar Instructor

[1] Lucas graduated in May 2019 from Duke University with High Distinction in Economics and a second major in Mathematics. Starting in the fall of 2019, he will pursue doctoral studies in economics at Harvard University. He can be reached at khuongdo1402@gmail.com.

Acknowledgements

I would like to gratefully acknowledge the Economics Department of Duke University for sponsoring my summer research with the Davies Fellowship, without which I would not have been able to finish my initial data work in a timely manner. To my advisors Professor James Roberts and Professor Michelle Connolly, thank you for your extremely useful feedback and encouragement throughout the writing of this paper. I also want to thank Professor Adam Rosen and Professor Leslie Marx for their help with a few technical details of my econometric specification in the early stages of this project, as well as my fellow students in the Honors Thesis seminar for their insightful suggestions during its middle stages. Moreover, I am grateful to the Michigan Department of Natural Resources officials with whom I corresponded through email, and who wish to remain anonymous, for generously providing me with the bidding data, their institutional knowledge and insider perspectives.

Finally, to my family and friends, thank you for always being there and believing in me. I am eternally indebted to my parents, Cong Do and Hoa Nguyen, without whose undying emotional and financial support I would never have gotten to where I am today.

Abstract

The state of Michigan administers oil and gas lease auctions semiannually through the Department of Natural Resources. In June 2012, the international news outlet Reuters published allegations of bid-rigging in the auctions following May 2010. This paper empirically investigates the validity of Reuters' allegations by analyzing auction bid sheets from 2008 to 2018 as well as other data reflecting market conditions over time. To this end, I first formulate a benchmark structural model of bidders' valuations and estimate it with auction data from a period during which I assume no collusion occurred. Then, I extend the benchmark model by endogenizing bidders' decision to collude. Using the extended model and the estimated benchmark parameters, I apply the simulated method of moments to solve for the collusive probability that "best" explains the observed bids during the alleged period of collusion. After discovering strong evidence for bid-rigging, I run counterfactual simulations to estimate the revenue damage caused to the state government by this anticompetitive bidding behavior. I find that the estimated revenue damage, summed over the entire alleged collusive period, totals over $450 million. However, although these findings lend support to Reuters' allegations and are contrary to the Department of Justice's conclusion in 2014 after they had probed the case, they should be approached only with caution, given the limitations of the available data on the potential bidders.

JEL Codes: L4, D44, L71.
Keywords: empirical auctions, oil and gas leases, industrial organization, collusion, bidding rings.

1. Introduction

Twice a year, once in May and once in October, the Michigan Department of Natural Resources (DNR) administers a public auction for oil and gas leases on state-owned land. The auction is oral ascending bid, wherein bidders take turns verbally stating their bids for a parcel of land starting at a reserve price set by the DNR. The parcel is then awarded to the highest-bidding participant, and the process repeats for the next parcel in the catalog until the auctioneer has run through all parcel offerings. After paying their bid amount for a parcel, the winner earns proprietary rights to the land for five years, during which they can apply for a drilling permit at the Department of Environmental Quality. Should they end up drilling the parcel, the lease will remain in effect for as long as oil and/or gas are produced in paying quantities, even when the five-year term has expired. As royalty, the state will also receive one-sixth of the revenues from any oil or gas produced on the parcel. These auctions are an important source of revenues for Michigan, having generated over $750 million for the state over the last 10 fiscal years.[2]

Of interest to my thesis is the May 2010 auction which, thanks to its all-time high average lease price of $1,413 per acre, alone raised as much money as Michigan had ever raised in its entire leasing history.[3] However, the following October auction raised just $28 per acre of land offered. Moreover, subsequent lease prices never returned to the May 2010 levels (see Figure 1a). This paper investigates whether bidder collusion was responsible for this sharp, persistent fall in the auction prices, during an otherwise auspicious period of shale boom that had driven the success of the May 2010 auction.

One possible explanation for the plunge in prices has to do with changes in the hype surrounding Michigan's Collingwood shale formation. Excitement about drilling Collingwood began in spring 2010, when Calgary-based Encana Corporation successfully tested a Collingwood well, aptly called "Pioneer", the first well in Michigan drilled using a modern technique called horizontal hydraulic fracturing. Pioneer's unprecedented initial production numbers, peaking at 3.2 million cubic feet of gas in a single day,[4] attracted a lot of attention from oil and gas firms in the May 2010 auction. Nonetheless, according to my email conversation with a DNR official, this hype waned over time due to disappointing well performance results from Encana's subsequent development efforts in the Collingwood shale, although the DNR official claims some people believe it still has good potential.

[2] See https://www.michigan.gov/documents/dnr/OG_FAQ_FINAL_401887_7.pdf.
[3] See https://www.reuters.com/article/us-chesapeake-antitrust-settlement-idUSKBN0NF1ZV20150424.
[4] See https://www.respectmyplanet.org/publications/michigan/michigan-oil-gas-monthly-july-2015.

Figure 1.1. Average lease price over time.

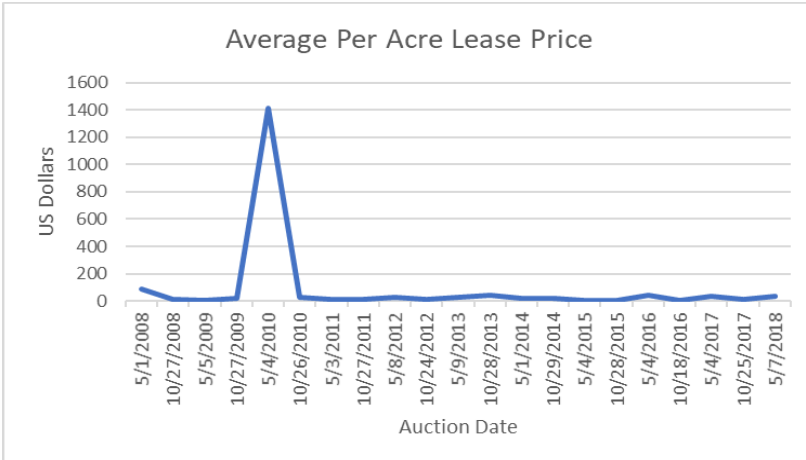
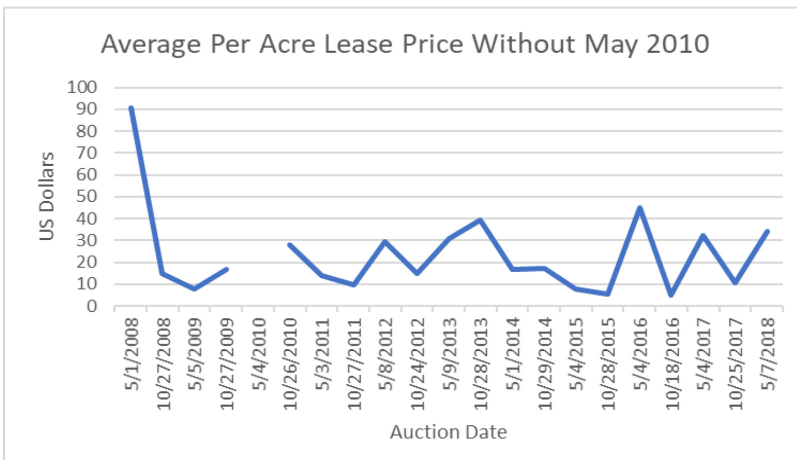


Figure 1.2. Average lease price over time, excluding the May 2010 auction.



Another possible reason for the precipitous fall in lease prices involves alleged collusive agreements between participating oil and gas companies. In June 2012, the international news agency Reuters published a report revealing that executives of Encana and Chesapeake Energy Corporation had exchanged emails about avoiding bidding each other up immediately following the May 2010 auction. Specifically, they discussed dividing the Collingwood shale into Michigan counties where each would be an exclusive bidder both in the upcoming public auction and in prospective deals with private landowners. Encana would claim Charlevoix, Cheboygan, Kalkaska and Crawford, whereas Chesapeake would claim Emmet, Presque Isle, Roscommon, Missaukee and Grand Traverse.[5] The report prompted a federal

[5] See https://www.reuters.com/article/us-chesapeake-encana-wire/exclusive-chesapeake-encana-plotted-to-suppress-land-prices-documents-idUSBRE85O0DU20120625.

antitrust investigation by the U.S. Department of Justice (DOJ) against the two companies later that month, who admitted to no wrongdoing. In April 2014, the DOJ concluded its probe without filing any charges.

However, Chesapeake and Encana continued to face state charges as Michigan accused them in March 2014 of colluding to suppress oil and gas lease prices. Despite having the DOJ's decision from the federal case in their favor, Encana agreed to pay $5 million as a civil settlement to Michigan in May. Chesapeake followed suit in April 2015, paying $25 million to settle the antitrust as well as other fraud and racketeering charges from private land owners. Because the state did not keep a bid-by-bid record of the auctions, it was ultimately unclear whether Chesapeake and Encana ever bid against each other following the email exchange.[6] However, if Reuters' allegations were true, then the bid-rigging might have contributed substantially to the drastic decrease in lease prices from May 2010. Furthermore, the fact that prices never shot back up in Figure 1b suggests that this coalition—which, given the size of the jump, possibly involved more entities than just Chesapeake and Encana—must have remained in effect for several auctions thereafter. This would imply that revenue damages from collusion were substantial.

Although the DOJ did not take any action against Chesapeake and Encana, the case is still worth investigating for several reasons. First, the DOJ never publicized any formal analysis of its probe. Also, in general, the DOJ will not file suit without sufficient proof to win in court, so their decision regarding the case does not necessarily imply that they thought the companies were completely innocent. Second, in April 2016, Chesapeake ex-CEO Aubrey McClendon was found guilty by the DOJ of rigging bids with an unnamed company in Oklahoma auctions for oil and gas leases between late 2007 and early 2012, when he was still working for Chesapeake.[7] Although this was in Oklahoma, the time frame curiously overlapped with the case in Michigan.

In addition, my research indicates that the industry's interest in the Collingwood shale stayed strong for several years. Even in late 2012 and early 2013, Michigan news outlets like Midland Daily News and Crain's Detroit Business were still reporting general optimism about Collingwood and on-going drilling activity in the shale.[8] It was not until September 2014 that Encana, presumably due to underwhelming production outcomes, decided to sell all of its Michigan leases to Marathon Oil Company

[6] My advisor James W. Roberts requested audio recordings of the October 2010 oral auction, but the files involving Chesapeake and Encana had mysteriously been voided.

