Firm Boundaries and External Costs in Shale Gas Production

Matthew O'Keefe [*†]

May 2, 2024

Abstract

Wastewater reuse in the shale gas industry reduces firms' private costs and mitigates many of the local environmental harms associated with fracking. Most reuse occurs within the firm boundary, but rival operators often exchange (or "share") wastewater prior to reuse. This paper considers how shale gas producers in Pennsylvania choose between internal reuse and sharing, and whether additional sharing would be environmentally beneficial. To quantify the costs of sharing, I build a novel empirical model of firms' wastewater management decisions. Estimating the model, I find that transaction costs associated with sharing are large — increasing total water-related costs by roughly 10% — but heterogeneous. Variation in the estimates reveals several channels for potential policy interventions to facilitate further sharing. However, counterfactual exercises suggest that such interventions may have limited environmental benefits.

1 Introduction

During fracking, millions of gallons of water are mixed with sand and chemicals and injected into the earth in the course of a few weeks. After a new well is completed, much of this water returns to the surface as wastewater, over a period that can last several years. Given the scale of fracking activity in the United States and the long-term nature of wastewater production, the transportation and disposal of fracking wastewater pose several challenges for environmental regulation. Conventional disposal wells permanently remove water from

[*] Northwestern University. Email: mfokeefe@u.northwestern.edu

[†] I am grateful for the generous support and guidance of Rob Porter, Mar Reguant, and Vivek Bhattacharya. I also thank Igal Hendel, Thomas Hubbard, Gaston Illanes, Ryan Kellogg, Gaston Lopez, Alex MacKay, Ameet Morjaria, Francisco Pareschi, Michael Powell, Bill Rogerson, Jin Yang, participants in the Northwestern IO Lunch, and participants in the 2023 IIOC Rising Stars Session for valuable feedback. Several individuals at the Pennsylvania Department of Environmental Protection and Susquehanna River Basin Commission provided helpful background on wastewater management. Any errors are my own.

the hydrologic cycle, and may contribute to drinking water contamination and localized seismic activity (Groundwater Protection Council, 2019). Moreover, wastewater is frequently transported by heavy water-hauling trucks, generating air pollution and greenhouse gas emissions while creating significant spill risk (EPA, 2016). Given these challenges, as well as intensive freshwater usage, state regulators typically encourage oil and gas producers to *reuse* wastewater from existing wells in place of freshwater when fracking new wells. Wastewater reuse in subsequent fracking reduces final disposal volumes, limits freshwater usage, and (in many cases) mitigates transportation-related environmental harms.

Wastewater reuse can occur either within the boundary of firm or through exchange between rival firms. Even in the absence of policy intervention, principal (or operating) firms often reuse wastewater in order to economize on wastewater disposal and freshwater acquisition costs. Exchange between rival firms — known as *sharing* — can further reduce firms' private costs in some circumstances. Because the volume and location of wastewater generation in the current period is determined by past drilling decisions, while new completions occur intermittently and are prone to delays, sharing can resolve temporary supply-and-demand mismatches, and can facilitate the realization of transportation, among other potential benefits. From a regulatory perspective, sharing can generate positive environmental spillovers both by increasing the extent of reuse and by facilitating reuse transactions that reduce transportation-related harms.

In this chapter, I analyze the wastewater sharing market in Pennsylvania, where 90% of wastewater is reused due to high conventional disposal costs, and 10% of wastewater is shared prior to reuse. Since many firms rely on a mix of internal reuse and sharing, the setting provides a rare opportunity to examine empirically how firms choose between insourcing and outsourcing (here, internal reuse and reuse via sharing). This enables me to provide new evidence on Coasean transaction costs (Coase, 1937) and a novel perspective on the role of firm structure in mediating environmental externalities. In the main part of the analysis I quantify the distribution *sharing frictions,* defined as transaction costs specific to sharing transactions, and assess their sources. Then, using the same empirical framework, I explore the equilibrium implications of sharing frictions for environmental outcomes and policy.

The primary data source for this study are wastewater disposal records maintained by the Pennsylvania Department of Environmental Protection (DEP). Using this data, I first establish that sharing frictions appear to play an important role in shaping firms' wastewater management decisions. This can be seen clearly in the context of the 2017 merger of EQT Corporation and Rice Energy, Inc., which created the largest natural gas producer in the United States. EQT and Rice never shared wastewater prior to merging despite clear geographic complementarities that were exploited by the merged entity immediately after consummation, suggesting that sharing frictions had been large ex ante.

While the merger evidence suggests that sharing frictions can be large, the prevalence of sharing among other pairs of firms suggest that they are often smaller, or that joint cost savings are not wholly attributable to distance-related transportation costs. To explore the latter possibility, I consider the extent to which observed shipment patterns are rationalized by differences in distance alone. Using a simple optimal transport model, I show that within-firm shipments realize only about 56% of distance-related transportation cost savings relative to random matching. Thus, other intrinsic costs of reuse that are not directly related to distance (such as treatment or labor costs) also appear to play an important role in determining patterns of reuse.

In order to empirically disentangle sharing frictions from other costs of reuse, I develop an empirical model of matching with transferable utility that captures key features of firms' wastewater management decisions. Firms simultaneously minimize wastewater disposal and water sourcing costs. Both the sharing market and firms' internal markets for wastewater are inherently two-sided: previously completed wells generate only a limited volume of wastewater in each period, while the capacity for reuse in completing a single well is constrained by engineering considerations. In this context, transactions in the sharing market can crowd out opportunities for internal reuse, and vice versa. I capture these dynamics by assuming that firms make disposal and reuse decisions on a truckload-by-truckload basis in a decentralized matching environment. In a matching game, truckloads of wastewater are matched to specific locations where disposal or reuse can occur. Stability of the match implies that all firms minimize total costs simultaneously.

In the model, the joint costs of reuse for a given truckload of wastewater consist of distance-related costs, other intrinsic costs (whether observed or latent) such as treatment costs, and sharing frictions incurred only when reuse occurs across the firm boundary. The sum of these costs is identified conditional on the distribution of the latent costs, as demonstrated by Choo and Siow (2006) and Galichon and Salanie (2022) in the context of marriage markets. Thus, shipments within firms' boundaries identify the intrinsic costs of reuse, while sharing frictions are identified by the difference between the intrinsic costs of sharing transactions and the shadow costs that rationalize market-wide matching patterns. This strategy is similar to that of Atalay et al. (2019), who identify the "net benefits of ownership" under vertical integration from establishment-level shipment data.

I estimate the model using a maximum likelihood estimator similar to the one proposed by Galichon and Salanie (2022). The mean estimated sharing frictions is equivalent to a cost of about \$6 per barrel, about three times as large as distance-related transportation costs for the average sharing transaction. This estimate is large, similar to prior results on the magnitude of distortions at the firm boundary (Masten et al., 1991; Atalay et al., 2019).

The estimates reveal significant heterogeneity in sharing frictions across observably similar transactions: the standard deviation in sharing frictions across transactions is almost \$3 per barrel. Contracting frictions appear to be an important source of sharing frictions. The estimates suggest a few potential mechanisms: for instance, sharing frictions are greater when inter-operator liability concerns and risks to well productivity are greater. Only a few pairs of firms share at close-to-integrated rates under the status quo, suggesting that firms are rarely able to circumvent sharing frictions through formal contracting or relationships. It follows that improvements in the contracting environment (such as clarification of liability rules) may be required to significantly improve the performance of sharing markets.[1,2]

In Section 7, I consider how the presence of sharing frictions mediates the external costs, or negative environmental externalities, created by wastewater management. I focus on wastewater trucking.[3] Wastewater management in Pennsylvania requires about 500,000 truck trips each year, at external costs of around \$7M. Counterfactual simulations suggest that sharing frictions *reduce* these costs by 13%, contrary to what might be expected. This occurs for two reasons: first, firms tend to operate in specific, circumscribed geographies, so that transactions within the firm are typically nearer in distance than those between firms. Second, firm boundaries inhibit matching on all components of the joint costs of reuse simultaneously, not just distance-related transportation costs. Under the status quo, the non-randomness of firm boundaries tends to prevent sorting on non-transportation costs at longer distances, softening the adverse effects of the underlying misalignment between private and social costs. However, if the cost of distance represented a larger share of private costs or if the distribution of firms' operations were more evenly distributed geographically, firm boundaries would tend to increase external costs instead. An implication is that market design interventions intended to make sharing easier may have limited net benefits, and could even be counterproductive.

Section 8 derives and analyzes the Pigouvian allocation, which minimizes the sum of private and external costs. The optimal allocation depends on the welfare-relevance of sharing frictions. If sharing frictions are welfare-irrelevant, the social planner optimally incentivizes participation in sharing market. One way to implement the optimal allocation is with tar-

[1] Recent legislation in Oklahoma created a liability shield for firms that process and transport wastewater for reuse. Moreover, this legislation included other measures that effectively reduced the transaction costs of sharing, such as clarifying that wastewater is the property of the operator rather than the landowner.

[2] Conversely, if search frictions are small relative to contracting frictions, interventions that target search frictions alone (e.g., the creation of a digital platform for wastewater sharing) may have limited impact.

[3] I focus on transportation-related externalities because reuse rates are already high in Pennsylvania, and because the data is better suited to analyzing the elasticity of transportation than the elasticity of reuse in general. However, the empirical model I develop can accommodate these margins.

geted sharing subsidies. Potential welfare gains are large: subsidies can reduce social costs by up to \$0.72 per barrel, or 64% of distance-related trucking costs for the average shipment. However, these potential welfare gains are entirely attributable to producer cost savings; trucking-related external costs *increase* by 12.6%. Absent large increases in reuse on the extensive margin, such a policy might be difficult to implement in practice.

In reality, it may be difficult for a regulator to distinguish between sharing fractions that are welfare-relevant (e.g., search and negotiation costs) and those that are not (e.g., managerial inattention).[4] Welfare losses from poorly calibrated policies can be substantial: if the social planner offers excessive subsidies due to a misinterpretation of the sharing frictions, social costs increase by up to \$2.43 per barrel before considering the cost of public funds. I conclude by considering how Pigouvian regulation can be made robust when welfare-relevance is ambiguous.

The remainder of Section 1 clarifies the relationship between this paper and prior work. Section 2 describes the setting. Section 3 presents motivating evidence on sharing frictions. Section 4 introduces the wastewater management model and Section 5 discusses identification and estimation. Section 6 describes the model estimates, focusing on the estimated sharing frictions and their interpretation. Section 7 discusses the relationship between sharing frictions and external costs. Section 8 discusses Pigouvian regulation. Section 9 concludes.

1.1 Related literature

This work most directly contributes to the policy literature on the local environmental impacts of fracking. In economics, Hausman and Kellogg (2015) and Black et al. (2021) survey local environmental issues as well as broader environmental and economic considerations.[5] Groundwater Protection Council (2019) reviews the current legal and regulatory frameworks applicable to wastewater management, including reuse within the oil and gas industry. I contribute to this literature by developing a structural framework that can be used to prospectively evaluate wastewater regulation.

In this chapter, sharing frictions are the transaction costs associated with wastewater sharing. In this way, sharing frictions are a special case of the transaction costs that govern the "make-vs-buy" decision analyzed by Coase (1937) and subsequently formalized in transaction cost economics (Williamson, 1971), property rights theory (Grossman and Hart,

[4]The distinction between welfare-relevant and welfare-irrelevant costs is most commonly encountered in the literature on switching costs in consumer markets (e.g., Handel and Kolstad, 2015). Previously, Buchanan and Stubblebine (1962) distinguished between Pareto-relevant and Pareto-irrelevant externalities.

[5]An important but distinct issue is whether the shale boom has increased or decreased global greenhouse gas emissions. See, e.g., Newell and Raimi (2014). For simplicity I do not consider the elasticity of drilling with respect to wastewater management costs, although this is an interesting avenue for future research.

1986), and elsewhere. I build on the existing transaction costs literature by embedding the make-vs-buy decision in a structural model of industry costs.

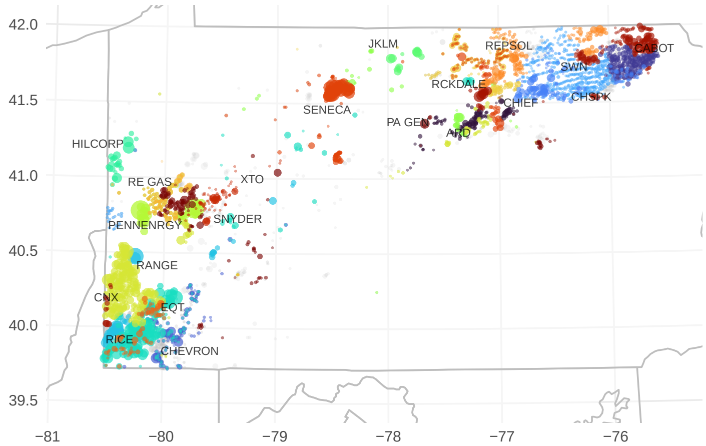
Few papers in empirical industrial organization have explicitly incorporated Coasean transaction costs, in large part due to the general scarcity of data on firms' internal operations. My empirical analysis is only possible because I observe transactions within- and between firms in the same narrowly defined market, which enables me to empirically separate production costs from transaction costs.[6] Demsetz (1988) emphasizes the potential heterogeneity in transaction costs within narrowly defined markets; although my analysis is limited in scope, my approach accounts for this richness to a greater extent than the closely related approach of Atalay et al. (2019), who use Census data to quantify the "net benefits of ownership" across a wide range of industries. I build on their work by obtaining more granular estimates of transaction costs and by incorporating transaction costs into a welfare analysis. Masten et al. (1991) quantify transaction costs using a selection model and relate their estimates to transaction characteristics. In comparison, I focus on heterogeneity within a market rather than across markets.

The Coasean view of the firm is complementary to the strategic view of the firm typically encountered in industrial organization (Bresnahan and Levin, 2012). In this way, the analysis in this chapter complements recent empirical work at the intersection of industrial organization and environmental economics that highlights the challenges of environmental regulation when firms act strategically (e.g., Mansur, 2007; Fowlie, 2009; Leslie, 2018; Preonas, 2023). In particular, evidence that transaction costs are large can help explain why some firms attain sufficient scale for strategic distortions to be significant.

Finally, this paper relates to a variety of recent papers in empirical industrial organization that study upstream frictions in the oil and gas industry. Kellogg (2011) and Covert (2015) study learning-by-doing spillovers within relationships and across the firm boundary, respectively. Vreugdenhil (2023) studies search and matching frictions in subcontracting with rig operators. Covert and Sweeney (2022) study information frictions relating to well productivity, also using data from Pennsylvania. I differ in focusing on the external costs that result from upstream frictions. My policy analysis is most closely related to that of Covert and Kellogg (2023), who study railroad transport of crude oil from the Bakken shale.

[6] Border costs in trade are closely related to sharing frictions in my framework (Anderson and van Wincoop, 2004; Head and Meyer, 2014); economic activity within the firm can be analogized to economic activity within a country. Note that many papers quantify the costs of market transactions, which are related to but conceptually distinct from the object of interest in my analysis (the sum of all costs incurred when a transaction occurs through exchange rather than internally). In recent work, MacKay (2022) and Hodgson (2022) estimate market transaction costs in comparison to the costs of inaction in long-term contracting and durable goods markets; in these environments, inaction can be interpreted as a form of insourcing.

Figure 1: Well pad locations for the twenty largest firms



2 Setting

In this section, I briefly describe the economics of wastewater reuse in Pennsylvania. Then I introduce the data, focusing on participation in the *sharing market*.

2.1 Wastewater reuse and sharing in Pennsylvania

Pennsylvania produces more natural gas than any state besides Texas and accounts for about 20% of total US natural gas production. This relatively recent development can be attributed to improvements in so-called "unconventional" drilling techniques, most notably horizontal drilling and hydraulic fracturing ("fracking"), which have enabled exploitation of the vast Marcellus and Utica shales. Oil and gas production in Pennsylvania is conducted by numerous operating firms ranging from small, independent firms operating only a few wells to the largest global energy firms (Small et al., 2014). Figure 1 shows the locations of well pads operated by each of the twenty largest operators (by disposal volume) in the period that I study. The clustering visible in the figure reflects economies of density in permitting, exploration, drilling, and marketing, as well as freshwater and wastewater management, which I discuss in this section.

The process of fracking is water intensive. A typical completion requires over a hundred thousand barrels of water (more than five million gallons), with longer wells requiring more. During the fracking process, water is blended with sand and various chemicals. Underground, the fracking fluid becomes mixed with minerals and pre-existing groundwater. After completion, a large proportion of this fluid returns to the surface as wastewater, commonly known as flowback or produced water. Wastewater production continues for the life of a well, in steadily diminishing volumes. Much like with hydrocarbons, the amount of wastewater that a given well will produce is difficult to predict, but typically amounts to around 50% of injected volume over the lifetime of a Marcellus well.[7]

In Pennsylvania, wastewater *reuse* creates significant surplus for operators because conventional disposal is costly. Wastewater is highly saline and may contain organic compounds, metals, and naturally occurring radioactive materials; consequently, federal regulations require careful handling and specialized disposal (Groundwater Protection Council, 2019). *Injection disposal*, which involves using specialized wells to inject wastewater deep below the surface of the earth, is the conventional method of wastewater disposal in the oil and gas industry. In Pennsylvania and West Virginia the underlying geology is not well suited to drilling injection wells (McCurdy, 2011). Injection wells are common in Ohio, but the distance between Pennsylvania gas wells and Ohio injection wells can be significant. This is illustrated in Figure 2, which shows the location of active injection wells relative to active gas wells in the data. The costs of injection disposal can be substantial: trucking costs to Ohio disposal wells are $2-3 per barrel for producing wells in southwestern Pennsylvania and $10-11 per barrel for producing wells in northeastern Pennsylvania, before disposal fees of $2-4 per barrel (Menefee and Ellis, 2020). The costs of wastewater reuse are small by comparison: only a limited amount of chemical treatment and filtering is needed, at a cost of around $0.25-0.50 per barrel or less, and trucking costs are often much lower.[8] Reuse also reduces the need for freshwater, which would otherwise need to be acquired and transported from local sources at a typical cost of around $2 per barrel. Note that reuse outside the oil and gas industry is extremely limited.[9]

Treatment prior to reuse can occur either directly on a well pad or at a centralized treatment facility (CTF). Treatment on a well pad is more prevalent than treatment at a CTF, but both are common (I provide market shares in Section 2.3). Some CTFs are operated by oil and gas producers, and others by third party treatment firms. Producer-affiliated CTFs are often little more than semi-permanent systems of tanks or impoundments where the same treatments conducted on a well pad can be conducted at a larger scale. Third party CTFs are constructed similarly but may also have technologies that can treat water to higher standards, although these technologies are rarely used in practice.[10]

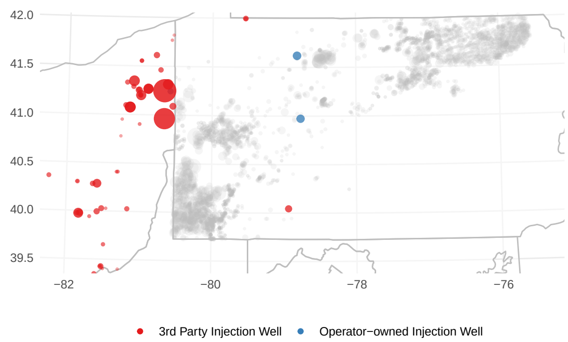
[7] The Marcellus and Utica shales (the main formations underlying Appalachia) are considered "dry" in the sense that relatively little water returns to the surface. In other regions, wastewater generation can be an order of magnitude larger (Kondash et al., 2018), substantially changing the economics of reuse.

[8] Pipeline transportation is also possible, both for disposal and reuse, but this is reportedly rare in the Appalachian Basin, perhaps due to challenging terrain (Groundwater Protection Council, 2019). According to the DEP, rail is occasionally used instead of truck prior to disposal.

[9] This primarily reflects a substantial difference in treatment requirements for reuse in fracking and reuse in other applications, as well as transportation costs (water being heavy).

[10] In practice, the choice between CTF and on-pad treatment primarily turns on a tradeoff between

Figure 2: Injection well locations



At any point in time, large firms have a stock of hundreds of completed wells producing gas and wastewater, in comparison to a handful of ongoing completions. Wells that are currently being completed by a firm may not be located near wells previously completed by that firm. Thus, a firm's demand for wastewater disposal at a particular point in time might be larger or smaller than its capacity for reuse, and the transportation costs associated with reuse inside the firm might be significant. In this context, wastewater *sharing* enables firms to reuse wastewater more efficiently, by resolving temporary supply and demand imbalances and exploiting transportation synergies. A wide range of prices is possible. I do not observe sharing prices, but anecdotally the sending firm may pay the receiving firm a "tipping fee" on the order of $1-3 per barrel in addition to paying for transportation. However, negative prices are possible (for example, if freshwater is scarce), and there is often no charge.[11]

2.2 Data

Wastewater disposal in Pennsylvania is regulated by the Pennsylvania Department of Environmental Protection (DEP). The DEP requires oil and gas operators to submit monthly reports indicating the disposal method and destination of all quantities of waste materials leaving every well pad, including each barrel of wastewater. These reports are publicly accessible on the DEP website and constitute the primary data source for this chapter. The

economies of scale and transportation costs. Regardless of ownership, the use of CTFs can increase transportation costs because wastewater must be transported twice – once to the CTF, and then again to a location where it can be reused. There are also differences in regulatory compliance costs that factor into this decision, such as differences in bonding requirements.

[11]One explanation for the prevalence of barter is that firms may seek to avoid a commercial designation for wastewater exchanges (Groundwater Protection Council, 2019).

reports clearly indicate whether a transfer was intended for reuse and, if so, provides further information identifying the destination facility. I use this information to distinguish internal reuse from sharing, based on whether the destination facility was associated with the sender.

The data have a few limitations worth highlighting. First, the data do not include prices, contract terms, or other details of the circumstances under which a shipment was intermediated.[12] This limits my ability to identify contractual mechanisms that contribute to sharing frictions. Furthermore, only the total volume of water transferred between two locations during a month is recorded, rather than the dates, modes, or volumes of particular shipments. Due to this limitation, I abstract from timing within months, mode choice, and less-than-full truckloads. In order to mitigate the impact of unobserved timing constraints within a month, I exclude from the estimation sample any shipments in month t leaving well pads that also received shipments of wastewater in month t.[13] Finally, the data do not indicate locations at which treatment processes occurred, or if these occurred in different stages at different locations. The location of reuse is not available prior to 2017, so I focus on the period from 2017 to 2020.[14] The data cleaning process is discussed in further detail in Appendix 10.

2.2.1 Centralized treatment facilities

One limitation of the data requires special treatment throughout the analysis. Specifically, the data records shipments to but not from CTFs. Lack of data on shipments from CTFs creates several empirical challenges. First, it is not possible to perfectly distinguish internal reuse from sharing in the case that wastewater is initially transferred to a CTF. Second, the data do not reveal how firms that accept wastewater substitute between direct shipments of wastewater and re-shipments from CTFs (because, for any facility, some unknown volume might have been received from CTFs). Third, I cannot directly calculate the substantial demand for trucking implied by re-shipments of wastewater from CTFs. Addressing these limitations requires ad hoc treatment, as I make clear throughout. Note that these considerations are not relevant for shipments to well pads, because re-transfer of wastewater from one well pad to another is prohibited by the DEP.

[12]For example, I do not observe whether outsourced transfers are mediated by direct interaction between two rival operators, or through a third party. Incentives might differ in each of these cases.

[13]In particular, this prevents me from allowing that the wastewater produced by one fracking event could have been (impossibly) reused as an input for that same fracking event. This restriction results in a loss of 11.4% of wastewater from the sample. I use the full sample for the remainder of Section 2.

[14]It is still possible to calculate aggregate reuse rates for earlier years, although I cannot calculate a sharing rate in this case. Figure 8 presents the full time series of data since 2010.

Table 1: Wastewater disposal market shares

Mode	Facility	% Mode	% Facility
Internal reuse	Own well pad	80.3	46.5
	Own CTF	-	21.9
	3rd party CTF	-	12.0
Rival reuse	Rival well pad	8.3	6.3
	Rival CTF	-	2.0
Injection well		8.1	8.1
Other		3.3	3.3

2.2.2 Other data sources

I supplement the DEP waste reports and related DEP databases with information from other sources. The DEP requires operators to file fracking records on the public FracFocus database maintained by the Ground Water Protection Council and the Interstate Oil and Gas Compact Commission. I rely on this database to understand the timing of fracking events and fracking fluid composition. DEP records include detailed latitude and longitude information for every well pad and disposal facility. I calculate driving distances and driving times between facilities using the Open Source Routing Machine (Luxen and Vetter, 2011) and data from OpenStreetMaps. One limitation of this approach is that I do not account for roadway-specific vehicle weight restrictions that could alter optimal shipment routes for trucks in comparison to passenger vehicles.

2.3 Description of the data

This section briefly summarizes the wastewater data. I present market shares and relate typical shipment patterns to the underlying economic and physical processes discussed above.

Table 1 presents market shares for each of the disposal and reuse methods described previously. I assume that all wastewater transferred to a CTF is ultimately reused. In total, 88.6% of wastewater was reused in the analysis period, while 8.1% was disposed in injection wells, and 3.3% was disposed by some other means.[15] 80.3% of wastewater was transferred to an internal well pad, internal CTF, or third party CTF, while 8.3% of wastewater was transferred to a rival well pad or a rival CTF.[16] 52.8% of wastewater was transferred directly

[15]The "Other" category encompasses (for example) shipments for reuse in West Virginia and landfill disposal of unusable sludges produced as a byproduct of treatment. I exclude these shipments from the main analysis.

[16]The DEP precludes firms from accepting water at one well pad and then later transferring it to another. Wastewater that is transferred directly to a well pad must be used on that well pad. This regulation is

to a well pad, while 35.9% was transferred to a CTF. Thus, the large majority of wastewater is reused, often but not always after being treated on the well pad, and a significant percentage of reuse occurs across the firm boundary.

In any month, there are many more well pads generating wastewater than facilities receiving wastewater for reuse. The first section in Table 2 shows the distribution of the number of well pads reporting wastewater transfers and the number of well pads and CTFs appearing as destinations each month. In the average month 1,712.6 distinct well pads reported wastewater transfers, encompassing transfers to 51.6 destination well pads, 11.0 producer-affiliated CTFs and 10.9 independent CTFs. Because wastewater production declines over time, the majority of well pads disposing of wastewater dispose of small volumes in comparison the volume of wastewater required to complete a well. The second section of Table 2 shows the distributions of monthly volumes by facility type. The mean volume of wastewater per well pad was 23.8 truckloads, but the median was just 3.7 truckloads; 34% of well pads disposed of fewer than two full truckloads.[17] In comparison, a well pad that received transfers in a given month received an average of 430.8 truckloads (median 29.5). Producer-affiliated and third party CTFs received 907.6 and 465.4 truckloads per month, respectively.