[7] See https://www.nytimes.com/2016/03/02/business/aubrey-mcclendon-is-charged-with-conspiracy-in-oil-and-natural-gas-bidding.html.

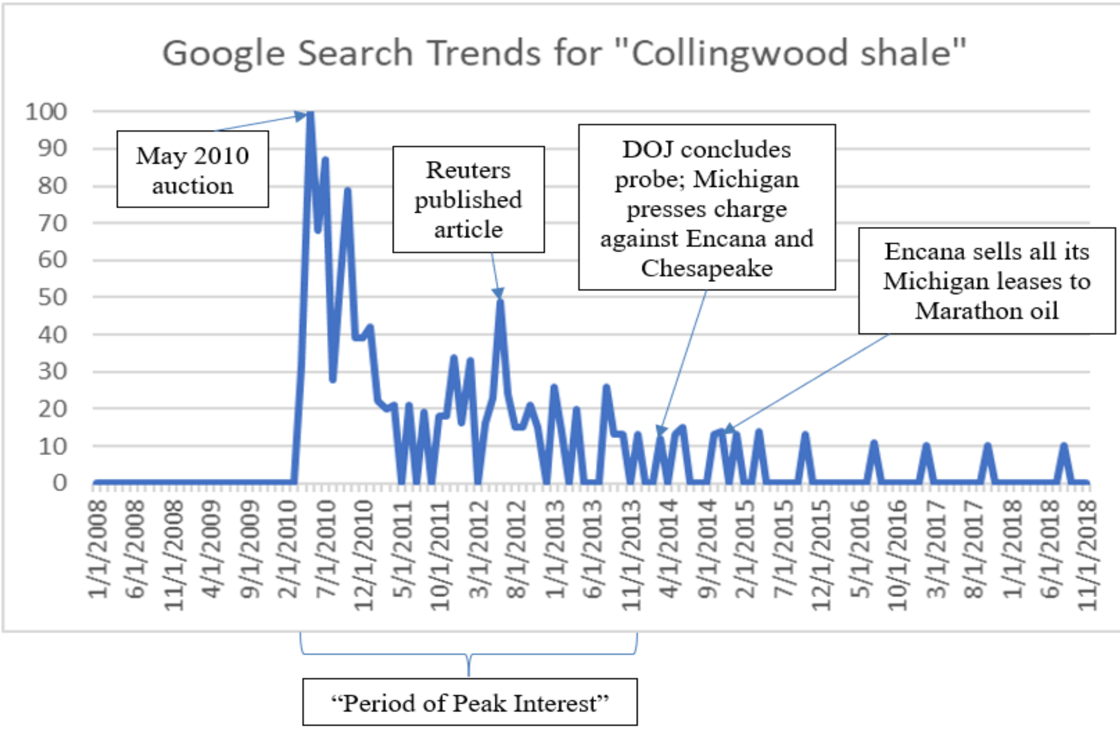
[8] See https://www.ourmidland.com/news/article/Fracking-technology-coming-to-Michigan-6966569.php and https://www.crainsdetroit.com/article/20130324/NEWS/303249962/hydraulic-fracturing-in-michigan-waiting-for-the-boom.

to "focus on more profitable operations elsewhere".[9] To approximate these hype patterns, I plot in Figure 1.3 the Google search trends for the phrase "Collingwood shale" and label the key events on the timeline. The graph suggests that the shale was still in people's minds for a few years after May 2010—at least until around 2013. Therefore, it is hard to fathom that the dwindling hype alone caused average lease prices to drop by 98% within a mere span of five months, from May to October 2010. These observations corroborate a comment made by the Michigan DNR official with whom I corresponded via email, that "2010-2013 [marked] the period of peak interest in the [shale]."

For these reasons, evaluating the validity of Reuters' allegations is the primary focus of this paper. My investigative strategy involves first constructing a benchmark model of bidder valuations. I estimate this model using only winning bid data from May 2008 to May 2010, and from October 2012 to May 2012—auctions which I suppose *were* competitive, given the above chain of events. Then, utilizing the estimated "benchmark" parameters and upon extending the model to allow for the possibility of collusion, I solve for the endogenous collusive probability which generates simulations that "best" match the data moments from October 2010 to May 2012, the assumed "collusive period" up to just before Reuters' publication. The explicit assumption is that if collusion *did* occur, then it did *not* happen before the Chesapeake-Encana email was sent and did *not* continue once Reuters' article had stirred the pot. Finally, if collusion is found to have been likely, I will measure the counterfactual revenue damages.

[9] See http://www.interlochenpublicradio.org/post/encana-leaving-michigan.

Figure 1.3. Google search trends for "Collingwood shale" from 2008 to present.



Note: "Numbers represent search interest relative to the highest point on the chart for the given region and time. A value of 100 is the peak popularity for the term. A value of 50 means that the term is half as popular. A score of 0 means there was not enough data for this term." Source: https://google.com/trends.

2. Literature Review

This section discusses the auction literature as it relates to the methodology and modeling choices made in this paper.

In the literature, auctions are typically classified as "independent private value" (IPV) or "common value" (CV). In an IPV auction, bidders' random valuations of the object are independently distributed. By contrast, the CV model assumes the existence of a random common component that, once realized, is equally valued by all the bidders, implying that their valuations are not independent.[10] In oil and gas lease auctions, the common component could be the physical amount of oil and/or natural gas underneath the parcel, whereas the private component could represent the firm's idiosyncratic technology. This technology is privately known to the firm itself but never to others, and affects its valuation of the parcel

[10] To be technically precise, CV in this paper refers to both the *pure* common value and the affiliated private value models, the latter assuming both a private value component and a common component in the valuations.

through, for example, its ex-ante knowledge of the area's geology, the costs of drilling and the ability to dig deep enough to extract all of the available resource.

For this thesis, I employ an Independent Private Values (IPV) paradigm to model the Michigan oil and gas lease auctions, as opposed to a Common Values (CV) one. Most papers in the empirical literature on oil and gas lease auctions adopt the latter approach. For example, Hendricks, Porter and Tan (1993) use a CV framework to analyze U.S. federal government auctions of offshore oil and gas leases on marginal tracts. Haile, Hendricks and Porter (2010) examine federally sold oil and gas leases on the Outer Continental Shelf with a CV assumption. More recently, Bhattacharya, Ordin and Roberts (2018) look at identification and estimation of a CV model of Permian Basin oil auctions organized by the New Mexico State Land Office. The two theoretical paradigms could have varying implications on auction outcomes. My assumption of IPV is mainly due to the fact that I only observe the winning bids in my data. Specifically, Athey and Haile (2002) establish that unless all bids are observed, neither the ascending bid nor the second-price auction model is identified as long as a common component is present. On the other hand, the same paper states that the IPV model is identified from just the transaction price.

Despite its potentially limiting exclusion of a common component, my modeling choice of IPV is precedented by Kong (2017a) who, like Bhattacharya et al. (2018), also looks at Permian Basin oil lease auctions. Kong justifies her IPV assumption with the following arguments. First, the Permian Basin has a long history of development and exploration, rendering public knowledge of the area's geology less noisy. Second, the period covered by the data coincides with the boom in horizontal hydraulic fracturing, which enhances production certainty. Both factors diminish the influence of the common component—the quantity of oil in the parcel—in the CV framework. Intuitively, if all bidders share roughly similar beliefs about the underlying oil quantity, then it is primarily the private components—the individual signals or technologies—that propel the differences between bidders' valuations. The above reasoning applies to my settings. Not only does Michigan have a similarly extended history of oil and gas development dating back to 1860, but the years covered by my data also span the horizontal fracking boom, which began in the early-to-mid 2000s.[11,12]

Within the IPV framework, DeMarzo, Kremer and Skryzpacz (2005)—hereafter abbreviated as DKS—provide the theoretical foundation for constructing my benchmark model. DKS introduce a model of contingent payment auctions, where the bids are securities whose total payments to the seller depend

[11] See http://www.michiganoilandgas.org/oil_and_gas_in_michigan_pre_1900.
[12] See https://www.ft.com/content/2ded7416-e930-11e4-a71a-00144feab7de.

on the item's realized cash flow. This model is suitable for studying the Michigan oil and gas lease auctions, which follow a "bonus-bid" format, an example of a security bid. In a bonus auction, the participant who bids the highest upfront cash amount (the "bonus") wins the object. If the object ends up generating any revenue for the winner, he then must pay an additional percentage (the "royalty rate") of that revenue to the auctioneer. Recall from earlier that this is precisely the way Michigan oil and gas lease auctions operate: the object here is just the parcel of land offered, and the royalty rate is one-sixth. More details of the DKS model are provided in Section 4.

When extending the benchmark DKS model to allow for collusion, I rely on the theoretical literature on bid-rigging. Graham and Marshall (1987) construct a model of collusive bidder behavior at an IPV second-price or oral ascending bid auction. They establish that any coalition is viable and more profitable than competitive bidding, provided that it can eliminate meaningful competition among members at the main auction, redistribute the benefits among its members, and conceal its existence from the auctioneer. Marshall and Marx (2007) outline how in an IPV second-price auction environment, a bidding ring will suppress the bids of all members except that of the member with the highest valuation. Intuitively, because any deviating ring member would have to compete with the highest ring and non-ring bidders, there is no incentive to do so. Hendricks and Porter (2007) summarize several incentive-compatible, ex-post efficient mechanisms that give rise to a bidding ring behaving as described above. Given the purpose of this paper and the lack of information beyond the winning bids, I abstract from the exact mechanism used by the ring. Instead, I take as given that whenever collusion is modeled at the bidding stage, the coalition will have only its highest-valuing member bid, while all the other ring members will effectively bid zero (or any sufficiently low amounts).

On the empirical side, Baldwin, Marshall and Richard (1997) guide my strategy of modeling bidders' collusive behavior. The authors specify five structural models lending support to allegations of bidder collusion at Forest Service timber sales in the Pacific Northwest in the 1970s. These five models include a noncooperative model with unit supply, a collusive model with unit supply, a noncooperative multiunit supply model, and two nesting models incorporating both multiunit supply and collusion. Mirroring the approach in their second model, I treat collusion as an outcome of the bidders' independent decisions. More precisely, each bidder decides whether or not to join a bidding ring with some probability, which could be parametrized to depend on several parcel-level characteristics.