The last section of Table 2 shows the distribution of shipment distance for each type of disposal facility. The mean shipment distance was 30.0 miles. Because firms' operations are spatially autocorrelated (shown in Figure 1), wastewater that is reused internally is typically shipped a shorter distance than wastewater that is shared. Internal shipments were 22.5 miles on average, while shipments to rivals were 45.0 miles on average. In comparison, shipments to injection wells (primarily from southwestern Pennsylvania) were 75.5 miles on average. Within the firm, shipments to CTFs are shorter than shipments to well pads (because, as the name implies, these facilities are centrally located), but these shipments would imply subsequent re-shipment at some unknown distance (likely of a similar magnitude).

2.3.1 Participation in the sharing market

I define the *sharing market* to encompass any bilateral exchange of wastewater by operating firms within the state of Pennsylvania for the purpose of reuse regardless of the means of intermediation. The data imply that sharing market volumes are substantial. In this section I briefly describe who shares wastewater, with whom, and under what circumstances. I rely on this evidence to justify key modeling assumptions later in this chapter.

intended to prevent excessive truck traffic.

[17] In the data, volumes are reported in barrels. I convert volumes to truckloads by assuming that water-hauling trucks have a capacity of 110 barrels (the modal shipment volume in the data; in practice, tanker capacity varies from about 80 to 130 barrels).

Table 2: Facility counts and shipment characteristics

	Mean	Std	5%	25%	50%	75%	95%
Facility count per month							
Well pads (origin)	1,712.6	76.6	1,587.2	1,659.8	1,707.5	1,763.0	1,831.0
Well pads (dest)	51.6	14.3	31.4	39.0	50.5	62.8	74.0
Producer CTFs (dest)	11.0	1.5	9.0	10.0	11.0	12.0	13.0
3rd party CTFs (dest)	10.9	1.3	9.0	10.0	11.0	12.0	13.0
Truckloads sent or received by facility-month							
Well pads (origin)	23.8	83.0	0.6	1.4	3.7	11.8	98.4
Well pads (dest)	430.8	915.4	0.9	4.0	29.5	363.0	2,347.7
Producer CTFs (dest)	905.1	1,429.1	2.9	51.4	271.8	1,071.6	4,683.4
3rd party CTFs (dest)	464.5	563.9	4.8	106.0	337.2	649.6	1,349.1
Miles per truckload by destination type							
Own pad or CTF	22.5	20.2	2.8	8.7	17.5	31.5	53.6
Rival pad or CTF	45.0	30.8	10.8	24.2	39.5	57.2	98.7
Injection well	75.5	54.0	18.1	30.1	68.0	88.6	215.9
3rd party CTF	31.4	29.8	4.4	10.6	24.4	44.6	76.0
All destinations	30.0	30.9	3.4	10.5	21.4	37.2	85.1

In any month, most large firms either send or receive wastewater through the sharing market. Among the twenty largest firms, all but five were active on both sides of the sharing market at some point during the sample, and only one never participated at all.[18] In the average month, 9.5 of the twenty largest firms sent wastewater to a rival, 7.0 firms accepted wastewater from a rival, and 3.3 firms did both.[19] Figure 11 plots the number of unique participants in each month. Participation varies considerably over time, but no fewer than 10 distinct large firms were active in each month of the sample period. Most firms share at least occasionally, and sharing status is not "sticky."

Conditional on participation, firms typically have many distinct sharing partners, suggesting that the marginal cost of reaching an agreement with a new counterparty is not large. Among the twenty largest firms, those that sent wastewater to a rival sent wastewater to 1.7 distinct rivals per month and 7.3 distinct rivals over the course of the sample (median 7.0), implying that sharing partners frequently change. Likewise, firms that received wastewater did so from 2.5 distinct rivals per month and 8.1 over the course of the sample (median 7.0). Trade within a firm-pair tends to be short-lived or intermittent, which would be surprising if long-term relationships were an important source of surplus: among pairs of firms that

[18]For these calculations, firms "participate" in the sharing market when sharing more than 1% of their wastewater and/or sourcing more than 1% of wastewater from the sharing market. For this exercise, I define the largest firms in terms of wastewater disposal volumes (rather than gas production volumes).

[19]Among all firms, 18.8 firms sent wastewater to a rival, 8.2 accepted wastewater, and 3.5 did both in the average month.

ever shared, sharing occurred in only 13.1 of 48 months on average, and only 9 pairs of firms shared in more than 24 months.[20]

These patterns are consistent with a model in which the sharing market functions as a spot market in which firms clear unanticipated wastewater imbalances at arms length and on short notice. Indeed, 88% of sharing market volume is between firms that are net suppliers and firms that are net demanders of wastewater in a particular month. Thus, the firms that supply most of the wastewater in the sharing market do so while accepting relatively little wastewater in return, and vice versa, to a greater degree than might be expected if firms fully coordinated wastewater management in advance. Unanticipated imbalances might arise due to the inherent lumpiness and unpredictability of wastewater generation and fracking activity.

2.4 External costs

In the policy analysis, I focus on negative environmental externalities created by wastewater trucking. Taken together, the data imply that about 500,000 truckloads of wastewater leave Pennsylvania well pads each year, at an average shipment distance of 30 miles. If all wastewater shipped to CTFs is subsequently re-shipped by truck, then a further 180,000 truckloads leave CTFs, at an average distance likely close to 30 miles (the mean shipment distance to CTFs). Emissions scale linearly with ton-miles. Under some assumptions regarding typical vehicle weight and emissions factors, it is possible to obtain estimates for the social costs of carbon emissions and air pollution associated with wastewater trucking using simple methodologies. In Appendix 2.4, I show that these costs are around $7M per year, or about 7% of the distance-related trucking costs incurred by firms.[21]

External damages from spills are more difficult to quantify. Spills pose serious ecological risks and can threaten drinking water resources (EPA, 2016). Maloney et al. (2017) find that transportation and storage contribute to 50 wastewater spills per year in Pennsylvania, but it is unclear how many of these spills occur on the road rather than at a well site (for example, during loading or unloading). Statistics from EPA (2016) suggest that the crash rate for tanker trucks is on the order of 100 per 100M truck-miles and the spill rate conditional on

[20]Some factors that make relationship formation difficult in this setting include the inherently competitive relationship between rival firms, and the relative thickness of the sharing market, both of which raise firms' incentives to defect. Hubbard (2001) and Harris and Nguyen (2023) are two notable empirical papers exploring the relationship between market thickness, contracting, and relationship formation.

[21]This is smaller than but similar in magnitude to the 20% external damages estimate for railroad transport of crude found in (Clay et al., 2019). There are two possible explanations for this difference. The first possibility is that marginal trucking costs are substantially less than $5 per mile – marginal trucking costs of $1.75 per mile would imply a 20% external damages estimate. A second possible explanation is that NOx and PM2.5 emissions are smaller for trucks than for locomotives, especially after model year 2010.

crashing is about 5-10%, implying a total of 1-2 spills a year. However, oil and gas-related trucking may differ from trucking in other industries, due to differences in demographics (Wilson, 2022) and hours of service regulations (Muehlenbachs et al., 2021). Several papers have found evidence that accident rates increase during shale development (e.g., Graham et al., 2015; Xu and Xu, 2020), but Muehlenbachs et al. (2021) fail to find causal evidence of elevated crash rates for trucks (primarily water-hauling trucks) in Pennsylvania specifically.

While I focus on environmental externalities, prior work has examined other unpriced externalities associated with heavy truck traffic created by the shale boom, such as elevated traffic fatalities (Muehlenbachs et al., 2021) and road damage (Abramzon et al., 2014).

3 Sharing frictions

I define *sharing frictions* as transaction costs specific to sharing transactions – costs incurred if a transaction takes place between firms, but not if a transaction takes place within a firm. This section presents evidence that sharing frictions play an important role in the allocation of wastewater for reuse.

3.1 The EQT-Rice merger

In November 2017 EQT Corporation and Rice Energy Inc. merged, creating one of the largest natural gas producers in the United States.[22] In this section I argue that changes in wastewater shipment patterns subsequent to the merger suggest that sharing frictions are economically large.

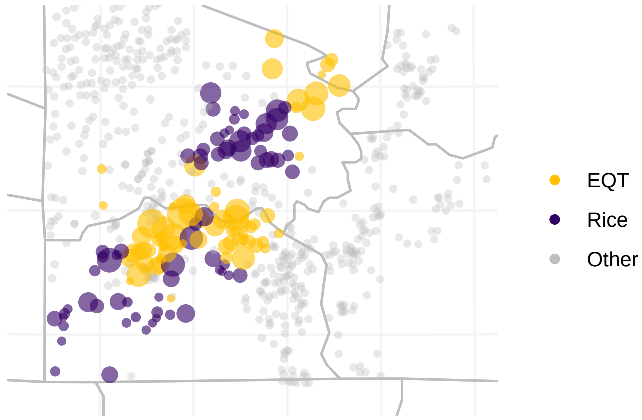
Figure 3 shows the locations of EQT and Rice well pads appearing in the data.[23] In the six months leading up to the merger announcement, 99% of Rice's wastewater volume originated at well pads within 22.5 miles of an EQT facility that accepted wastewater, and 68% of EQT's wastewater volume originated at well pads within 22.5 miles of a Rice facility that accepted wastewater (22.5 miles being the mean distance of internal shipments in the data). The exceptional geographic proximity between EQT and Rice was the primary stated rationale for the merger, in part because of potential synergies in "rig allocation, pad sites, water, access roads, etc." and pipeline access, but primarily because of the potential to drill longer wells by amalgamating existing leases.[24]

[22]EQT and Rice were respectively the 4th and 8th largest producers in Pennsylvania by gas production in the year leading up to the merger announcement. The merger was announced in June 2017 and completed in November 2017.

[23]EQT also had a significant presence in West Virginia, while Rice was present in Ohio.

[24]See EQT Corporation (2017).

Figure 3: EQT and Rice pre-merger well pad locations



Despite apparently large geographic synergies, EQT and Rice never shared prior to the merger. The first two columns of the left panel of Table 3 summarize disposal market shares for EQT and Rice well before the merger.[25] 91.1% of EQTs wastewater was reused internally or sent to a third party CTF, while 8.9% was sent to rivals other than Rice. The right panel indicates that during this period EQT received some wastewater from rivals other than Rice (about 2.7% of total wastewater received). In contrast, all of Rice's wastewater was reused internally or sent to a third party CTF. Rice received 4.1% of wastewater from a single rival firm.[26]

After the merger, transfers between formerly-unintegrated EQT and Rice facilities increased dramatically. The second two columns of the left panel of Table 3 indicate post-merger market shares for former Rice- and EQT well pads. After the merger, 22.5% of wastewater generated at EQT-linked well pads was transferred to Rice-linked facilities, and 62.4% of wastewater generated at Rice-linked facilities was transferred to EQT-linked facilities. 40.1% of wastewater received at EQT-linked facilities came from Rice-linked well pads, and 49.1% of wastewater received at Rice-linked facilities came from EQT-affiliated well pads. Thus, the removal of the firm boundary was followed by a significant increase in "sharing," consistent with the elimination of significant ex ante sharing frictions.[27]

[25]The pre-merger period spans January to June 2017, while the post-merger figure spans December 2017 to December 2020.

[26]This firm, Alpha Shale Resources, had previously been involved in a joint venture with Rice. Rice bought out Alpha's joint venture stake in 2014, before the sample period. Thus, these shipments may reflect reporting errors in the wastewater data, if Rice was the de facto operator of the sending well pads when the shipments were observed. Alternatively, sharing frictions between Rice and Alpha might have been particularly low as a result of their previous joint venture, facilitating exchange.

[27]It is also possible these patterns are the result of unobserved changes in the joint entity's completion

Table 3: EQT and Rice pre- and post-merger market shares

Share of Wastewater Leaving Well Pad				
	Pre-merger		Post-merger	
Destination	EQT	Rice	EQT	Rice
EQT pad	83.4	0.0	65.1	62.4
Rice pad	0.0	70.6	22.5	31.0
Other rival	8.9	0.0	2.7	0.6
3rd party CTF	7.7	29.3	8.5	4.6
Injection well	0.0	0.0	1.2	1.4

Share of Wastewater Received				
	Pre-merger		Post-merger	
Source	EQT	Rice	EQT	Rice
EQT pad	97.3	0.0	59.1	50.5
Rice pad	0.0	95.9	40.1	49.1
Other rival	2.7	4.1	0.8	0.4

3.2 Sharing friction heterogeneity and other intrinsic costs

The EQT-Rice merger suggests that sharing frictions can have significant effects on wastewater allocations. However, the data indicate many firms frequently participate in the sharing market when gains from trade are presumably smaller, suggesting that sharing frictions may exhibit significant heterogeneity. Thus, it is unclear what the aggregate impacts of sharing frictions might be.

Because transportation costs are a primary driver of wastewater management costs, one way to understand the scale of aggregate impacts is to compare the observed allocation of wastewater to the minimum-distance allocation. If sharing frictions and distance-related transportation costs are the only sources of variation in cost, this exercise identifies the aggregate impact of sharing frictions.

The minimum-distance allocation can be obtained as follows. Consider the flow of wastewater in month t between all well pads K and all disposal facilities D. For each well pad $\kappa \in K$ and disposal facility $\delta \in D$, the data records the actual shipment volume $\hat{\mu}_{\kappa\delta}$. Suppose all truckloads of wastewater shipped within a month are perfectly substitutable. Holding fixed the total disposal volume $Q_\kappa = \sum_{\delta \in D} \hat{\mu}_{\kappa\delta}$ at well pad κ and the total volume received $C_\delta = \sum_{\kappa \in K} \hat{\mu}_{\kappa\delta}$ at facility δ, the minimum-distance allocation μ is a solution to the

activity or other unobserved changes in relative costs, rather than the removal of sharing frictions specifically. To address this concern, I revisit the EQT-Rice merger from an equilibrium perspective later in this chapter.

following problem:

$$\begin{aligned} \min_{\mu \geq 0} \quad & \sum_{\kappa \in K} \sum_{\delta \in D} \mu_{\kappa\delta} d_{\kappa\delta} \\ \text{s.t.} \quad & \sum_{\delta \in D} \mu_{\kappa\delta} = Q_\kappa \quad \forall\, \kappa \in K \\ & \sum_{\kappa \in K} \mu_{\kappa\delta} = C_\delta \quad \forall\, \delta \in D \end{aligned} \tag{1}$$

where $d_{\kappa\delta}$ is the distance between κ and δ.[28] Because this problem is a linear program, it can be solved easily. Key results are summarized in the rightmost column of Table 4. The mean trucking distance decreases by 16% under μ relative to the status quo, while the aggregate sharing rate increases by 21.4 percentage points to 31.9%, suggesting that the aggregate impacts of sharing frictions may be substantial.

In reality, not all truckloads of wastewater are perfectly substitutable. Many factors can shift the real or shadow costs of potential shipments, even within the firm. For instance, the timing of when wastewater must be removed from one well pad might not align well with the timing of when can be accepted at another. Transportation costs unrelated to distance might differ unobservably depending on the structure of trucking contracts, loading and unloading times, or opportunities for backhauls. Treatment costs could vary depending on wastewater composition and the particular characteristics of a completion (fracking fluid formulation, well construction, target formation). Thus, even if all well pads K and disposal facilities D were controlled by a single firm, the observed shipment plan would not necessarily correspond to the minimum-distance allocation.

To understand the empirical significance of imperfect substitutability, I compare observed shipments for reuse *within* firms to two simple benchmarks. The first is random matching from origins K_f to destinations D_f, irrespective of distance. The second implements the minimum-distance allocation within each firm f conditional on observed interfirm shipment volumes. Summary statistics for each benchmark are provided in the second and third columns of Table 4. Comparing the data to these benchmarks, I find that within-firm shipments realize 56% of possible transportation cost savings relative to random matching. Thus, differences in distance only partially explain firm shipment decisions.

Repeating this exercise for the market as a whole, I find that observed shipments realize 95% of possible transportation cost savings relative to random matching. This rate exceeds the within-firm rate, suggesting that reductions in sharing frictions could lead to less efficient

[28] I focus on over-the-road trucking distance, but one could also consider trucking time. The more appropriate choice depends on the structure of trucking contracts, which may vary in practice.

Table 4: Minimum-distance benchmarks

		Within-firm		Market-wide	
	Data	Random	Optimal	Random	Optimal
Miles per truckload					
All	24.9	27.9	22.6	147.0	18.6
Internal	22.7	26.1	20.1	29.0	15.9
Rival	43.6	43.6	43.6	168.9	24.5
Sharing rate	10.5	10.5	10.5	84.4	31.9

transportation overall. Because firms are geographically segregated from one another, even relatively inefficient shipments within the firm (compared to the minimum-distance benchmark) result in a relatively efficient outcome when viewed from the perspective of the market as a whole.

4 Model

In this section I develop a model of wastewater management to quantify sharing frictions and other intrinsic costs of reuse. To do so, I adapt the matching framework of Choo and Siow (2006) and Galichon and Salanie (2022) to the setting of the Pennsylvania wastewater market.

Let K denote the finite set of well pads generating wastewater in month t, and D the finite set of facilities accepting wastewater for reuse (both well pads and operator-affiliated CTFs).[29] For simplicity, I assume that each well pad $\kappa \in K$ and facility $\delta \in D$ exists for a single month and then vanishes.

The firm is a coalition of managers who make decisions independently from one another. Each wellpad $\kappa \in K$ is controlled by a manager m_κ, while each facility $\delta \in D$ is controlled by a manager m_δ. m_κ ships each truckload of wastewater i to the least cost destination, whether reuse at a facility $\delta \in D$ or injection disposal (the outside option, $\{0\}$). At the same time, m_δ sources each truckload of completion water j from the least cost source, whether wastewater from a well pad $\kappa \in K$ or freshwater (the outside option, $\{0\}$). In the core of the matching game that I describe, this representation of the firm is without loss, since a coalition of managers cannot achieve lower costs than a collection of managers acting independently.

A mass of Q_κ truckloads of wastewater are generated at $\kappa \in K$, while a mass of C_δ truckloads are needed at $\delta \in D$. Let $r_{\kappa\delta}$ denote the systematic cost of reusing wastewater

[29] For the purpose of estimation, I exclude shipments to third party CTFs from the sample.

from κ at δ, which encompasses transportation costs and any other costs relevant to firms. $r_{\kappa\delta}$ is *systematic* in the sense that it is incurred for all truckloads shipped between κ and δ. The form of $r_{\kappa\delta}$ differs depending on whether κ and δ belong to the same firm or to rival firms. Let $\mathcal{I}$ denote the set of $\kappa\delta$ pairs within firms, and $\mathcal{R}$ the set of $\kappa\delta$ pairs between firms, such that $\mathcal{I} \cup \mathcal{R}$ partitions $K \times D$. Then:

$$r_{\kappa\delta} = \begin{cases} r^{\mathcal{I}}_{\kappa\delta} & \text{if } \kappa\delta \in \mathcal{I} \\ r^{\mathcal{I}}_{\kappa\delta} + \phi_{\kappa\delta} & \text{if } \kappa\delta \in \mathcal{R} \end{cases}$$

where $r^{\mathcal{I}}_{\kappa\delta}$ is the systematic cost of reuse if κ and δ belong to the same firm, and $\phi_{\kappa\delta}$ represents *sharing frictions* incurred when they do not. $r^{\mathcal{I}}_{\kappa\delta}$ may depend on the observable characteristics $X_{\kappa\delta}$ and unobservable characteristics $\xi_{\kappa\delta}$ of a potential match, but does not depend on whether κ and δ belong to the same firm. Thus, sharing frictions are the difference between systematic costs when firm boundaries exist and when they do not.

In practice, the cost of reusing wastewater from κ at δ varies from truckload to truckload due to *latent cost heterogeneity.* The true cost of supplying the jth unit of demand at δ with truckload i from κ is:

$$\tilde{r}_{ij} = r_{\kappa\delta} - \epsilon_{i\delta} - \eta_{\kappa j}$$

where $\epsilon_{i\delta}$ and $\eta_{j\delta}$ represent latent truckload-specific cost savings for managers m_κ and m_δ. These capture any idiosyncratic costs that are relevant for manager's decisions regarding specific truckloads i and j, but distinct from the systematic costs affecting all shipments between κ and δ. Latent costs are additively separable across i and j.[30]

If i is not reused, it can be sent to an injection disposal well at cost $\tilde{r}^K_{i0} = r^K_{\kappa 0} - \epsilon_{i0}$, where $r^K_{\kappa 0}$ is the systematic cost of injection disposal from κ. Likewise, freshwater can be obtained for j at cost $\tilde{r}^D_{0j} = r^D_{0\delta} - \eta_{0j}$, where $r^D_{0\delta}$ is the systematic cost of obtaining freshwater at δ. $D_0 = D \cup \{0\}$ is the set of all locations to which wastewater can be shipped (for disposal or reuse) and $K_0 = K \cup \{0\}$ is the set of all locations from which source water (wastewater or freshwater) can be shipped for use in completions.

I assume that $\epsilon_{i\delta}$ is drawn iid from $P_K(X_{\kappa\delta};\theta)$ and $\eta_{\kappa j}$ is drawn iid from $P_D(X_{\kappa\delta};\theta)$. These distributions may depend on observables $X_{\kappa\delta}$ and a parameter vector θ, but not on

[30] This implies that m_κ is indifferent as to whether i is used as input j or j' at δ, and that m_δ is indifferent as to whether truckload i or i' is received from κ. These implications are empirically reasonable: the composition of wastewater changes little on short time horizons at κ, and wastewater is unlikely to segregated at δ. Separability does not exclude "matching on unobservables" altogether: m_κ's cost of shipping to δ or might vary from i to i'; and symmetrically, m_δ's cost of reusing wastewater from κ might vary from j to j', resulting in occasional matches.

unobservables. Importantly, they do not depend on whether κ and δ are operated by the same firm. Thus, I assume that latent costs are not distributed differently across internal transaction and sharing transactions except insofar as latent costs correlate with observed facility and match characteristics that are independent of the firm boundary. The main purpose of this assumption is to simplify the interpretation of the model: under this assumption, $\phi_{\kappa\delta}$ is the sole source of cost differences between internal reuse and the sharing market.[31] I assume that P_K and P_D have full support and finite expectations.

The core of the matching game consists of the set of all *stable*, *feasible* matchings of wastewater from K to D. A matching μ is *stable* if no manager m_κ would prefer to ship a truckload i allocated for reuse under μ to a disposal well, no manager m_δ would prefer to replace wastewater received under μ with freshwater, and no m_κ and m_δ would privately agree to match any i and j that are not matched under μ. A matching is *feasible* if every truckload i is matched to some $\delta \in D_0$, and every j is allocated to a truckload of wastewater or freshwater from some $\kappa \in K_0$. The matching μ can be represented as a probability mass function, where $\mu_{\kappa\delta}$ represents the probability of observing a shipment of wastewater between κ and δ. Using this convention, the set of feasible matchings $\mathcal{M}(\mathbf{Q}, \mathbf{C})$ consists of all matchings μ such that $\sum_{\delta \in D_0} \mu_{\kappa\delta} = Q_\kappa$ for all $\kappa \in K$ and $\sum_{\kappa \in K_0} \mu_{\kappa\delta} = C_\delta$ for all $\delta \in D$, and $\mu_{\kappa\delta} \geq 0$ for all $\kappa\delta \in K \times D$.

Gretsky et al. (1992) establish that the core of the matching game is equivalent to the set of Walrasian equilibria of an exchange economy. In a Walrasian equilibrium characterized by transfer matrix τ, each manager m_κ solves a discrete choice problem for truckload i:

$$\min_{\delta \in D_0} \quad r^K_{\kappa\delta} + \tau_{\kappa\delta} - \epsilon_{i\delta} \tag{2}$$

where $r^K_{\kappa\delta}$ represents the portion of the costs of reuse incurred by the sender and $\tau_{\kappa\delta}$ is a (possibly negative) transfer of utility to m_δ. Symmetrically, each manager m_δ solves a discrete choice problem for truckload j:

$$\min_{\kappa \in K_0} \quad r^D_{\kappa\delta} - \tau_{\kappa\delta} - \eta_{\kappa j} \tag{3}$$

where $r^D_{\kappa\delta}$ are the costs of reuse incurred by the receiver and $\tau_{\kappa\delta}$ is the utility transfer received from m_κ. In equilibrium, the mass of transfers $\mu_{\kappa\delta}$ between κ and δ is equal to m_κ's demand for disposal at δ and to m_δ's supply of capacity to wastewater from κ when choices are

[31] Relaxing this assumption amounts to allowing for the possibility that sharing frictions differ from truckload to truckload for the same origin-destination pair. This might be plausible if, for example, information frictions specific to sharing transactions (e.g. adverse selection on low quality wastewater) vary at the level of the truckload within the well pad-month or facility-month. While this type of variation is plausible, presumably most of the variation in sharing frictions exists at the level of the facility-pair-month.

made according to (2) and (3). Note that the systematic costs $r^{\mathcal{I}}_{\kappa\delta}$ are exactly equal to the sum $r^K_{\kappa\delta} + r^D_{\kappa\delta}$ for all potential transactions. Furthermore, utility is perfectly transferrable. This assumption is reasonable because firms can readily exchange cash, but $\tau_{\kappa\delta}$ does not necessarily represent a cash transfer.[32]

Galichon and Salanie (2022) show there is a unique equilibrium matching μ^* in the core. Moreover, μ^* minimizes a social cost function:

$$\mu^* = \arg\min_{\mu\in\mathcal{M}(\mathbf{Q},\mathbf{C})} \sum_{\kappa\in K}\sum_{\delta\in D}\mu_{\kappa\delta}\left\{r_{\kappa\delta} - r^K_{\kappa 0} - r^D_{0\delta}\right\} - \mathcal{E}\left(\mu,\mathbf{Q},\mathbf{C}\right) \tag{4}$$

In this expression, the first term captures all systematic cost savings from reuse. The match entropy term $\mathcal{E}$ captures the surplus contribution of latent cost savings. Intuitively, $\mathcal{E}$ quantifies the amount of latent cost heterogeneity required to rationalize a given match μ conditional on the distributions of P_K and P_D. In the special case that P_K and P_D are extreme value type 1 distributions with dispersion parameters σ_K and σ_D,

$$\mathcal{E}\left(\mu,\mathbf{Q},\mathbf{C}\right) = -\sigma_K\sum_{k\in K}\sum_{\delta\in D_0}\mu_{\kappa\delta}\log\left\{\frac{\mu_{\kappa\delta}}{Q_k}\right\} - \sigma_D\sum_{\delta\in D}\sum_{k\in K_0}\mu_{\kappa\delta}\log\left\{\frac{\mu_{\kappa\delta}}{C_\delta}\right\}$$

I provide a general formulation of $\mathcal{E}$ in Appendix 11.