Finally, to solve for the collusive probability that "best" matches the data for the alleged collusive period, I employ the simulated method of moments technique introduced by McFadden (1989). I then evaluate whether collusion was a likely driver of the low lease prices in the few auctions after May 2010.

3. A Brief Overview of Relevant Theory for IPV Ascending Auctions

Before providing empirical specifications, I summarize some basic theoretical results about IPV ascending auctions as they relate to certain simplifying assumptions inherent in my model.

3.1 Equivalence to Second-Price Auctions

In its most basic form, an ascending auction can be modeled as a button auction. Imagine an auction room where all bidders have their fingers on a button. Starting from zero, the seller continuously raises the price of the item, and a bidder can push the button whenever he wants to exit the auction. The price at which a bidder exits is treated as his bid, and the auction automatically concludes once there is only one bidder left. A classic result in auction theory states that in an IPV button auction, it is a weakly dominant strategy for each bidder to bid his true valuation (Hendricks & Porter, 2007). It follows that the auction ends as soon as the bidder with the second-highest valuation drops out, implying that the winner must pay the seller an amount equal to the second-highest valuation among the bidders. This in turn makes the button auction strategically equivalent to an IPV second-price auction.

Of course, a typical oral ascending auction like that for Michigan oil and gas leases is not necessarily equivalent to a button auction. The main reason is that prices in practice are not continuously rising; instead, bidders tend to place "jump bids". As a result, the winner might have to pay more than he would expect in the button auction. To illustrate, suppose there are two bidders. Bidder 1's valuation is 10, while bidder 2's valuation is 8. Bidder 2 goes first and announces a bid of 6. Bidder 1, overestimating bidder 2's valuation and wanting expedite the process, announces a "jump bid" of 9. Obviously, bidder 2 will stop bidding because the current price already exceeds his valuation. Bidder 1 then has to pay 9 to the seller and gain a surplus of 10 – 9 = 1. Had the auction been in the button format, bidder 2 would have pressed the button at 8, and bidder 1's surplus would have been 10 – 8 = 2.

This discrepancy arises only because the bidder 1's jump from 6 to 9 is "too big". If every bidder is assumed to make a sufficiently small increment on each other's bid each time, then the outcome of the oral ascending auction will converge to that of the button auction. To see why, suppose bidder 2 above starts with a bid of 0, followed by bidder 1's bid of 0.01, followed by bidder 2's bid of 0.02, and so on.

Then the auction will stop as soon as it is bidder 1's turn to submit a bid of 8.01 (or 8.00, if bidder 2 went first), which is approximately the revenue generated by the button or second-price auction.

For analytical convenience, I therefore impose this behavioral assumption of "small bid increments". In doing so, I can approximate the Michigan oral ascending auctions with the more tractable button or second-price version and assume truth-telling as each bidder's equilibrium strategy. This simplifying assumption, though strong, is commonly made in empirical papers dealing with IPV oral-bid auctions, such as Kong (2017) and Baldwin et al. (1997). Moreover, in the Michigan lease auctions, bidders are the ones calling out the bids, so it may be reasonable to assume that they never want to "jump" by too much lest they end up paying more than they need to. In subsequent sections, my auction model will thus be referred to simply as "second-price".

3.2 Collusion in IPV Second-Price Auctions

As mentioned in the literature review, without specifying the precise collusive mechanism, I assume that any coalition forces all but its highest-valuing member to bid zero—or just low enough to never affect the final outcome. Furthermore, I use the following stylized fact from Graham and Marshall (1987), which is implicitly assumed in Baldwin et al. (1997): when two or more distinct rings form at the same auction, they will always merge into a single coalition to maximize profits. I also qualify that for each bidder, ring participation does not necessarily occur with certainty despite that an all-inclusive ring is most profitable to the bidders (Graham & Marshall, 1987). This makes sense practically because a ring may not want to divulge its existence to everyone due to the increased risk of getting caught and punished.

4. The Models of Michigan Oil and Gas Lease Auctions

4.1 The Benchmark DeMarzo-Kremer-Skrzypacz (DKS) Model

4.1.1 Setup

The setup closely follows the Michigan lease auctions. A group of n potential bidders vie for a parcel of land in a second-price auction. The seller sets a known first reserve price r, which was \$13 per acre from May 2008 to October 2011, \$12 for the May 2012 auction, and \$10 for all subsequent auctions.[13] Should no bids be received at r or above, the seller might lower the minimum bid by \$8 (i.e. to \$5, \$4 and

[13] From here on, the term "auction" does not refer to the auctioning of any specific parcel, but to the larger, twice-a-year auctions where many parcels are being leased at once.

\$2 per acre, respectively) and reoffer the parcel at the new reserve price. According to another DNR official I talked to, the decision to reoffer when there is no bid is randomized through a computer and differs from one auction to another, depending how many parcels have been sold.

Each bidder i submits a cash bid c_i, and the bidder who submits the highest bid is rewarded ownership of the lease. Provided that bidder j wins, the parcel yields a stochastic total future payoff P_j. Denote by S_{zj} his private signal about this payoff conditional on a vector of observed parcel-specific characteristics z, to be discussed in Section 5. Each bidder i is assumed to know his own signal s_{zi}, but not those of his competitors. The following assumptions are taken from DKS:

DKS ASSUMPTION 1. For all z in the support, the conditional private signals $S_z \equiv (S_{z1}, \dots, S_{zn})$ and payoffs $P \equiv (P_1, \dots, P_n)$ satisfy the following properties:

(a) Symmetric IPV: Each S_{zi} is independently drawn from the same distribution $G(s|z)$.

(b) $\mathrm{E}[P_i|S_{zi}] = S_{zi}$. That is, a bidder's signal given the covariates is normalized to be the expected payoff of the parcel in the event that he wins.

Denote by $B(P)$ the winner's total payment to the seller. In the Michigan auctions, which are bonus-bid auctions with royalty rate $\frac{1}{6}$, this is the sum of the cash bid c and the contingent payment $\frac{1}{6}P$:

$$B(P) = c + \frac{1}{6}P. \tag{1}$$

The model consists of a bidding and a post-bidding stage, which will be discussed in reverse chronological order.

Post-bidding stage. Define $ER(s) \equiv \mathrm{E}[B(P)|S_{zi} = s)]$ to be the seller's expected revenue given that the winner's conditional private signal is s. In the case of Michigan lease auctions, we have:

$$ER(s) = \mathrm{E}\left[c + \tfrac{1}{6}P\middle|S_{zi} = s,\right] = c\ + \frac{1}{6}s, \tag{2}$$

where the first equality comes from (1), and the second equality follows from DKS assumption 1(b).

Bidding stage. For bidder i whose conditional private signal is $S_{zi} = s$, the expected profit provided that he wins and pays the seller a cash amount c is:

$$\begin{aligned} ExpectedProfit(s, c) &= \mathrm{E}[P_i|S_{zi} = s] - ER(s) \\ &= s - \left(\frac{1}{6}s + c\right) \end{aligned} \tag{3}$$

Recall that in an IPV second-price auction, it is every bidder's weakly dominant strategy to bid his valuation. Therefore, although this is not the final outcome for this type of auctions, *if* bidder i's bid cleared the auction, and *if* he had to pay his valuation to the seller, then truth-telling equilibrium would imply that the expected payoff in (3) is zero. In other words, the bidder neither gains nor loses from paying his entire valuation to obtain a parcel that, by definition, is worth as much as he values it. Solving for the cash bid *c* as a function of *s* gives:

$$c(s) = \frac{5}{6}s \tag{4}$$

In the sense that everyone's bid reveals his true valuation, the cash bid could be interpreted as the bidder's private valuation of the parcel, given his signal *s*. This interpretation has an intuitive economic meaning: a bidder's valuation of a parcel in (4) equals how much money he expects it to generate (s), net the royalty he would then have to pay to the seller ($\frac{1}{6}s$).

Define $V_z \equiv c(S_z)$ as the random valuation of the parcel conditional on *z*, with cdf $F(v|z)$ and density function $f(v|z)$. DKS Assumption 1(a) implies that the bidders' conditional valuations are also independent and identically distributed, as functions of independent random variables are themselves independent.

4.1.2 Deriving the Conditional Density Function for the Observed Winning Bid

Let *W* be the winning bid, and $F_w(w|z,n,r,Auc)$ the corresponding conditional cdf, where Auc denotes the auction in question. Recall that in a second-price auction, the winner pays the second-highest valuation, provided that it is above the reserve price *r*. Let $V_{z1} \leq V_{z2} \leq \cdots \leq V_{zn}$ denote the order statistics from n independent draws from V_z, and $f_{(k:n)}(v|z)$ the conditional density function of the k^{th} order statistic. For example, the second-highest valuation corresponds to $V_{z(n-1)}$. The relationship between $f_{(k:n)}(v|z), f(v|z)$ and $F(v|z)$ is summarized below:

$$f_{(k:n)}(v|z) = nf(v|z)\binom{n-1}{k-1}\left(1-F(v|z)\right)^{n-k}F(v|z)^{k-1}. \tag{5}$$

To derive the conditional density of the winning bid $f_w(w|z)$, it helps to enumerate all the possible scenarios. These scenarios are summarized in terms of order statistics in Table 4.1 below.

Table 4.1. Benchmark winning bids. V_{zk} = k^{th}-order statistic of the conditional valuation.

Scenario #	Scenario	Winning Bid
1	$V_{z(n-1)} > r$	$V_{z(n-1)}$
2	$V_{zn} \geq r$ and $V_{z(n-1)} < r$	r
3	$V_{zn} < r$	$< r$

To expound the scenarios more clearly, I say that a potential bidder *participates in the parcel offering* if his valuation exceeds the minimum bid, and that the parcel *fails* if it has no participants. In the first scenario, the winning bid $w > r$ if two or more bidders participate in the offering, in which case w equals the second-highest valuation. Second, $w = r$ if the winner is the only participant, in which case his dominant strategy is to pay the minimum bid required. Third, $w < r$ if the parcel fails the first time, in which case the highest valuation must be lower than r.