4.1 Discussion

Modeling the firm as a collection of managers is non-standard, but without loss in the context of a matching model: the firm can be viewed as a coalition of managers, and in the core no coalition of managers can do achieve lower costs than managers acting independently. Equilibrium transfers within the firm capture the shadow costs of shipments that crowd out more efficient internal shipments or profitable exchanges in the sharing market.

This decentralized approach excludes various forms of strategic behavior by firms. For example, a firm cannot earn more surplus by threatening to abstain from sharing (as recognized by Shapley and Shubik (1971)). In contrast, I implicitly assume that all firms are always willing to trade with all other firms at the margin, even if it is usually too costly to do so. Some important forms of strategic behavior could be captured in a model of multilateral Nash bargaining (such as the so-called Nash-in-Nash model of Horn and Wolinsky (1988)).[33] A challenging in adapting the Nash-in-Nash approach to this setting is the difficulty of constructing counterfactual "sharing networks." In this setting, sharing networks are potentially

[32] For example, $\tau_{\kappa\delta}$ could represent a "favor" (as in Samuelson and Stacchetti (2017)).

[33] Carlton (2020) notes the absence of transaction costs in prior applications of Nash-in-Nash bargaining.

large, and existing approaches to simulating network formation are not obviously applicable due to the lack of a clear distinction between upstream and downstream firms.[34]

Another limitation of the model is that sharing frictions are exogenous and incurred in proportion to the number of truckloads sent. Linearity of the sharing frictions excludes the possibility that sharing frictions might be amortized over many similar truckloads (although, presumably, estimated frictions would be smaller in this case). In this way, the model differs from trade models that stipulate fixed costs of importing and exporting (Antrás and Chor, 2022). From a theoretical perspective, it would be natural to endogenize sharing frictions by modeling relationship dynamics (e.g., Chassang, 2010; Gibbons and Henderson, 2012) or favor trading (Samuelson and Stacchetti, 2017). Recent empirical work in the development literature has made progress on quantifying relationship dynamics (e.g., Macchiavello and Morjaria, 2015, 2021), but I do not pursue this approach here, primarily because I find limited evidence of relationship-driven patterns of trade.

5 Identification and estimation

This section discusses identification and estimation.

5.1 Identification

Let Δr denote the $K \times D$ matrix of systematic cost savings with typical element $\Delta r_{\kappa\delta} = r_{\kappa\delta} - r^K_{\kappa 0} - r^D_{0\delta}$. Galichon and Salanie (2022) establish that the systematic cost savings Δr are identified from the observed match μ^* (i.e., from the shipments of wastewater) conditional on $(P_K, P_D, \mathbf{Q}, \mathbf{C})$.[35] However, the systematic costs of internal reuse $r^{\mathcal{I}}$ and sharing frictions ϕ are not separately identified without further restrictions. Thus, I assume that $r^{\mathcal{I}}$ takes the following form:

$$r^{\mathcal{I}}_{\kappa\delta} = g_{\kappa\delta}\left(X_{\kappa\delta};\theta\right) + u^{\mathcal{I}}_{\kappa} + u^{\mathcal{I}}_{\delta} \tag{5}$$

where $g_{\kappa\delta}\left(X_{\kappa\delta};\theta\right)$ is a known function of observables $X_{\kappa\delta}$ and parameters θ, and $u^{\mathcal{I}}_{\kappa}$ and $u^{\mathcal{I}}_{\delta}$ are additively separable unobservables.[36] Separability implies that *within the firm*, any

[34] Ho and Lee (2019) and Ghili (2022) develop models of network formation in Nash-in-Nash environments. These papers exploit institutional differences between upstream and downstream firms to simplify the strategy space (e.g., by assuming that one side of the market can pre-commit to a particular network).

[35] This result is known for the special case of the logit from Choo and Siow (2006).

[36] In general, I assume that parameters are constant across months of the sample. However, the assumption that κ and δ exist for a single month implies that $u^{\mathcal{I}}_{\kappa}$ can differ over time for the same physical well pad (and likewise for $u^{\mathcal{I}}_{\delta}$). This assumption is reasonable: in practice, the characteristics of wastewater generated at a particular well pad are constantly evolving, as are drilling needs at facilities where reuse can occurs.

unobserved systematic costs of reusing wastewater at δ are independent of the source of the wastewater $\kappa \in K$. Symmetrically, any unobserved systematic costs of sourcing wastewater from κ are independent of the facility where reuse occurs $\delta \in D$. Thus, reuse may be generally more costly at some facilities than at others, and wastewater from some well pads may be generally more costly to accept, but there are no unobserved complementarities in the cost of reuse that shift the costs of all truckloads between κ and δ.

Under this assumption, $r^{\mathcal{I}}$ is identified from observed shipments within each firm, and ϕ is non-parametrically identified conditional on $r^{\mathcal{I}}$ (because Δr itself is identified, and $\phi_{\kappa\delta} = \Delta r_{\kappa\delta} - r^{\mathcal{I}}_{\kappa\delta}$). In practice, I do not observe a sufficient amount of data to non-parametrically estimate $\phi_{\kappa\delta}$ for every $\kappa\delta$. I therefore assume that sharing frictions ϕ take the following functional form:

$$\phi_{\kappa\delta} = h_{\kappa\delta}\left(X_{\kappa\delta};\theta\right) + \pi_b$$

where $h_{\kappa\delta}\left(X_{\kappa\delta};\theta\right)$ is a parametric function of observables and π_b is a bilateral fixed effect for sharing between a pair of rivals firms b (the operators of κ and δ). This structure allows for firm pair-specific unobserved heterogeneity (for example, the presence of a relationship), but in a restricted form: sharing frictions are constant within firm pairs, except insofar as they differ with observable match characteristics. Thus, there are no unobserved sources of heterogeneity in match-specific sharing frictions within the set of facilities operated by each pair of firms. This excludes the possibility that a certain pair of firms has a comparative advantage in coordinating certain types of shipments but not others.

5.1.1 Welfare-relevance

Sharing frictions are a type of transaction costs. While some authors have interpreted transaction costs as strictly real costs, this is not required: transaction costs might instead be viewed as "choice costs" that differ from "true costs" in the same way that "choice utility" differs from "true utility" when behavioral frictions are incorporated into models of consumer choice. Under this interpretation, the welfare-relevant component of the sharing frictions is not separately identified from the welfare-irrelevant component. Examples of welfare-relevant costs could include wages expended in search and negotiation, or quantifiable risks to future profits (e.g, production risks). Examples of welfare-irrelevant costs include shadow costs arising from managerial inattention, loss aversion, or excessive secrecy. This distinction plays an important role in the policy analysis below.

5.2 Parameterization

To estimate the model, I further assume that the observable component of systematic disposal costs $g_{\kappa\delta}\left(X_{\kappa\delta};\theta\right)$ is linear in transportation costs and other observables:

$$g_{\kappa\delta}\left(X_{\kappa\delta};\theta\right)=d_{\kappa\delta}+x'_{\kappa\delta}\beta \tag{6}$$

where $d_{\kappa\delta}$ is the linear component of transportation costs (e.g., distance- or driving time-related costs) between κ and δ and $x_{\kappa\delta}$ is a vector of origin-destination-specific observables. The coefficient on $d_{\kappa\delta}$ is one, so that all other parameters are interpretable in terms of the marginal cost of distance, which I assume is known. I focus on the case of linear (rather than log) over-the-road distance because trucking contracts in similar industries often have a per mile component, and because linear distance appears to deliver a good model fit.[37] $x_{\kappa\delta}$ includes controls for differences between sending well pads and receiving facilities that could (possibly) shift the cost of reuse. Specifically, $x_{\kappa\delta}$ includes proxies intended to capture potential differences in fracking fluid formulation and wastewater composition.

Similarly, I assume that the common component of sharing frictions is linear in observables:

$$h_{\kappa\delta}\left(X_{\kappa\delta};\theta\right)=z'_{\kappa\delta}\alpha$$

where $z_{\kappa\delta}$ is a vector of match-specific observables. $z_{\kappa\delta}$ includes facility-level characteristics to disentangle potential sources of sharing frictions (e.g., wastewater characteristics, well pads vs. CTFs).

In practice, I cannot feasibly estimate π_b for every pair of firms. Instead, I keep only the 27 firm pairs with the largest bi-directional sharing volume (encompassing 50% of all sharing volume) and aggregate all remaining firms into 6 groups of observably similar firm pairs (for example, large firms in northwestern Pennsylvania sharing with small firms anywhere). I make this decision for two reasons. First, it is difficult to estimate a large number of parameters using the estimation procedure I describe in the next section. Second, an estimate for each firm pair fixed effect π_b need not exist in a finite sample if firms are never observed to share. A partial aggregation strategy ensures the existence of an estimate for each pair of firms. The cost of this assumption is that any resulting violations of the assumption that

[37] In addition to distance, I also consider driving time, log distance, and non-linear distance measures (e.g., 30 mile increments); see the discussion of alternative specifications below. I do not explicitly model other components of transportation costs (for example, labor expenses incurred while loading and unloading), but to the extent that these costs are specific to particular sending and receiving facilities rather than particular routes, they are captured in the facility fixed effects $u^{\mathcal{I}}_{\kappa}$ and $u^{\mathcal{I}}_{\delta}$.

π_b is constant within groups introduce bias into the estimates.

Finally, I assume that the latent cost distributions $P_K(X_{\kappa\delta};\theta)$ and $P_D(X_{\kappa\delta};\theta)$ are type 1 extreme value error distributions with mean zero and scale parameters σ_K and σ_D, respectively. I make this choice for simplicity and computational convenience: computation of the equilibrium is significantly more efficient in this case than with richer forms of heteroskedasticity. Moreover, this assumption enables me to estimate the main parameters of interest without ad hoc assumptions regarding the shares of the outside options, as I discuss in the next section. Alternatively, one could consider scale parameters that vary depending on the facility type (for example, for larger and smaller well pads in K and between larger and smaller well pads and CTFs in D), or a similarly constructed nesting structure.

I discuss covariate construction in Appendix 10 and additional details concerning the full specification of the model in Appendix 12.

5.3 Estimation

For each month t, I observe the total shipment volume $\hat{\mu}_{\kappa\delta}$ for each $\kappa \in K$ and $\delta \in D$, as well $\hat{\mu}_{\kappa 0}$ for each $\kappa \in K$ and (in auxiliary data) $\hat{\mu}_{0\delta}$ for each $\delta \in D$. Using these data, Galichon and Salanie (2022) derive a maximum likelihood estimator for the true parameter vector $\theta_0 \in \Theta$ under the assumption that the data reveal $\mathbf{Q}$ and $\mathbf{C}$. This last assumption is analogous to the conventional assumption that market shares are observed without error in demand models. In my setting, such an assumption is unattractive because $\mu_{\kappa 0}$ and $\mu_{0\delta}$ are observed with noise.[38] At least under the logit assumption, it is possible to derive an asymptotically equivalent estimator that does not rely on noisy estimates of $\mu_{\kappa 0}$ or $\mu_{0\delta}$. Under the assumption that $\mathbf{Q}$ and $\mathbf{C}$ are known and ϵ and η follow Gumbel distributions, a consistent (but inefficient) maximum likelihood estimator for θ_0 is:

$$\hat{\theta} = \arg\max_{\theta\in\Theta} \sum_{t\in T}\sum_{\kappa\in K}\sum_{\delta\in D} \hat{\mu}_{\kappa\delta} \log\left(\frac{\mu_{\kappa\delta}(\theta;\mathbf{Q},\mathbf{C})}{\sum_{\kappa\delta\in K\times D}\mu_{\kappa\delta}(\theta;\mathbf{Q},\mathbf{C})}\right)$$

[38]When δ is a well pad, $\mu_{0\delta}$ can be estimated from injection volumes recorded in FracFocus. However, the timing of recorded wastewater shipments does not perfectly align with the timing of recorded fracking events, so this requires further assumptions. Moreover, when δ is a CTF, estimating $\mu_{0\delta}$ requires further assumptions about when and where wastewater shipped to CTFs is ultimately re-used. On the other hand, some percentage of shipments to injection wells do not actually represent shipments of reusable wastewater. For example, sludges produced as a byproduct of the treatment process are also shipped to injection wells as liquid wastes, but these volumes (though small) are indistinguishable from reusable water in the data.

In Appendix 12, I establish that this estimator is equivalent to:

$$\hat{\theta} = \arg\max_{\theta \in \Theta} \sum_{t \in T} \sum_{\kappa \in K} \sum_{\delta \in D} \hat{\mu}_{\kappa\delta} \left(\sigma_K + \sigma_D\right)^{-1} \left\{-d_{\kappa\delta} - x'_{\kappa\delta}\beta - z'_{\kappa\delta}\alpha - \pi_b + \tilde{u}_\kappa + \tilde{v}_\delta\right\} \tag{7}$$

where $\tilde{u} \in \mathbb{R}^K$ and $\tilde{v} \in \mathbb{R}^D$ are latent mean utility parameters that satisfy a system of conditional market clearing equations for each $t \in T$:

$$\sum_{\delta \in D} \exp\left\{(\sigma_\kappa + \sigma_\delta)^{-1} \left\{-d_{\kappa\delta} - x'_{\kappa\delta}\beta - z'_{\kappa\delta}\alpha - \pi_b + \tilde{u}_\kappa + \tilde{v}_\delta\right\}\right\} = \sum_{\delta \in D} \hat{\mu}_{\kappa\delta} \quad \forall\, \kappa \in K \tag{8}$$

$$\sum_{\kappa \in K} \exp\left\{(\sigma_\kappa + \sigma_\delta)^{-1} \left\{-d_{\kappa\delta} - x'_{\kappa\delta}\beta - z'_{\kappa\delta}\alpha - \pi_b + \tilde{u}_\kappa + \tilde{v}_\delta\right\}\right\} = \sum_{\kappa \in K} \hat{\mu}_{\kappa\delta} \quad \forall\, \delta \in D \tag{9}$$

Importantly, (7) does not depend on $\hat{\mu}_{\kappa 0}$ or $\hat{\mu}_{0\delta}$. Due to Sinkhorn's Theorem, $\tilde{u}$ and $\tilde{v}$ satisfying (8) and (9) exist and are unique up to scale (see, e.g., Idel, 2016). In Appendix 12, I describe how I solve this system for a given θ. Together, $\tilde{u}$ and $\tilde{v}$ rationalize the observed marginal market shares conditional on θ analogously to how additively separable terms representing unobserved heterogeneity rationalize observed market shares in the Berry et al. (1995) (or "BLP") setting. Due to this similarity, the implementation of (7) is similar to the implementation of BLP-style demand models (and many of the computational suggestions of Conlon and Gortmaker (2020) are directly applicable).[39] I implement the estimator using KNITRO and obtain standard errors under maximum likelihood assumptions.

5.3.1 Counterfactuals

Substitution to the outside goods depends on the model primitives. Without incorporating data on the outside options, it is not possible to estimate substitution to the outside goods in counterfactuals. I therefore construct all counterfactuals holding fixed the total volume of wastewater generated and reused at each well pad K and facility D.[40]

6 Estimates

This section presents the main estimates. I focus on the estimated sharing frictions and their interpretation before briefly describing other estimated parameters, model fit, and alternative specifications.

[39] Chiong et al. (2016) formalize an equivalence between demand inversion and equilibrium computation in two-sided matching models. Bonnet et al. (2022) extends this result to a richer class of demand models.

[40] This can be formalized with a timing assumption: first, managers choose between reuse and their outside options; then, in a second stage, the reuse market clears. The counterfactuals that I construct can be viewed as counterfactuals of the second stage of this game.

6.1 Sharing frictions

Table 5 presents the main estimates. For the average truckload of wastewater shipped to a rival firm, sharing frictions are equivalent to the cost of shipping a truckload of wastewater 135.1 miles. If trucking costs \$5 per mile, this implies a cost of about \$6.14 per barrel of wastewater shared. In comparison, the mean shipment distance is 43.6 miles. This implies that bringing a typical sharing transaction inside the firm has the same effect on private costs as a threefold decrease in distance.

From an aggregate perspective, the estimates imply that firms incur sharing frictions equivalent to a cost of about \$22M/year. At the same time, the presence of sharing frictions leads to less efficient matching of wastewater from old wells to new wells.[41] Comparing the estimated model to a counterfactual with no sharing frictions suggests that sharing frictions indirectly raise firms' private costs by about \$27M/year (for example, through increased transportation costs). In comparison, firms incur about \$500M/year in freshwater sourcing, wastewater transportation, and final disposal costs. This back-of-the-envelope suggests that sharing frictions raise firms' total water-related costs by roughly 10%.

The model implies that these costs could be eliminated through integration. To validate this implication of the model, Figure 4 plots the EQT-Rice "sharing rate" against the rate implied by the model before and after the merger.[42] After the merger, sharing frictions are assumed to be zero. Model fit is reasonably good: the mean absolute error in sharing rates is 4.7% for the pre-period and 4.2% for the post-period. This provides suggestive evidence that counterfactuals which eliminate sharing frictions offer a reasonable prediction of firm behavior under integration (or efficient contracting).

Motivated by this result, Figure 5 plots the actual sharing rate for each pair of firms that traded frequently against the rate that would have been obtained if sharing frictions between that specific pair of firms were eliminated (holding fixed sharing frictions between all other pairs of firms). At the median, the actual sharing volume was 78% lower than the optimal sharing volume, and only three pairs of firms shared at more than 75% of the optimal volume. Although any pair of firms can achieve optimal levels of bilateral coordination through formal or informal (i.e., relational) contracting, these findings suggest that doing so may be difficult

[41] To illustrate the magnitude of this selection effect, Table 5 reports mean outcomes under the if the distance-minimizing allocation of wastewater; under this allocation, mean sharing frictions would be 20% greater than mean for observed shipments. From another perspective, only 49% of reuse would occur within firms absent sharing frictions. See Appendix 14.1 for further details.

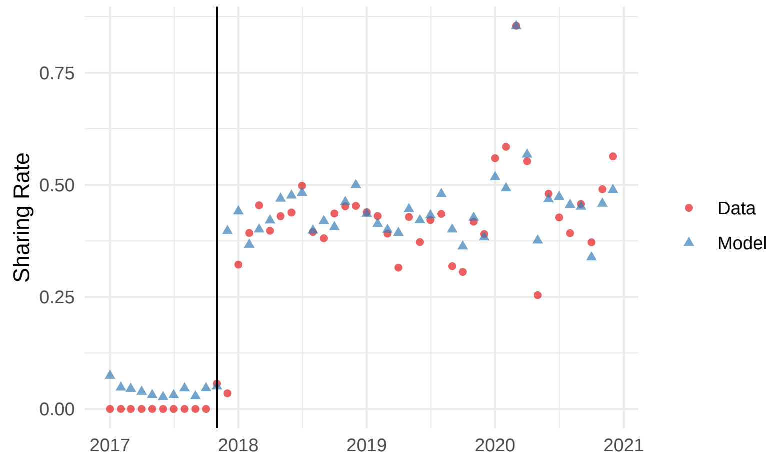
[42] For the purpose of this exercise, I define the "sharing rate" as the proportion of wastewater volumes leaving Rice or EQT well pads that were shipped to facilities within the joint entity previously affiliated with the other party. Since EQT and Rice never shared, a firm-pair specific fixed effect π_b cannot be estimated before the merger, so π_b in this case is estimated from one the aggregate categories (specifically, the category encompassing shipments between large firms in southwestern Pennsylvania).

Table 5: Key parameter estimates (in miles)

	Est	SE	\$/bbl
Mean $\phi_{\kappa\delta}$			
weighted by data	125.7	0.072	5.71
weighted by benchmark	154.2	0.081	7.01
Standard deviation $\phi_{\kappa\delta}$			
weighted by data	48.1	0.090	2.18
weighted by benchmark	36.3	0.039	1.65
Sharing market cost shifters α			
rival $\times$ poor $\rightarrow$ good env record	-	-	-
rival $\times$ good $\rightarrow$ poor env record	8.5	0.110	0.39
rival $\times$ gel $\rightarrow$ slickwater	-28.6	0.103	-1.30
rival $\times$ slickwater $\rightarrow$ gel	85.3	2.996	3.88
rival $\times$ large κ $\rightarrow$ well pad	-	-	-
rival $\times$ large κ $\rightarrow$ CTF	25.2	0.044	1.15
rival $\times$ small κ $\rightarrow$ well pad	4.4	0.151	0.20
rival $\times$ small κ $\rightarrow$ CTF	29.6	0.261	1.35
Within-firm cost shifters β			
gel $\rightarrow$ slickwater	6.7	0.092	0.31
slickwater $\rightarrow$ gel	-8.7	0.046	-0.39
small κ $\rightarrow$ CTF	-5.7	0.129	-0.26
$\sigma_\kappa + \sigma_\delta$	22.5	0.006	1.02
Mean distance (sharing market)			
weighted by data	43.6	-	1.98
weighted by benchmark	24.5	-	1.12

Notes: SE indicates the MLE standard error. Point estimates are converted into dollars per barrel (\$/bbl) under the assumption that marginal transportation costs are \$5/mile and that each water-hauling truck holds a full capacity of 110 barrels. The "benchmark" refers to distance-minimizing allocation. Note that $u_\kappa^{\mathcal{I}}$ and $u_\delta^{\mathcal{I}}$ are not reported, and that κ- and δ-specific covariates in x (such as small well pad or CTF indicators) are not separately identified from $u_\kappa^{\mathcal{I}}$ and $u_\delta^{\mathcal{I}}$.

Figure 4: Monthly sharing rate across EQT-Rice firm border



in practice.

6.2 Sources of sharing frictions

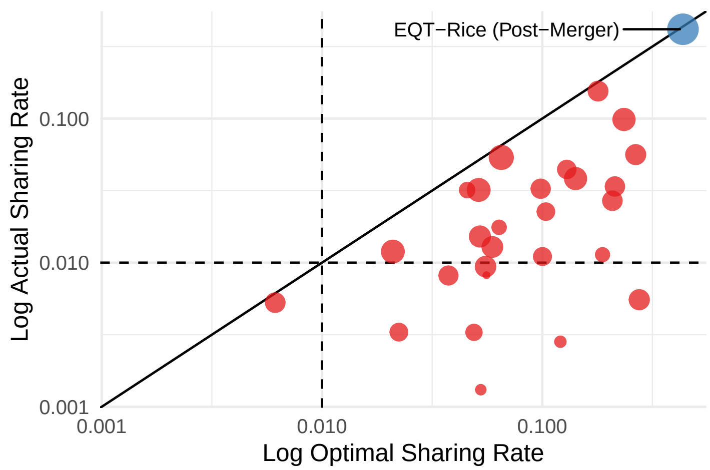
The estimated sharing frictions exhibit significant heterogeneity: the standard deviation across transactions is approximately \$2.18 per barrel, about a third of the mean. In this subsection, I leverage this heterogeneity in order to explore potential sources of sharing frictions.

6.2.1 Wastewater composition

Certain types of wastewater may be less well suited to reuse than others, either because they require more treatment or because they create risks to well productivity. Fracking fluids typically belong to one of two broad categories: "gels" and "slickwaters." The distinction between the two is not always sharp, but certain constituents commonly found in slickwater formulations can inhibit the formation of gels in gel-based formulations (Montgomery, 2013; Walsh, 2013). Using the FracFocus data, I construct an indicator for whether a given well was likely to have been fracked with a gel or a slickwater-based fracking fluid, which I infer from the disclosure of certain key chemicals.[43] I construct a similar measure for wells subsequently completed with wastewater shipped to facility δ used gels or slickwaters (aggregating across a firms' wells pads in the case that facility δ is a CTF). The estimates imply that sharing frictions are more than \$5.18 per barrel greater when slickwaters are used at the sending well pad and gels are used at the receiving well pad than in the opposite case. One potential

[43] Guar in the case of gels, and acrylamide in the case of slickwaters (Montgomery, 2013).

Figure 5: Pairwise no-friction sharing rates vs. data (in logs)



Notes: Each circle represents a firm pair. The x axis indicates the log of the sharing right implied if there were no frictions between a particular pair of firms, ceteris paribus. The y axis indicates the log of the actual sharing rate for the same firms. The size of each circle corresponds to the actual sharing volume (distinct from the sharing rate). Note that EQT-Rice pre-merger is not included, since there was no sharing (implying that the log actual sharing rate does not exist). Aside from EQT-Rice, I include only the 27 firm pairs for which I obtained an estimate of π_b.

explanation is that sharing frictions may be greater when risks to well productivity are elevated.

6.2.2 Inter-operator environmental liability

When a firm shares wastewater with another firm, it assumes some non-contractible risk of civil liability, regulatory fines, or reputational damage if an adverse event (such as a spill) occurs at the recipient's facility. Using DEP records, I construct a measure of firms' environmental records using the relative rates of spills and well site inspection failures (see Appendix 10.2 for details). If inter-operator liability concerns are significant, then sharing frictions should be greater when the firm affiliated with δ has a poor environmental record. Consistent with this hypothesis, I find that sharing frictions are \$0.39 per barrel greater when the firm at κ has a good environmental record and firm at δ has a poor one than in the opposite case.

6.2.3 Scheduling frictions

The patterns in Figure 5 are consistent with a model in which opportunistic ex post coordination occurs regularly, but surplus-maximizing ex ante coordination occurs rarely. Ex ante coordination can be relatively difficult for at least two reasons. First, barriers to information sharing (e.g., secrecy) can prevent firms from anticipating the gains from future coordination. Second, even if gains from trade are known to both parties, firms may not be able to credibly commit to future wastewater deliveries, since ex post defection to internal reuse or the spot market is possible. In the literature on relational contracting, the first problem is known as the clarity problem while the second is known as the credibility problem Gibbons and Henderson (2012).

6.3 Additional results

This section briefly summarizes model fit, additional parameter estimates, and alternative specifications.

6.3.1 Model fit

The first two rows of Table 8 indicate that the model provides a close fit for the aggregate sharing rate and the mean shipment distance observed in the data. Figure 13 presents several additional model fit diagnostics at the firm and monthly level. Panels (a) and (c) indicate predicted and observed sharing market participation rates, both as sender and receiver. Panels (b) and (d) indicate the corresponding mean shipment distances for sharing transactions. Overall, the model fits the data reasonably well, largely on account of the flexibility of the fixed effects $u_\kappa^{\mathcal{I}}$ and $u_\delta^{\mathcal{I}}$. The final row shows the time series of the sharing rate and aggregate mean shipment distance. This provides visual evidence that model fit is similar throughout the sample period.