As an extension of the model, I could further divide the third scenario into three subcases: (1) $r - 8 < w < r$ if the parcel fails the first time, the reserve price is reduced, and two or more bidders participate in the reoffering; (2) $w = r - 8$ if only one person participates in the reoffering; and (3) $w = 0$ if the parcel fails without being reoffered, or if the parcel fails, is reoffered, and then fails again at the reduced reserved price. However, such a specification requires estimating the probability of reoffer as a model parameter, one for each auction. Having to estimate so many parameters could result in model overfitting and cause maximum likelihood estimation in MATLAB to be numerically unstable. For this reason, the three subcases above are here subsumed under a more general scenario, in which the parcel fails at the first reserve price. Put another way, I treat all observed winning bids below r as if I did not have any other information than the mere fact that they failed to sell at the first round. Considering that only 1% of the parcels in my time frame of interest ended up being sold at the reoffered round, this simplification of the specification is reasonable, and is perhaps necessary for circumventing the computational difficulty of optimizing over a higher-dimensional parameter space.

Piecing these facts together, I can write the conditional density function for the winning bid W:

$$f_w(w|z,n,r) = \begin{cases} f_{(n-1:n)}(w|z), & w > r \\ nF(r|z)^{n-1}(1 - F(r|z)), & w = r \\ F(r|z)^n, & w < r \end{cases} \tag{6}$$

Since it is neither differentiable nor continuous, using standard nonparametric methods to estimate $f_w(w|z,n,r)$ is difficult. To help with estimation, I impose a parametric structure on the conditional distribution of the private valuations, $F(v|z)$.

4.1.3 Parametric Specification

Similar to Baldwin et al. (1997), I assume that the per-acre valuation V_z is lognormally distributed whose associated normal distribution has mean $\mu(z)$, which depends on the covariates z, and variance σ^2. Specifically, the mean is parametrized to have a linear index, so that $\mu(z) = z \cdot \gamma$. The model's parameter vector is then $\theta \equiv (\gamma', \sigma)'$. Although the identification result for IPV auctions in Theorem 1 of Athey and Haile (2002) does not assume a reserve price, given the above parametrization, it is straightforward to show that θ is identified under the standard rank condition:

PROPOSITION 1. Suppose $\mathrm{E}[Z'Z]$ is invertible. Then the benchmark parametric model is identified.

Proof. Assume knowledge of the population distribution of (W, Z). Fix any z. Knowing the joint distribution, I can identify the conditional density $f_w(w|z, n, r)$ in (6). Then I can use the first line of (6) to recover $f_{(n-1:n)}(w|z)$ for all $w > r$. Using $f_{(n-1:n)}(w|z)$ for different values of n and the relationship in (5), I can recover $F(v|z)$ for all $v > r$.

Given $F(v|z)$ for $v > r$ and the assumption that V_z is Lognormally distributed, I can recover $\mu(z)$ and σ because there can exist one Lognormal distribution whose tail matches the identified part above of $F(v|z)$. Since the parametrization of the mean parameter holds for all values of *z*, I can identify:

$$\mu(Z) = Z\gamma. \tag{7}$$

Because knowledge of the population distribution of Z is assumed and $\mu(Z)$ has been recovered, I know $\mathrm{E}[Z'Z]$ and $\mathrm{E}[Z'\mu(Z)]$. It then follows from (7) the rank condition that γ is identified:

$$\gamma = \mathrm{E}[Z'Z]^{-1}\mathrm{E}[Z'\mu(Z)]. \tag{8}$$

■

This parametrization allows me to express the density function of the conditional winning bid in (6) as $f_w(w|z, n, r, \theta)$, replacing $F(v|z)$ with the associated Lognormal cdf. Assuming that the conditional winning bids are i.i.d. across observations, I can then estimate the parameter vector θ by maximizing the conditional log-likelihood function:

$$l_b(\theta) \equiv \frac{1}{T}\sum_{t=1}^{T} ln\, f_w(w_t|z_t, n_t, r_t, \theta)\,, \tag{9}$$

where each $t = 1, 2, \ldots, T$ corresponds to an observation in my data set.

4.2 Endogenizing Collusion

As discussed in Section 3.2, I assume that a bidding ring suppresses the bids of all but its highest-valuing member. Then, following Baldwin et al. (1997), I model bidder collusion as a decision made symmetrically and independently by each bidder: prior to the auction, a given bidder joins a bidding ring with probability π_x, which is allowed to depend on covariates *x*, to be discussed in Section 5. Specifically, I parametrize the probability π_x to be $\pi_x(\beta) = \frac{e^{x\beta}}{1+e^{x\beta}} \in (0,1)$. Lastly, consistent with the benchmark case when mapping valuations to bids, I avoid dealing with the unknown reoffer probability by combining all winning bids below the first reserve price *r* into a single category represented by bids of zero (i.e. all such parcels are simply assumed to fail rather than ever get reoffered).

5. Choice of Covariates

5.1 Determinants of the Valuation Distribution

Incorporated into my model is the assumption that the distribution of a bidder's per-acre valuation of a given parcel varies with *z*, a vector of observables. In particular, ***Z* = (1, *Resource_Price, Size, Non_Development, Collingwood, Hype, Collingwood*Hype*)**, where:

- ***Resource_Price***: Bidders' valuations of a parcel depend on the market price of the particular resource underneath, whether it is natural gas or oil. In general, the Michigan basin is capable of producing both oil and natural gas, depending on the depth of the formation. Figure 5.1, taken from Banas (2012), provides a map of Michigan counties, the span of the Collingwood shale and the estimated oil and gas zones. Because data do not directly reveal which resource underlies a given parcel, I treat all parcels located in counties inside the oil (resp. gas) zone as oil (resp. gas) parcels. Then, as an estimate of their "true" oil price, I use the West Texas Intermediate crude oil price per barrel at the month of the auction. Likewise, for those located in the gas zone, I use the Henry Hub natural gas spot price per barrel of oil equivalent (BOE) at the month of the auction. Given little other information, everything "in between" (i.e. in counties intersected "considerably" by the demarcation line in the figure) is discarded from the analysis to reduce noise.[14] Ignoring these data points may not be as consequential as it sounds: of the 22,765 total observations, only 1,070 belong to the last category, compared to 6,895 in the oil zone and 14,800 in the gas zone. Figure 5.2 displays the numbers of oil versus gas parcels for each auction.
- ***Size***: The size of the parcel, measured in the total number of acres being offered. There may be increasing returns to owning a larger parcel.
- ***Non_Development***: A dummy for whether the lease is a development or non-development lease. A non-development lease, so-classified to protect sensitive lands like public parks and recreational

[14] Due to the subjectivity of this classification rule, for replicability, the following counties are considered by this paper to be in the gas zone: Oscoda, Crawford, Kalkaska, Wexford, Lake, Osceola, Missaukee, Roscommon, Ogemaw, Iosco, Arenac, Gladwin, Clare, Mecosta, Isabella, Midland, Bay, Montcalm, Gratiot, Saginaw, Tuscola, Shiawassee, Clinton, and Ionia. The following counties are considered to be "in between": Alcona, Grand Traverse, Manistee, Newaygo, Kent, Eaton, Ingham, Genesee, Lapeer, Huron, and Sanillac. The rest are considered to be in the oil zone.

areas, does not allow the parcel's land surface to be used for oil and gas development purposes without a separate written permission. For example, the lease owner cannot have a well, pipelines, roads, etc. physically located on the parcel. This means any drilling would need to be done diagonally from wellheads in adjacent areas. Because of the extra costs associated with either directional drilling or negotiating with the Department of Environmental Quality for surface usage, non-development leases should be less valuable than development leases, all else held equal.

- ***Collingwood***: A dummy for whether the parcel is in a county spanned by the Collingwood shale, according to Figure 5.1. Because I am unable to pin down the precise geographic span of the formation, I label a county as "Collingwood" if the shaded region representing the extent of the shale overlaps "sufficiently" with the county.[15] This may lead to a systematic overcount of Collingwood parcels in the data set—something to keep in mind when interpreting the results.
- ***Hype***: A variable that captures the temporal patterns of hype surrounding Collingwood, inferred from the Google search trend values for the phrase "Collingwood shale", whose interpretation has previously been described in detail in Figure 1.3. In particular, since May 2010 has the peak value of 100, I set the trend *index* of the corresponding auction to 100. For all other auctions, I calculate the average Google trend values in the three months leading up to the auctions and use those as their respective trend indexes, as valuations are likely affected by recent as well as current hype.
- ***Collingwood*Hype***: An interaction term between *the* ***Collingwood*** dummy and ***Hype***. This is intended to reflect the reasonable assumption that the impact of the Collingwood hype on valuation is different for Collingwood versus non-Collingwood parcels.

[15] Again, due to the subjectivity of this classification rule, for replicability, the following counties are considered by this paper to be in the Collingwood shale: Emmet, Cheboygan, Presque Isle, Alpene, Montmorency, Charlevoix, Antrim, Otsego, Oscoda, Crawford, Kalkaska, Grand Traverse, Benzie, Leelanau, Wexford, Missaukee, Roscommon, Gladwin, Clare, Osceola, Isabella, Midland, Montcalm, and Gratiot.

Figure 5.1. Estimated oil and gas zones for the Collingwood shale calculated from Trenton Formation Conodont Color Alteration Index Data and pyrolysis data. The shaded region represents the span of Collingwood. Source: Banas, 2012.