6.3.2 Non-transportation costs within the firm

Table 7 reports point estimates for within-firm cost shifters $x_{\kappa\delta}$. In general, κ- and δ-specific characteristics are not separately identified from the fixed effects in (5). Consequently, I only consider a small number of covariates that interact characteristics of κ with characteristics of δ: two pertaining to wastewater composition, and one pertaining to facility type. Costs are \$0.26 lower per barrel of wastewater when the sending well pad is small and the receiving facility is a CTF. Surprisingly, costs are \$0.70 greater per barrel of wastewater when shipping from a "gel" origin to a "slickwater" destination than in the opposite case, which is the reverse of what one would expect if wastewater from gel-based completions requires more intensive treatment. The magnitude of these costs are small in comparison to the distribution of transportation costs across potential shipments.

6.3.3 Latent costs

The point estimate for $\sigma_K + \sigma_D$ indicates how much of the total match surplus is created by matching on latent costs rather than systematic costs.[44] The point estimate of 22.5 miles implies that the standard deviation of latent costs across potential shipments is between 20.4 and 28.8 miles. Two simple counterfactuals provide a clearer sense for the relative significance of latent costs in comparison to distance-related costs. Table 8 indicates the mean shipment distance as $\sigma_K + \sigma_D \to 0$ and as $\sigma_K + \sigma_D \to \infty$. In the first case, the mean shipment distance decreases by 13%. This implies that heterogeneity in latent costs has the same effect on transportation efficiency as increasing all distances by 15%. In the second case, the mean shipment distance increases by almost 500%. Comparison of these results implies that observed match is far from random. Thus, latent cost heterogeneity is an important and empirically significant component of firms' private costs, but the systematic component of costs $r_{\kappa\delta}$ is nevertheless the primary source of match surplus.

[44] Note that my preferred estimation approach does not separately identify σ_K and σ_D (separate identification is not needed for the counterfactuals of interest). Moreover, without specifying the shares of the outside options, I cannot directly construct $\mathcal{E}(\mu)$ (although I know $\mathcal{E}(\mu) - \mathcal{E}(\mu')$ for matches μ and μ', which is sufficient for the welfare analysis).

6.3.4 Alternative specifications

Table 6 reports the main coefficient estimate, log likelihood, and other model fit information for several additional model specifications. The main results correspond to specification (2) in the table. Specification (1) is a simplified model with no cost shifters x, no firm-pair fixed effects π, and only a constant included in z. Relative to this model, my preferred specification significantly improves model fit, as evidenced by improvements in log likelihood and the other reported model fit statistics.[45] In the main specification $d_{\kappa\delta}$ is represented by the over-the-road distance in miles between κ and δ. Because distance $d_{\kappa\delta}$ is the primary source of cost variation in the model, I consider three alternative specifications for $d_{\kappa\delta}$: over-the-road driving time in hours, the log of the over-the-road distance in miles, and a non-linear representation of over-the-road distance (specifically, a series of indicators for 30-mile increments). The specification with over-the-road driving time provides the best model fit. However, because the difference in model fit between over-the-road distance and over-the-road driving time is small, I prefer the former, which is better suited to the policy analysis. Finally, it is natural to suppose that sharing frictions differ with distance. z does not include distance in the main specification, since the relevant variation should already be captured in the firm-pair fixed effects π_b. To validate this modeling choice, specifications (3) and (4) incorporate distance into z linearly and non-linearly, respectively. This results in little improvement in model fit.

7 External costs

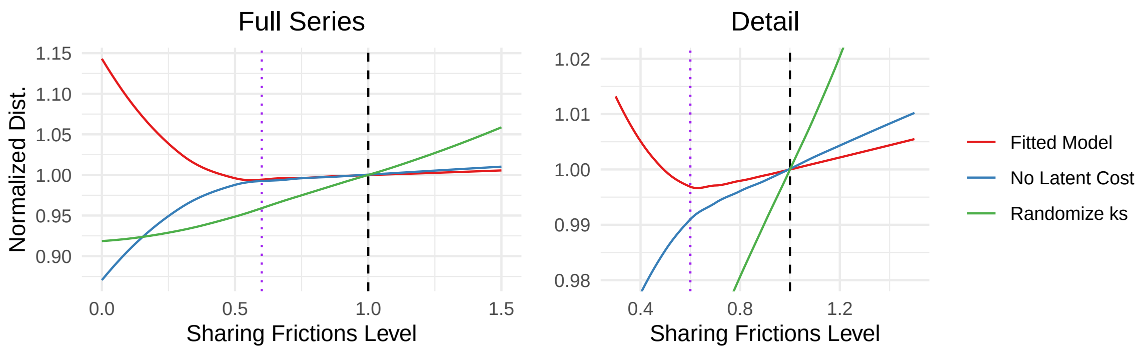
In this section, I analyze the relationship between sharing frictions and industry transportation efficiency, and hence the level of external costs from wastewater transportation.

The red line in Figure 6 shows the change in the mean shipment distance under proportional scaling of the estimated sharing frictions. The estimated level is indicated by the dashed black vertical line. For small reductions in sharing frictions, the mean shipment distance falls slightly, attaining a maximal reduction of less than 0.3% when sharing frictions are approximately 40% below the estimated level (indicated by the dotted purple line). For larger reductions, the mean shipment distance increases.

In the absence of sharing frictions, the mean shipment distance would be 15% greater than under the status quo, not lower as might have been expected. This occurs due to a combination of two features of the market: the non-random distribution of firms' locations on the one hand (depicted in Figure 1), and the significance of costs unrelated to distance

[45] Table 7 presents the estimates from Table 5 side-by-side with estimates from Specification (1), which yields a similar estimate for the mean friction.

Figure 6: Shipment distance as a function of sharing frictions



Notes: Shows change shipment distance under proportional scaling of all sharing frictions. The black vertical line indicates the estimated level of the parameters (without scaling); the purple vertical line indicates the scale at which shipment distances are minimized in the fitted model. For each series, shipment distances are normalized to the level implied by the parameter estimates.

on the other. Absent either of these factors, eliminating sharing frictions altogether would instead tend to reduce shipment distances. The blue line in the figure indicates the mean shipment distance at the estimated parameters if the scale of latent costs $\sigma_K + \sigma_D$ were assumed to be zero. The green line indicates the mean shipment distance if the ownership of each well pad κ were randomly re-assigned. Setting sharing frictions to zero reduces the mean shipment distance by 13% in the first case and 8% in the second. Thus, the main finding would reverse if firms were geographically dispersed or if costs unrelated to distance were relatively less important. Under the status quo, eliminating sharing frictions increases matching on transportation and non-transportation cost simultaneously. Because firms are geographically clustered, the elimination of sharing frictions leads a to a resorting that results in relatively more shipments at greater distances.

Although the implication that eliminating sharing frictions could increase shipment distances is counterintuitive, the significance of the underlying mechanisms can be seen in the case of EQT-Rice merger. Although unrealized ex ante geographic complementarities were presumably large, I find that the elimination of sharing frictions in particular generated only a modest improvement in transportation efficiency. The lower panel of Table 8 compares the actual and fitted post-merger shipment patterns to those that would have prevailed in a counterfactual in which sharing frictions had persisted (i.e., if the merger had not occurred), holding facility-level supply and demand for wastewater fixed at the observed post-merger levels. In the fitted model, the mean shipment distance within the EQT-Rice joint entity fell by 3.7% after the merger.[46] In the counterfactual with no merger, shipment distances would have fallen by 2.7%. Thus, the elimination of sharing frictions alone explains less than a third of the predicted reduction in shipment distances, while the rest is attributable to changes in the location of wastewater generation and reuse, which are taken to be exogenous.[47]

These results have a few immediate policy implications. First, market design improvements that aim to make sharing easier may not lead to large reductions in trucking-related external costs, and may even lead to increases in these costs. Even if it were possible to scale all sharing frictions by the optimal common factor, doing so would only lead to a modest reduction in shipment distance in comparison to the scale of potential increases. This highlights that environmental regulators face a quantity-quality tradeoff when it comes to sharing. A regulator concerned about trucking-related external costs should not promote all sharing indiscriminately, but only the types of sharing for which private and social costs are well aligned; otherwise, policy-induced sharing could exacerbate external costs.

8 Optimal regulation

In this section, I derive and analyze the optimal allocation of wastewater, emphasizing how this allocation depends on the welfare-relevance of the sharing frictions. Under the status quo, firms fail to internalize the external costs of trucking. Limited participation in the sharing market due to sharing frictions can also be interpreted as a market imperfection, depending on the nature of transaction costs. Optimal regulation must correct for both market imperfections at once.

8.1 Pigouvian taxation

As discussed in Section 5.1.1, the welfare-relevance of the sharing frictions is not identified. Let $s \in [0,1]$ index the welfare-relevance of $\phi_{\kappa\delta}$ (for the purpose of illustration, I assume that s does not depend on $\kappa\delta$). This means that for each $\kappa\delta$, the social planner should internalize $s\phi_{\kappa\delta}$ but not $(1-s)\,\phi_{\kappa\delta}$. Suppose the social planner observes the estimated cost parameters, the marginal external cost of trucking γ, and the welfare relevance parameter s. Holding

[46] In the data (as opposed to the fitted model), the mean shipment distance within the EQT-Rice joint entity fell by 4.2% after the merger. Note that this represents a subset of all wastewater shipments leaving EQT and Rice well pads. Across all shipments, the mean shipment distance fell by 18%, primarily as a result of reduced shipments to third party CTFs and other rivals.

[47] I do not predict how Rice and EQT's drilling activity would have evolved absent the merger. Nevertheless, this finding demonstrates that integration can entail many changes in economic activity aside from the elimination of sharing frictions, which fall outside the scope of the model. Thus, the elimination of sharing frictions should not be interpreted as a "merger-to-monopoly."

fixed the quantities $(\mathbf{Q}, \mathbf{C})$, a Pigouvian match μ_s^* in state s solves:

$$\min_{\mu \in \mathcal{M}(\mathbf{Q},\mathbf{C})} \gamma \times \left\{ \sum_{\kappa\delta} \mu_{\kappa\delta} d_{\kappa\delta} \right\} + C_s(\mu) \tag{10}$$

where $C_s(\mu)$ represents the welfare-relevant component of private costs under μ in state s. In Appendix 13, I show that the optimal match μ_s^* is implemented by Pigouvian tax on truck-miles where the optimal tax (or subsidy) on shipments between κ and δ in state s is:

$$tax_{\kappa\delta}^{(s)} = \gamma - (1-s) \, d_{\kappa\delta}^{-1} \phi_{\kappa\delta} \tag{11}$$

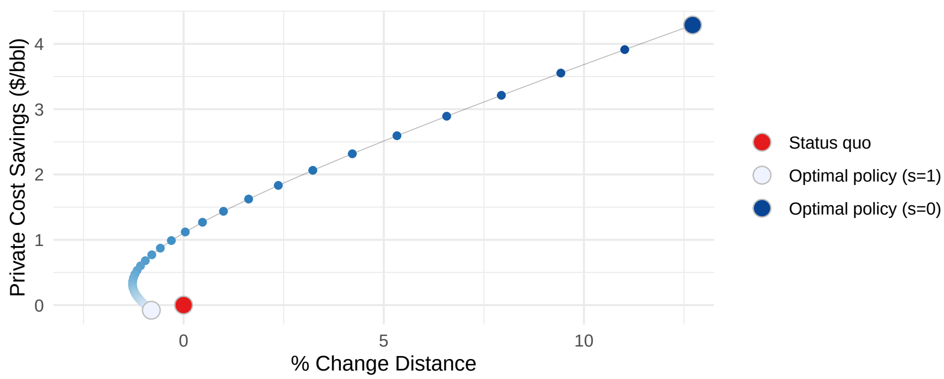
If sharing frictions are fully welfare-relevant ($s = 1$), the social planner implements a uniform tax equal to the marginal external cost of trucking γ. If sharing frictions are not fully welfare-relevant, the social planner augments this uniform tax with subsidies to incentivize greater levels of sharing, possibly leading to net subsidies on some shipment paths. The expression in (11) implies that subsidies are greater when sharing frictions are greater, and when shipment distances are smaller.

The optimal policy does not necessarily lead to a reduction in shipment distance. If (11) is implemented, the sum of the external costs of trucking and firms' private costs (the objective in (10)) is weakly lower. However, depending on the value of s, one of either external costs or firms' private costs can increase. To illustrate magnitudes, I calibrate $\gamma = 0.07$ based on the calculation in Section 2.4. Figure 7 shows the percentage change in trucking distance and the difference in firms' private costs relative to the baseline (in dollars per barrel) under the optimal policy for different values of s. When $s = 1$, the optimal policy reduces mean shipment distances by 0.8%, while increasing firms' private costs by about \$0.08 per barrel. When $s = 0$, the optimal policy reduces firms' private costs by \$4.29 per barrel, while increasing mean shipment distances by 12.7%. Moreover, the relationship between s and mean shipment distances is non-monotonic.

8.2 Pigouvian taxation under ambiguity

Implementation of (11) requires knowledge of s, but in practice a social planner may not have access to this information. The form of (11) implies that firms have an incentive to overstate s in order to earn larger subsidy payments. Incorrect inference of s can lead to large welfare losses. To illustrate the scale of potential policy regret, Figure 14 shows the producer cost savings and social cost savings (net of tax revenue) with full subsidies $tax_{\kappa\delta}^{(0)}$ relative to the uniform tax $tax_{\kappa\delta}^{(1)}$ alone, assuming that the marginal cost of public funds is zero ($\lambda = 0$).