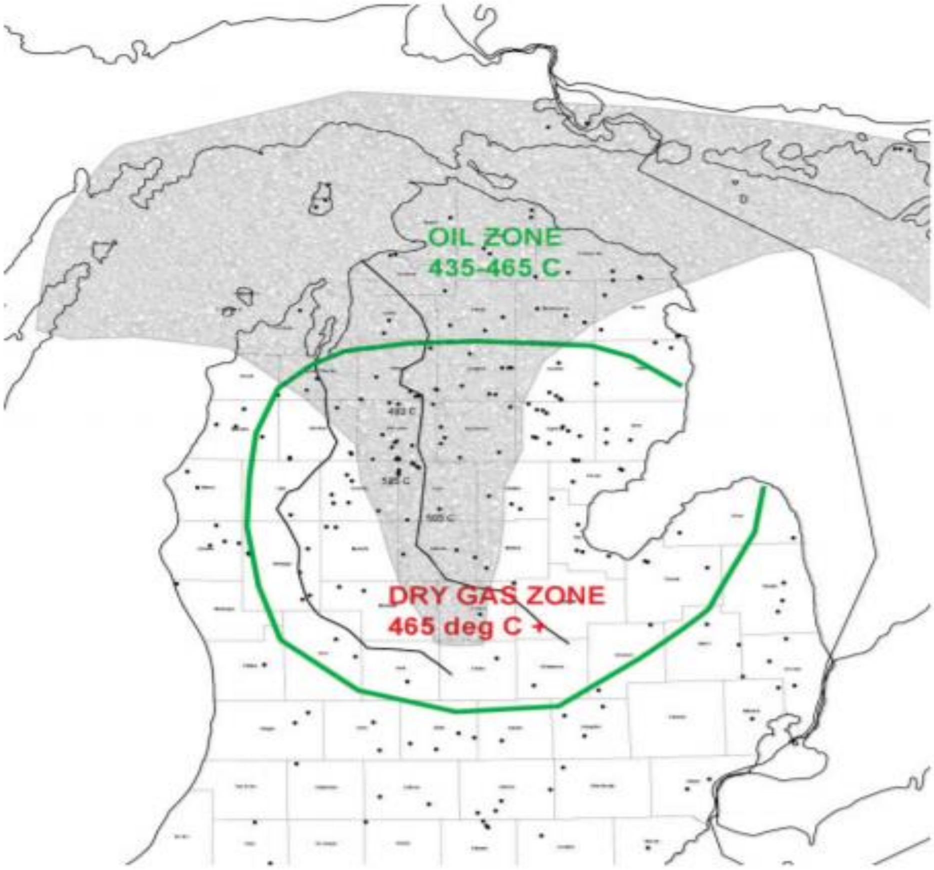
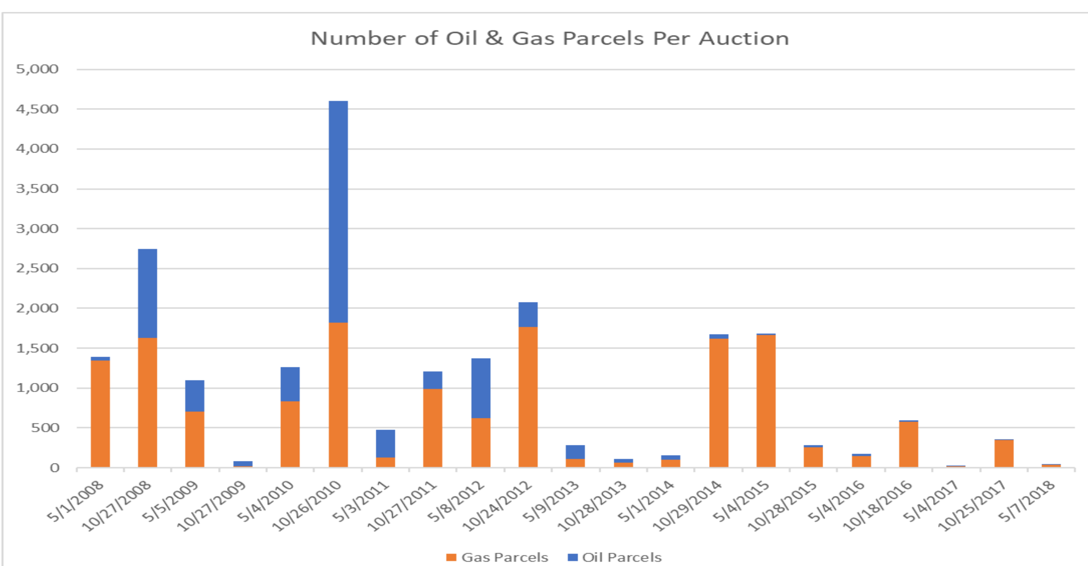


Figure 5.2. Total number of oil and gas parcels for each auction.



5.2 Determinants of the Collusion Probability

Some observed covariates may affect the collusive probability π_x, in the sense that bidders may be more likely to collude on parcels with certain characteristics than on others. When examining bidder collusion at Forest Service Timber Sales, as determinants of the collusive probability, Baldwin et. al (1997) considers a dummy for a specific forest reputed to be more competitive than others and a "bidder proximity" dummy for whether the highest and second-highest bidders were located in the same county. Given the fact that I only observe the winning bids, the most pertinent and natural factor for my setting is whether the parcel is in a county mentioned in the email between Chesapeake and Encana.

In particular, ***X* = (1, *Email_County*)**, where ***Email_County*** is a dummy variable for whether the county is one of Charlevoix, Cheboygan, Kalkaska, Crawford, Emmet, Presque Isle, Roscommon, Missaukee and Grand Traverse. These are the counties that Chesapeake and Encana allegedly wanted to split among themselves, according to the email recovered by Reuters.

Table 5.1 tabulates the number of offerings during the alleged collusive period, grouped by whether they are labeled as ***Collingwood*** and/or ***Email_County***. As expected, all of the counties mentioned in the email are located on the Collingwood shale. Among the Collingwood parcels, over half do not lie in an "email county". It is also noteworthy that over three-quarters of the parcels offered during this period belong in Collingwood, reflecting the contemporary hype around the shale.

Table 5.1. Parcel counts between October 2010 and May 2012 by their ***Collingwood*** and ***Email_County*** designations.

	Email_County	**Not Email_County**	**Total**
Collingwood	4,444	1,360	5,804
Not Collingwood	0	1,863	1,863
Total	4,444	3,223	7,667

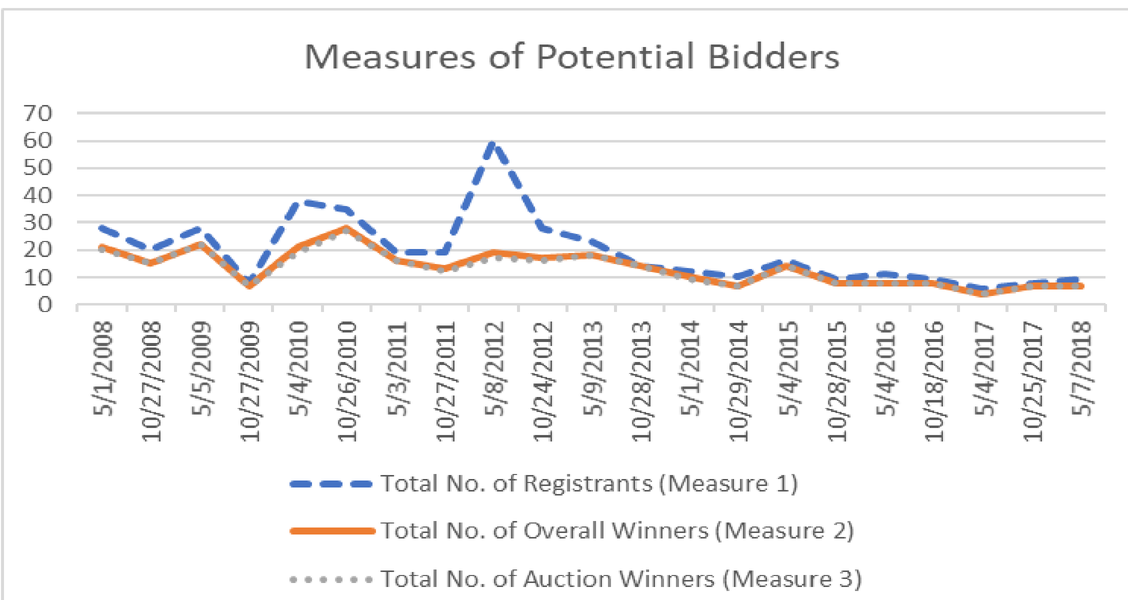
5.3 Measure of Potential Bidders for a Parcel

There are 186 unique registered bidders in my data set, which comprises a total of twenty-one auctions. A bit surprisingly, only 96 of them have ever been a winning bidder. For example, Rich Patterson was a registered bidder in nine of the twenty-one auctions, but never won a single parcel. It is possible that these are bidders who register only to scout out the industry's general level of activity, without the intention to ever participate in the bidding. Due to this large number of possibly "non-serious" bidders, and because I do not observe all the bidders for a given parcel, the number of potential bidders n could be approximated in three ways. First, n could just be the total number of registrants for an auction in question.

Second, n could be the number of registrants who won at least one parcel in that particular auction. Third, n could be the number of bidders, out of the 96 overall winners, who registered for that same auction. As evident in Figure 5.3 below, the latter two measures are mostly identical, so I consider only the second while ignoring the third in my analysis.

As for the first and second measures, both are overestimates of the number of potential bidders for a given parcel because not all registrants, winners or not, desire every parcel in the catalog. However, the second measure, which by definition is always more conservative than the first, could correct some of the upward bias in using the registered bidder list as a proxy for the potential bidders for each parcel in the auction. The downside of this measure is that it ignores "serious" bidders who just never happened to draw a high enough valuation to compete with the others. Nonetheless, this concern is unlikely in my model given the large number of offerings per auction, as illustrated in Figure 4 above. That is, because valuations are assumed independently drawn across parcels in the symmetric IPV framework, even if a bidder has an unlucky valuation draw for a particular parcel, compound probability implies it is extremely unlikely that he is unlucky for all hundreds of them. For these reasons, for the remainder of the paper, I focus solely on the second one as my main measure of the number of potential bidders.

Figure 5.3. Three candidate measures of the number of potential bidders for each auction.



6. Data

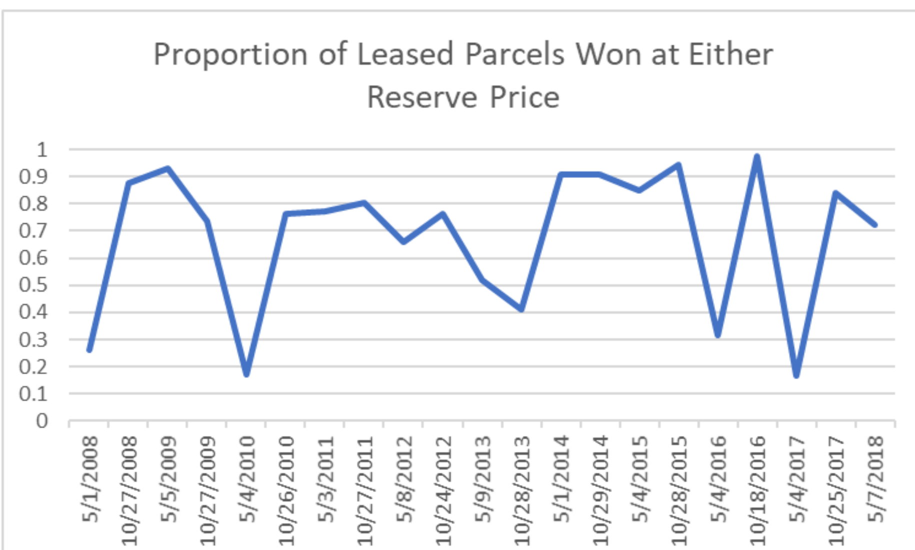
This paper combines data from several sources. Google Trends supplies the online search patterns for the term "Collingwood shale". Monthly WTI crude oil prices per barrel are collected from FRED at the St. Louis Federal Reserve. Monthly natural gas spot prices per million BTU are downloaded from Henry Hub and converted to dollars per barrel of oil equivalent by multiplying by 5.8, an approximation used by the U.S. Internal Revenue Service. Lastly, bidding data were acquired from the Michigan DNR through a Freedom of Information Act request, encompassing all twenty-one of the biannual public oil and gas lease auctions administered from May 2008 to May 2018. There are 22,765 observations in total, each corresponding to a parcel being offered. After dropping the parcels that are "between" the oil and gas zones, 21,695 observations are left. Of these 6,895 are oil and 14,800 are gas parcels. For each parcel offering, I observe the transaction price, the identity of the winning bidder, its geographic location represented by the Public Land Survey System, and the other covariates in Z. From the DNR website, I also obtain a list of all registered bidders for every auction, including the aggregate number of acres they each won and their total immediate (i.e. "bonus") payments to the DNR.

Table 6.1 provides summary statistics for the key variables. The first row of each variable represents the corresponding statistics over the "period of peak interest" in the Collingwood play (i.e. 2010–2013, as per the Michigan official); the second over the same peak period *excluding* the May 2010 auction; and the third over all the auctions outside of the peak period. From the table, there are several points to note about economic activity, interest and implied hype. First, on the supply side, although the peak period comprises less than 40% of all administered auctions (8 out of 21), it contains over half of the offerings in the data set. This difference appears to be driven by the October 2010 auction, which stands out as much larger than the rest at over 4,500 offerings—or around 40% of the total peak-period number of parcels. Second, on the demand side, for both measures of potential bidders, n is likewise greater during the peak period than otherwise. Third, the 3-month average Google search trends are also much higher during the peak period. The same observations hold even if the "anomalous" May 2010 auction is omitted. Thus, these three measures of activity and interest provide evidence for the hypothesis that the hype did not vanish immediately after May 2010, despite eventually dwindling.

However, consistent with the Reuters report, the average winning bids do not match what one would expect from the above hypothesis of an initially persistent hype. From Figure 1.1 and Table 6.1, the inflated average price during the peak period appears mostly driven by the May 2010 auction. If May 2010 is excluded, the lease prices—both overall and for Collingwood parcels—plummet to the non-peak-period levels, although the variation among the latter is smaller. Moreover, to acquire a sense of the general level of competition at each auction, I plot in Figure 6.1 the proportion of all leased parcels sold at either the first or reduced reserve price.[16] Being sold at a reserve price means the parcel is won uncontested. Therefore, the relatively small percentage of leases won at the minimum bid in May 2010 implies that lease winners faced more competition there than in the subsequent auctions. On the one hand, this observation might just be a function of low bidder valuations after May 2010: intuitively, if valuation draws were generally low, then most would be realized below the reserve price, thus causing that one lucky draw above the reserve price to win uncontested. This would be plausible if the Collingwood hype did in fact rapidly vanish. On the other hand, given the evidence for an initially persistent hype, the reduced level of competition during the allegation period is also compatible with the hypothesis of bidder collusion.

Finally, it is interesting to compare the average winning bid for Collingwood and non-Collingwood parcels across the periods. Outside of the peak period, Collingwood parcels seem to be worth nearly 2.5 times as much as non-Collingwood parcels; however, during the peak period excluding May 2010, this ratio decreases to roughly 1.5. Considering that many Collingwood counties are included in the Chesapeake-Encana email and thus are prime suspects for collusion, the deflated Collingwood to non-Collingwood price ratio during this period further corroborates to the collusive hypothesis. That is, if collusion was indeed occurring more to Collingwood parcels, it would likely cause their prices to decrease *relative* to their non-Collingwood counterparts.

Figure 6.1. The proportions of leased parcels won at either the first or reoffered reserve price.



[16] Recall that if a parcel fails to sell at the first reserve price or above, then under the discretion of the DNR, it may be reoffered at a reduced reserve price.

Table 6.1. Summary statistics of key variables. The first row of each variable represents the summary statistics for the peak period (May 2010 – October 2013), the second row for the peak period excluding May 2010, and the third row for all auctions held outside of the peak period.						
Variable	**Obs.**	**Mean**	**SD**	**Min**	**Median**	**Max**
Auction-Specific Variables						
# Registered Bidders for Auction	8	29.75	14.82	14	25.5	60
	7	28.29	15.57	14	23	60
	13	13.38	7.46	6	10	28
# Bidders Who Win at Auction	8	17.38	4.47	12	16.5	27
	7	17.14	4.78	12	16	27
	13	10.46	5.53	4	8	22
Oil Price at Time of Auction ($/barrel)	8	90.26	9.37	73.74	92.00	100.90
	7	92.62	7.11	81.89	94.51	100.90
	13	68.87	24.02	46.22	59.27	125.40
Gas Price at Time of Auction ($/BOE)	8	20.97	3.44	14.09	21.03	25.00
	7	20.53	3.48	14.09	20.71	25.00
	13	23.70	14.36	11.14	18.27	65.37
Google Trends 3-Month Average Before Auction	8	29.3	32.1	6.7	15.3	100.0
	7	19.1	15.8	6.7	13.7	54.0
	13	1.9	2.8	0	0	8.3
Parcel-Specific Variables						
Overall Winning Bid ($/acre)	11,400	179.56	584.79	0	12.00	5,500.00
	10,139	21.95	73.98	0	10.00	2,600.00
	10,295	23.07	43.97	0	10.00	525.00
Winning Bid for Collingwood Parcels ($/acre)	7,259	272.00	715.92	0	13.00	5,500.00
	6,175	25.02	88.56	0	13.00	2,600.00
	9,288	24.41	44.61	0	10.00	500.00
Winning Bid for non-Collingwood Parcels ($/acre)	4,141	17.53	42.39	0	10.00	1300.00
	3,964	17.17	41.75	0	10.00	1300.00
	1,007	10.67	35.24	0	0	525.00
Parcel Size (acres)	11,400	90.38	53.63	1	80	246
	10,139	90.10	54.14	1	80	246
	10,295	90.16	51.89	1	80	228
Proportion of Collingwood Parcels	11,400	0.64	0.48	0	1	1
	10,139	0.61	0.49	0	1	1
	10,295	0.90	0.30	0	1	1

7. Estimation Results

I begin this section by briefly restating my research design. In the first stage, using observations corresponding to a period during which I assume there was no collusion, I estimate the benchmark model in Section 4.1 to obtain $\hat{\theta}$. In the second stage, I treat $\hat{\theta}$ as the "true" value of θ during the collusive period. I then use the simulated method of moments to find the value of the parameter β of the collusive probability, described in Section 4.2, that "best" explains the observed winning bids over this period, assuming that any resulting coalition behaves according to Section 3.2. To this end, for each β, I simulate a sequence of auction sheets using $\hat{\theta}$ and the data to compute a set of "simulated moments". The "best" β is the value that minimizes the optimally weighted distance between the simulated and sample moments.

7.1 The First Stage: Maximum Likelihood Estimation (MLE)

The parameters of the benchmark model are estimated by maximizing equation (9), separately for gas and oil parcels, and only for auctions from May 2008 to May 2012 and from October 2012 to May 2018—periods assumed to have no collusion. The maximization procedure is done with MATLAB's *fminunc* function, with several different initial guesses being tried to ensure that the global maximum is achieved. The resulting $\hat{\theta}$ is reported in Table 7.1. All standard errors and 95% confidence intervals are based on 1,000 bootstraps, using the actual sample sizes for oil and gas parcels as the respective bootstrap sample sizes.[17]

[17] Although *fminunc* does provide a numerical estimate of the Hessian matrix, which theoretically allows us to directly compute the standard errors of the parameters, this estimate is known to often be inaccurate for non-smooth objective functions, of which mine is an example. For this reason, inference here is conducted using the bootstrap method.

Table 7.1. MLEs of the benchmark parameter θ, for gas and oil parcels.			
Variable	$\hat{\boldsymbol{\theta}}$	**Bootstrapped Standard Error**	**Bootstrapped 95% CI**
Gas Parcels (obs = 11,240)			
Constant	-5.0652	0.2089	[-5.5794, -4.7095]
Gas Price	0.0472	0.0011	[0.0438, 0.0482]
Parcel Size	-0.0005	0.0003	[-0.0011, 0.0001]
Non-Development	-0.1146	0.0346	[-0,1919, -0.0547]
Collingwood	0.4055	0.1762	[0.2154, 0.9306]
Hype	-0.0034	0.0108	[-0.0207, 0,0239]
Collingwood*Hype	0.0538	0.0109	[0.0257, 0,0707]
σ	3.9439	0.0761	[3.7667, 4.0689]
Oil Parcels (obs = 2,788)			
Constant	-7.9310	0.5453	[-8.8273, -6.6471]
Oil Price	0.0515	0.0036	[0.0443, 0.0585]
Parcel Size	0.0018	0.0006	[0.0006, 0.0031]
Non-Development	-0.0403	0.0837	[-0.1858, 0.1400]
Collingwood	0.8536	0.1051	[0.5697, 0.9791]
Hype	0.0131	0.0012	[0.0088, 0.0134]
Collingwood*Hype	0.0429	0.0017	[0.0417, 0.0481]
σ	3.7281	0.1717	[3.3271, 3.9957]

Most of the MLEs exhibit the expected sign. For example, resource prices and being in Collingwood significantly predict a higher valuation mean, while having a non-development classification lowers valuation due to the anticipated costs of addressing the associated restrictions. Likewise, adding together the coefficients for hype and the interaction term reveals that the hype measure increases the valuation mean for parcels in the Collingwood shale, as it should. Regarding parcel size, the coefficient estimate works in the expected direction for oil parcels but is not statistically different from zero for gas parcels. The latter may be a consequence of not having sufficiently controlled for the geographical location of the tracts beyond the Collingwood dummy, due to inadequate information in the data. For instance, a huge parcel offered in a remote, dry area will have a lower valuation relative to a smaller parcel in another region bustling with oil and gas activity. Finally, it is noteworthy that the number of oil parcels in this subsample is much lower than that of gas parcels, which partially explains the relatively larger standard errors of the latter's coefficient estimates.

7.2 The Second Stage: Simulated Method of Moments (SMM)

The second stage aims to estimate β, the parameter of the collusive probability $\pi_x(\beta) = \frac{e^{x\beta}}{1+e^{x\beta}}$, using the simulated method of moments proposed by McFadden (1989). The method proceeds as follows. Plugging in data from October 2010 to May 2012[18], and for any given value of β, I run 1,000 simulations of auction outcomes generated by the estimated $\hat{\theta}$ from Section 7.1. From these simulations, I compute a vector $\boldsymbol{m^{sim}}(\beta) \equiv (m_1^{sim}(\beta), \ldots, m_8^{sim}(\beta))'$ of six selected moments, summarized in Table 7.2 below.

Table 7.2. Selected moments for auction outcomes.
M1. Average winning bid for gas parcels
M2. Probability that the winning bid for gas parcels is less than the first reserve price
M3. Probability that the winning bid for gas parcels is equal to the first reserve price
M4. Average winning bid for oil parcels
M5. Probability that the winning bid for oil parcels is less than the first reserve price
M6. Probability that the winning bid for oil parcels is equal to the first reserve price

More specifically, denote by $m_j^s(\beta)$ the j^{th} moment in Table 7.2 from simulation s. The j^{th}*simulated population moment* $m_j^{sim}(\beta)$ is computed by averaging over all 1,000 simulations:

$$m_j^{sim}(\beta) = \frac{1}{1000}\sum_{s=1}^{1000} m_j^s(\beta)\,. \tag{10}$$

Furthermore, let m_j^{samp} be the corresponding j^{th} *sample moment*, computed as usual from the actual data set. Denote the vector of six sample moments by $\boldsymbol{m^{samp}} \equiv (m_1^{samp}, \ldots, m_6^{samp})'$. The simulated method of moment estimator $\hat{\beta}_{SMM}$ is then given by:

$$\hat{\beta}_{SMM} = argmin_\beta[\boldsymbol{m^{samp}} - \boldsymbol{m^{sim}}(\beta)]'\ \Omega_{opt}\ [\boldsymbol{m^{samp}} - \boldsymbol{m^{sim}}(\beta)], \tag{11}$$

where Ω_{opt} is a 6-by-6 optimal weight matrix. For a consistent estimate of Ω_{opt}, Gourieroux, Monfort and Renault (1993) proposes using the inverse of the variance-covariance matrix of the sample moments. Here, I estimate this variance-covariance matrix using 100,000 bootstrapped samples, whose sample sizes are set equal to the respective oil and gas sample sizes from the actual data.

Equation (11) is then solved using MATLAB's *fminsearch* function, again setting several different initial guesses to ensure that the global maximum is achieved. Table 7.3 reports the resulting SMM

[18] To reiterate what has been said in the introduction, this strategy of analysis assumes that if collusion did occur, then it did *not* happen before the Chesapeake and Encana exchanged the email; nor did it persist after Reuters' article was published in June 2012 and a federal investigation was subsequently launched.

estimates of the two components of β. Standard errors are calculated with the formula in DeBacker (2016), where the partial derivatives of the vector of moments are approximated numerically.

Table 7.3. SMM estimates of the parameter β of the collusive probability $\pi_x(\beta)$.		
Variable	$\boldsymbol{\hat{\beta}_{SMM}}$	**Standard Error**
Constant	2.2539	0.0846
Email_County	2.5233	0.0056

To interpret these coefficients, recall the parametrization $\pi_x(\beta) = \frac{e^{\beta_0+\beta_1+Email\_County}}{1+e^{\beta_0+\beta_1+Email\_County}}$. Plugging in the values from Table 7.3, for parcels in counties not included in the email between Chesapeake and Encana, the estimated collusive probability is $\frac{e^{2.2539}}{1+e^{2.5233}} = 0.905$. However, for those in counties mentioned in the email, the estimated collusive probability soars to $\frac{e^{2.2539}}{1+e^{2.2539+2.5233}} = 0.991$. This means that each bidder during the collusive period participates in the bidding ring with near certainty. Thus, the SMM estimates—*if my empirical assumptions going into the model are correct*—provide strong evidence for collusion during this period. Interestingly, the probability of colluding remains high even for parcels that do not belong to an "email county", suggesting either that collusion was widespread and likely not restricted to just the counties mentioned in the email, or that there is a bias in my estimation.

Indeed, one major concern regarding my approach lies in the lack of granularity in the data for the measure of potential bidders for each parcel, which is assumed to be uniform across all parcels in a given auction. In reality, there may be significant variation in the number of potential bidders each parcel attracts, and as acknowledged in Section 5.3, it is very likely that the chosen measure consistently overstates the true values. A consequence is that the estimated collusive probabilities above could be inflated. A simple thought experiment demonstrates the intuition for this: suppose the number of potential bidders for a set of observations according to the chosen measure is 22 (this is the actual average for each parcel over the collusive period), but the true (unknown) value is 4. Now suppose the collusive probability which "best" fits the data by moment matching is one that forces there to remain a single effective bidder for each parcel, who then wins it at the reserve price. Clearly, the collusive probability required to ensure these outcomes is going to be much lower if there are 4 bidders in the pool of potential bidders than if there are 22, since fewer bidders in total are needed in the former case to be in the coalition to induce transaction at the reserve price. In other words, if the data suggest the presence of an all-inclusive bidding ring, getting 4 bidders to collude requires a lower collusive probability than does having all 22 be in the ring. Thus, given the limited available data, the evidence here for collusion should not be interpreted as conclusive. Regardless, these results could still be valuable as a comparison point against future research using alternative approaches and/or better data. In part A of the Appendix, I confirm the findings' sensitivity to the number of potential bidders by repeating the second stage with the number of potential bidders fixed at a lower value throughout the collusive period—and finding the subsequent estimated collusive probabilities to be smaller than those reported above.

Nonetheless, assuming that my chosen measure of potential bidders does not differ too much from the true numbers, Table 7.4 displays the values of the six moments generated by the model and $\hat{\beta}_{SMM}$, averaged over 1,000 simulations, alongside their empirical counterparts calculated from the data. The SMM estimates overpredict the average winning bids and the proportions of parcels sold at the first reserve price, but underpredict the proportions of parcels that fail to sell at the first reserve price. However, these differences do not appear outrageously large, suggesting that the evidence for collusion provided by the estimates may be sufficiently reasonable in terms of how well the induced moments match the actual data.

Table 7.4. Values of simulated and sample moments.

Moments	**Data**	**Model**
Average winning bid for gas parcels	14.46	23.70
Probability that the winning bid for gas parcels is less than the first reserve price	0.452	0.317
Probability that the winning bid for gas parcels is equal to the first reserve price	0.439	0.523
Average winning bid for oil parcels	33.17	39.66
Probability that the winning bid for oil parcels is less than the first reserve price	0.362	0.282
Probability that the winning bid for oil parcels is equal to the first reserve price	0.458	0.642

Of course, aside from the empirical issues raised above about the measure of potential bidders, these results are also based on several arguably restrictive modeling assumptions. Therefore, going forward, it may be worthwhile to explore more flexible modeling choices and see whether the empirical evidence for collusion still holds. For example, I could develop a model that accounts for entry costs, bidder asymmetry and unobserved heterogeneity. Endogenizing the entry costs would also potentially allow me to achieve a more conservative measure of potential bidders, thereby partially correcting for the problem I have with my current measure. On the data side, my analysis would benefit greatly from more precise classifications of oil and gas and Collingwood parcels, finer geographical partitions than just the Collingwood designation to reflect "economically desirable" regions, and of course, a more accurate measure of the number of potential bidders. If given more detailed data about the bids, I might also be able to deviate from the IPV framework, incorporate a common values component in the bidder valuations a la Bhattacharya et al. (2018), and compare between the two modeling paradigms. Lastly, it is worth checking how my conclusions compare to an alternative research design, as mine relies on a rather strong premise about which auctions are perfectly competitive and which ones are suspects for collusion.

7.3 Brief Empirical Justification for the Research Design

As a rough sanity check that my research design is justifiable, I split my non-collusive period into two parts: the May 2008 auction (the first in my data set chronologically), and the rest of the auctions which are assumed to be competitive. I then run the first-stage procedure over the latter set of auctions before applying the resulting estimates to solve for the collusive probability in May 2008 via the second stage. The idea is that since there is no reason to suspect collusion in May 2008, the second-stage estimates should yield relatively low collusive probabilities—if my research design is indeed sensible.

Table 7.5 and Table 7.6 display, respectively, the first-stage MLE and second-stage SMM estimates for the parameters for May 2008, solved according to the procedure above. The signs of first-stage MLEs still work in the expected directions after removing May 2008 from the benchmark subsample. Moreover, the second-stage results suggest that in May 2008, for counties not mentioned in the email, each bidder colludes with probability $\frac{e^{-3.3425}}{1+e^{-3.3425}} = 0.0341$, which is essentially the same as the collusive probability for counties mentioned in the email due to the negligible coefficient for the *Email_County* variable. That the coefficients are small and identical is indeed not surprising because the email about splitting the counties to suppress competition was not written until the summer of 2010. In other words, my approach estimates that for any bidder in the May 2008 auction, there is approximately a 3% chance that he colludes for *any* given parcel. Therefore, collusion seems relatively unlikely in May 2008, which is consistent with my expectations of no collusion and a research design that “makes sense”.

Table 7.5. First-stage MLEs of the benchmark parameter θ, excluding May 2008.	
Variable	$\hat{\boldsymbol{\theta}}$
Gas Parcels (obs = 9,896)	
Constant	-8.5604
Gas Price	0.0618
Parcel Size	0.0016
Non-Development	-0.1144
Collingwood	0.8334
Hype	0.0112
Collingwood*Hype	0.0449
σ	3.7257
Oil Parcels (obs = 2,741)	
Constant	-5.1026
Oil Price	0.0213
Parcel Size	-0.0005
Non-Development	-0.1536
Collingwood	0.8088
Hype	0.0107
Collingwood*Hype	0.0401
σ	4.1282

Table 7.6. Second-stage SMM estimates of the parameter β of the collusive probability for May 2008.	
Variable	$\hat{\boldsymbol{\beta}}_{\boldsymbol{SMM}}$
Constant	-3.3425
Email_County	-0.0004

8. Counterfactuals

The evidence for collusive behavior in Section 7.2 naturally begs the question: how much revenue would have been generated had there been no collusion? Put differently, what would the simulated auction outcomes look like under perfect competition?

To examine this counterfactual scenario, I repeat the simulation procedure described in the first paragraph of section 7.2, fixing the collusive probability π_x at zero to produce a competitive setting. Then, for each simulation s, I compute the mean and median winning bid, separately for oil and gas parcels. Finally, I calculate the *average simulated competitive mean winning bid* and the *average simulated competitive median winning bid* by taking the averages of those means and medians over all 1,000 simulations, respectively. Table 8.1 displays the counterfactual simulation outcomes over the collusive period, juxtaposed with their sample counterparts calculated from the actual data. Note that no simulation is run for the May 2010 auction since it is not contained in the collusive period.

Table 8.1. Simulated competitive auction outcomes versus actual outcomes.

Auction	**Actual Mean Bid**	**Average Simulated Competitive Mean Winning Bid**	**Actual Median Bid**	**Average Simulated Competitive Median Winning Bid**	**# Parcels**	**Total Acres Offered**
Gas Parcels						
May 2010	1323.72	N/A	1,350.00	N/A	833	71710
October 2010	15.34	574.01	13.00	136.70	1,818	207070
May 2011	14.87	46.65	13.00	13.00	133	10596
October 2011	8.76	17.86	0.00	11.36	990	97417
May 2012	20.90	21.83	12.00	12.00	619	46628
Oil Parcels						
May 2010	1686.48	N/A	2025.00	N/A	428	45060
October 2010	36.24	1487.30	13.00	417.07	2,785	235591
May 2011	12.02	82.36	13.00	14.03	347	24511
October 2011	14.82	26.69	13.00	12.97	219	18911
May 2012	36.85	37.46	12.00	12.00	759	60967

Assuming that collusion *did* in fact happen, the counterfactual results indicate that much revenue damage was incurred by bidders' anticompetitive behavior. Most strikingly, in the October 2010 auction, the actual averages lease prices of gas and oil parcels were about 40 times lower than their predicted competitive values. In all the other auctions during the collusive period, the average winning bids under the imposition of no collusion are still greater than the observed means, though by not as much as in October 2010. Overall, the total revenue loss over the period due to collusion amounts to over \$450 million—\$16,521,252 collected in real life by the state of Michigan versus the \$471,723,305 generated by the competitive simulations. In particular, the October 2010 auction was responsible for 99% of this discrepancy. This is consistent with Figure 1.3, which shows that the hype around the Collingwood shale remained strong in the months leading up to October 2010 before subsequently dropping to lower levels. This means that valuations must have remained high during this auction, contrary to the observed depressed winning bids. Therefore, the incurred revenue damage for Michigan might have been tremendous—especially for the auction immediately following the email exchange.

As a final note, the average simulated competitive winning bids are computed using the simplified auction rules described in Section 4.2, according to which there is only the first reserve price and no subsequent reoffer stage, whereas the observed winning bids include revenues gained from selling the reoffered parcels. In other words, in my simulations, the auction ends at the first round if no bid above the first reserve price is placed, whereas the chance of reoffering in the actual auctions makes it possible for the auctioneer to sell the parcel in the reoffer stage, even when the first stage fails.[19] Therefore, the damage calculated above is actually an *underestimate*. That is, had I incorporated the possibility of reoffering the failed parcels in my simulations, the simulated mean winning bids would have been even higher—and larger would have been the gap between the simulated competitive revenues and their sample counterparts.

9. Conclusion

This paper employs a structural model of bidding to examine an alleged incident of collusion in the Michigan public oil and gas lease auctions between 2010 and 2012. The empirical investigation proceeded in two stages. In the first stage, I use data from a period during which bidding was likely competitive to estimate, via maximum likelihood, the parameters for various determinants of bidders' valuations. In the second stage, I restrict my attention to the alleged collusive period. I incorporate into the model a "collusive parameter", treat as given the estimates from the first stage as the parameter values for the valuation determinants in the extended model, and estimate the collusive parameter using the simulated method of moments. My analysis reveals strong empirical evidence for bid-rigging during the collusive period. Finally, I conduct counterfactual analysis to find that collusion induced over $450 million in revenue damage for the state of Michigan, a large portion of which was associated with the October 2010 auction immediately following the Chesapeake-Encana email correspondence.

My findings could potentially have important policy implications. Given the massive revenue loss predicted above, the Michigan Department of Natural Resources may deem it worthwhile to invest in anti-collusive measures. The vulnerability of ascending-bid or second-price auctions to bidder collusion is widely acknowledged in the literature (Marshall & Marx, 2009). Accordingly, various counteractive auction designs have been suggested to inhibit or mitigate the impact of collusion by encouraging deviation from bidding rings. These include non-transparent registration, suppression of the identities of

[19] As an example, suppose I have a parcel that fails to sell at the first reserve price—at $10 per acre. In my simplified model where there is no reoffer round, the auction ends there, and the seller receives zero in revenue. But in reality, the parcel *could* get reoffered at the reduced reserve price of $2 per acre, and someone may wind up buying it at $7 per acre. In this case, my simulations, which use the simplified auction rules, *underestimate* the revenue the seller could have received ($0 versus $7).

active and winning bidders (Marshall & Marx, 2009), utilization of the less collusion-susceptible first-price sealed-bid format (Marshall & Marx, 2007), and implementation of an optimal rather than static reserve price to minimize the seller's loss (Graham & Marshall, 1987).

However, the evidence for collusion in this paper should be considered with caution. As aforementioned, my modeling approach and the data I have available—particularly the measure of potential bidders—are not without flaws. Thus, additional comparisons, using alternative research designs and better parcel-level data about the participating bidders, are needed to warrant increased confidence in the accuracy of my results.

References

Athey, S. & Haile, P. A. (2002), 'Identification of Standard Auction Models', *Econometrica* **70**(6), 2107-2140.

Baldwin, L., Marshall, R. & Richard, J-F. (1997), 'Bidder Collusion at Forest Service Timber Sales', *Journal of Political Economy* **105**(4), 657-699.

Banas, Ryan M. (2012), 'Geophysical and geological analysis of the Collingwood Member of the Trenton Formation', *Master's Thesis, Michigan Technological University*.

Bhattacharya, V., Ordin, A. & Roberts, J. W. (2018), 'Bidding and Drilling Under Uncertainty: An Empirical Analysis of Contingent Payment Auctions', *Working Paper*.

Cramton, P. & Schwartz, J. A. (2000), 'Collusive Bidding: Lessons from the FCC Spectrum Auctions', *Journal of Regulatory Economics* **17**, 229-252.

DeBacker, J. (2016). *Econ 7130 – Microeconomics. Spring 2016. Notes for Lecture #10*. Retrieved from https://www.jasondebacker.com/classes/Lecture10_Notes_SMM.pdf.

DeMarzo, P. M., Kremer, I. & Skrzypacz, A. (2005), 'Bidding with Securities: Auctions and Security Design', *American Economic Review* **95**(4), 936-959.

Gourierroux, C., Monfort, A. & Renault, E. (1993), 'Indirect Inference', *Journal of Applied Econometrics* **8**, S85-S118.

Graham, D. & Marshall, R. (1987), 'Collusive Bidder Behavior at Single-Object Second-Price and English Auctions', *Journal of Political Economy* **95**(6), 1217-1239.

Haile, P. A., Hendricks, K. & Porter, R. H. (2010), 'Recent U.S. Offshore Oil and Gas Lease Bidding: A Progress Report', *International Journal of Industrial Organization* **28**(4), 390-396.

Hendricks, K. & Porter, R. H. (2007), 'An Empirical Perspective on Auctions', *Handbook of Industrial Organization* **32**(3), 2073-2143.

Hendricks, K., Porter, R. & Tan, G. (2008), 'Bidding ring and the winner's curse', *The RAND Journal of Economics* **39**(4), 1018-1041.

Kong, Y. (2017a), 'Selective Entry in Auctions: Estimation and Evidence', *Working Paper*.

Marshall, R. C. & Marx, L. M. (2007), 'Bidder collusion', *Journal of Economic Theory* **133**(1), 374-402.

Marshall, R. C. & Marx, L. M. (2009), 'The Vulnerability of Auctions to Bidder Collusion', *The Quarterly Journal of Economics* **124**(2), 883-910.

McFadden, D. (1989), 'A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration', *Econometrica* **57**(5), 995-1026.

Appendix A: The Impact of the Number of Potential Bidders on the Estimated Collusive Probability

In the following sensitivity analysis, I take the parameter estimates for the determinants of bidder valuations computed in the first stage and rerun the second stage, fixing the number of potential bidders at 4 for every parcel offered during the collusive period. This is in lieu of using the chosen measure of potential bidders, the mean of which over that period is around 22. Since I am not aware of any economic research that has hitherto been done on the Michigan oil and gas lease auctions, the number 4 is selected from browsing previous empirical research on oil and gas auctions held in other states. In particular, in their analysis of the New Mexico State Land Office's oil auctions, Bhattarachya et al. (2018) estimate the average number of potential bidders in their sample to be 3.87.

Table A1 displays the SMM estimates of the parameters of the collusive probability using this alternative measure. As hypothesized, my results are incredibly sensitive to changes in the number of potential bidders. According to the estimates, the probability that a given bidder decides to collude for parcels not mentioned in the email is $\frac{e^{-2.3905}}{1+e^{-2.3905}} = 0.0839$, whereas this probability increases to $\frac{e^{-2.3905+2.5662}}{1+e^{-2.3905+2.5662}} = 0.5438$ for counties that get mentioned in the email. While these results still support the hypothesis of collusion being likely in the suspect counties, the values of the estimates are drastically lower than those reported in my main results. Therefore, the credibility of my findings depends heavily on how accurate my chosen measure of the number of potential bidders is. It is for this reason that I am hesitant to consider my evidence for collusion conclusive.

Table A1. SMM estimates of the parameter β of the collusive probability $\pi_x(\beta)$, fixing *n* = 4.

Variable	$\hat{\boldsymbol{\beta}}_{SMM}$
Constant	-2.3905
Email_County	2.5662