Figure 7: Distance vs. private cost savings under optimal policy



Notes: Shows private cost savings per barrel vs. the percentage change in shipment distance relative to the status quo under the optimal (Pigouvian) tax and subsidy scheme. Note that the level of private costs is not identified, so I report a difference in levels on the y-axis.

If sharing frictions are welfare-relevant ($s = 1$) but the full subsidy is implemented anyway, social costs increase by $2.34 per barrel even for $\lambda = 0$.

To formalize the notion that the social planner may not be able to infer s, suppose the social planner faces Knightian (i.e., unquantifiable) uncertainty over s. Faced with this form of uncertainty, it is natural to consider maxmin policies across the set of feasible states $\mathcal{S} = [0, 1]$. A robust Pigouvian allocation $\mu^*_{\mathcal{S}}$ is the solution to a maxmin problem:

$$\max_{\mu \in \mathcal{M}(Q,C)} \min_{s \in \mathcal{S}} -\gamma \times \left\{ \sum_{\kappa\delta} \mu_{\kappa\delta} d_{\kappa\delta} \right\} - C_s(\mu)$$

For the estimated cost parameters, it is straightforward to show that $\mu^*_{\mathcal{S}}$ is the allocation obtained under the uniform tax $tax^{(1)}$.[48] In this sense, the uniform tax $tax^{(1)}$ may be more robust than the adjusted tax $tax^{(s)}$ or similar tax rules. Moreover, even if s were known, a uniform tax is simpler to implement than a complicated tax-and-subsidy scheme tax: the regulator need not determine $\phi_{\kappa\delta}$ for each potential shipment, and other administrative costs may be lower than in the case of a non-uniform tax.

[48]This is the case for any cost parameters such that social costs are strictly decreasing in s. A proof of this claim is provided in Appendix 13. The proof follows from an application of the minimax theorem, exploiting the concavity of $\mathcal{E}$.

9 Conclusion

The oil and gas industry's freshwater usage and wastewater management practices create significant local environmental impacts in oil and gas producing states. In many oil and gas plays, low costs of entry and lack of product differentiation lead to fragmented market structures in which sharing frictions may exacerbate these impacts. I find robust evidence that sharing frictions are economically large. Developing policy responses to reduce wastewater sharing frictions can potentially deliver immediate benefits in Pennsylvania and may prove useful in other settings, particularly if fracking becomes more widespread outside the United States in the future. The model estimates suggest that policy responses which focus on contracting frictions (for example, clarification of liability rules) may be more effective than other approaches (for example, those that focus on search). In Pennsylvania specifically, the benefits of any such interventions are likely modest at best, and could even be negative due to the geographic configuration of firms.

The finding that interfirm transaction costs are empirically significant speaks to a long literature on the theory of the firm (originating in Coase (1937)). Using a novel structural approach, I find clear evidence that integration can eliminate transaction costs. This approach also allows me to demonstrate that interfirm transaction costs exhibit significant heterogeneity even within a narrowly defined market, in large part due to heterogeneity in contracting frictions across transactions. Even though transaction costs are low for some transactions, few firms are able to obtain surplus-maximizing allocations through contracting alone. Depending on the welfare-relevance of transaction costs, policy interventions to mitigate sharing frictions may be justified on welfare grounds independently of any environmental benefits.

The policy analysis can be extended in a few ways. First, with improved data it would be straightforward to quantify the impact of sharing frictions on injection well disposal rates and freshwater withdrawals. Freshwater withdrawals in particular are an important issue in shale basins in the western United States, and the framework in this paper can be used to model regulation in this context. Second, it would be valuable to link sharing frictions to firms' drilling decisions in order to understand how sharing frictions affect shale gas production in general. I leave these questions for future research.

References

Abramzon, S., C. Samaras, A. Curtright, A. Litovitz, and N. Burger (2014) "Estimating the Consumptive Use Costs of Shale Natural Gas Extraction on Pennsylvania Roadways," *Journal of Infrastructure Systems*, 20 (3).

Anderson, J.E. and E. van Wincoop (2004) "Trade Costs," *Journal of Economic Literature*, 42 (3), pp. 691–751.

Antrás, P. and D. Chor (2022) "Global Value Chains," *Handbook of International Economics*, pp. 297–376, eds. G. Gopinath, E. Helpman, and K. Rogoff. Elsevier.

Atalay, E., A. Hortaçsu, C. Syverson, and M.J. Li (2019) "How Wide is the Firm Border?" *Quarterly Journal of Economics*, 134 (4), pp. 1845–1882.

Berry, S., J. Levinsohn, and A. Pakes (1995) "Automobile Prices in Market Equilibrium," *Econometrica*, 63 (4), pp. 841–890.

Black, K.J., A.J. Boslett, E.L. Hill, L. Ma, and S.J. McCoy (2021) "Economic, Environmental, and Health Impacts of the Fracking Boom," *Annual Review of Resource Economics*, 13, pp. 311–334.

Bonnet, O., A. Galichon, Y. Hsieh, K. O'Hara, and M. Shum (2022) "Yogurts Choose Consumers? Estimation of Random-Utility Models via Two-Sided Matching," *The Review of Economic Studies*, 89 (6), pp. 3085–3114.

Bresnahan, T.F. and J.D. Levin (2012) "Vertical Integration and Market Structure," NBER Working Paper.

Buchanan, J.M. and W.C. Stubblebine (1962) "Externality," *Economica*, 29 (116), pp. 371–384.

Carlton, D.W. (2020) "Transaction costs and competition policy," *International Journal of Industrial Organization*, 73.

Chassang, S. (2010) "Building Routines: Learning, Cooperation, and the Dynamics of Incomplete Relational Contracts," *American Economic Review*, 100 (1), pp. 448–465.

Chiong, K.X., A. Galichon, and M. Shum (2016) "Duality in dynamic discrete-choice models," *Quantitative Economics*, 7, pp. 83–115.

Choo, E. and A. Siow (2006) "Who Marries Whom and Why," *Journal of Political Economy*, 114 (1), pp. 175–201.

Clay, K., A. Jha, N. Muller, and R. Walsh (2019) "External Costs of Transporting Petroleum Products: Evidence from Shipments of Crude Oil from North Dakota by Pipelines and Rail," *The Energy Journal*, pp. 55–72.

Coase, R.H. (1937) "The Nature of the Firm," *Economica*, 4 (16), pp. 386–405.

Conlon, C. and J. Gortmaker (2020) "Best practices for differentiated products demand estimation with PyBLP," *The RAND Journal of Economics*, 51 (4), pp. 1108–1161.

Covert, T.R. (2015) "Experiential and Social Learning in Firms: The Case of Hydraulic Fracturing in the Bakken Shale," Working Paper.

Covert, T.R. and R. Kellogg (2023) "Environmental Consequences of Hydrocarbon Infrastructure Policy," Working Paper.

Covert, T.R. and R.L. Sweeney (2022) "Secrecy Rules and Exploratory Investment: Theory and Evidence from the Shale Boom," Working Paper.

Decker, C., E.H. Lieb, R.J. McCann, and B.K. Stephens (2013) "Unique equilibria and substitution effects in a stochastic model of the marriage market," *Journal of Economic Theory*, 148 (2), pp. 773–792.

Demsetz, H. (1988) "The Theory of the Firm Revisited," *Journal of Law, Economics, & Organization*, 4 (1), pp. 141–161.

EPA (2016) "Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States (Final Report)," Report of the Environmental Protection Agency.

EQT Corporation (2017) "Investor Presentation, dated as of June 19, 2017," U.S. Securities and Exchange Commission. Form 8-K, Exhibit No. 99.2, 0001104659-17-040193.

Fowlie, M.L. (2009) "Incomplete Environmental Regulation, Imperfect Competition, and Emissions Leakage," *American Economic Journal: Economic Policy*, 1 (2), pp. 71–112.

Galichon, A. and B. Salanie (2022) "Cupid's Invisible Hand: Social Surplus and Identification in Matching Models," *Review of Economic Studies*, 89 (5), pp. 2600–2629.

Ghili, S. (2022) "Network Formation and Bargaining in Vertical Markets: The Case of Narrow Networks in Health Insurance," *Marketing Science*, 41 (3), pp. 501–527.

Gibbons, R. and R. Henderson (2012) "Relational Contracts and Organizational Capabilities," *Organization Science*, 23 (5), pp. 1350–1364.

Graham, B.S. (2011) "Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers," *Handbook of Social Economics*, 1, pp. 965–1052, eds. J. Benhabib, A. Bisin, and M.O. Jackson. North-Holland.

——— (2013) "Comparative Static and Computational Methods for an Empirical One-to-one Transferable Utility Matching Model," *Advances in Econometrics*, pp. 153–181, in Structural Econometric Models (Vol. 31), Emerald Group Publishing Limited.

Graham, J., J. Irving, X. Tang et al. (2015) "Increased traffic accident rates associated with shale gas drilling in Pennsylvania," *Accident Analysis and Prevention*, 74, pp. 203–209.

Gretsky, N.E., J.M. Ostroy, and W.R. Zame (1992) "The Nonatomic Assignment Model," *Economic Theory*, 2 (1), pp. 103–127.

Grossman, S.J. and O.D. Hart (1986) "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration," *Journal of Political Economy*, 94 (4), pp. 691–719.

Groundwater Protection Council (2019) "Produced Water Report: Regulations, Current Practices, and Research Needs," Report.

Handel, B.R. and J.T. Kolstad (2015) "Health Insurance for 'Humans': Information Frictions, Plan Choice, and Consumer Welfare," *American Economic Review*, 105 (8), pp. 2449–2500.

Harris, A. and T.M.A. Nguyen (2023) "Long-Term Relationships and the Spot Market: Evidence from US Trucking," *Working Paper*.

Hausman, C. and R. Kellogg (2015) "Welfare and Distributional Implications of Shale Gas," *Brookings Papers on Economic Activity*, pp. 71–125.

Head, K. and T. Meyer (2014) "Gravity Equations: Workhorse, Toolkit, and Cookbook," *Handbook of International Economics*, 4, pp. 131–195, eds. G. Gopinath, E. Helpman, and K. Rogoff. Elsevier.

Heo, J., P.J. Adams, and H.O. Gao (2016) "Reduced-form modeling of public health impacts of inorganic PM2.5 and precursor emissions," *Atmospheric Environment*, 137, pp. 80–89.

Ho, K. and R.S. Lee (2019) "Equilibrium Provider Networks: Bargaining and Exclusion in Health Care Markets," *American Economic Review*, 109 (2), pp. 473–522.

Hodgson, C. (2022) "Trade-ins and Transaction Costs in the Market for Used Business Jets," *American Economic Journal: Microeconomics*, 15 (4), pp. 350–391.

Horn, H. and A. Wolinsky (1988) "Bilateral Monopolies and Incentives for Merger," *The RAND Journal of Economics*, 19 (3), pp. 408–419.

Hubbard, T.N. (2001) "Contractual Form and Market Thickness in Trucking," *The RAND Journal of Economics*, 32 (2), pp. 369–386.

Idel, M. (2016) "A review of matrix scaling and Sinkhorn's normal form for matrices and positive maps," arXiv:1609.06349v1.

Kellogg, R. (2011) "Learning by Drilling: Interfirm Learning and Relationship Persistence in the Texas Oilpatch," *The Quarterly Journal of Economics*, 126 (4), pp. 1961–2004.

Kondash, A.J., N.E. Lauer, and A.V. Vengosh (2018) "The intensification of the water footprint of hydraulic fracturing," *Science Advances*, 4 (8).

Leslie, G. (2018) "Tax induced emissions? Estimating short-run emission impacts from carbon taxation under different market structures," *Journal of Public Economics*, 167, pp. 220–239.

Luxen, D. and C. Vetter (2011) "Real-time routing with OpenStreetMap data," pp. 513–516, in Proceedings of the 19th ACM SIGSPATIAL International Conference on Advances in Geographic Information Systems. Association for Computing Machinery.

Macchiavello, R. and A. Morjaria (2015) "The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports," *American Economic Review*, 105 (9), pp. 2911–2945.

——— (2021) "Competition and Relational Contracts in the Rwanda Coffee Chain," *The Quarterly Journal of Economics*, 136 (2), pp. 1089–1143.

MacKay, A. (2022) "Contract Duration and the Costs of Market Transactions," *American Economic Journal: Microeconomics*, 14 (3), pp. 164–212.

Maloney, K.O., S. Baruch-Mordo, L.A. Patterson et al. (2017) "Unconventional oil and gas spills: Materials, volumes, and risks to surface waters in four states of the U.S.," *Science of the Total Environment*, 2017 Mar 1;581-582:369-377.

Mansur, E.T. (2007) "Do Oligopolists Pollute Less? Evidence from a Restructured Electricity Market," *The Journal of Industrial Economics*, 55 (4), pp. 661–689.

Masten, S.E., J.W. Meehan, and E.A. Snyder (1991) "The Costs of Organization," *Journal of Law, Economics, & Organization*, 7 (1), pp. 937– 976.

McCurdy, R. (2011) "Underground Injection Wells For Produced Water Disposal," in Proceedings of the Technical Workshops for the Hydraulic Fracturing Study: Water Resources Management. EPA.

Menefee, A.H. and B.R. Ellis (2020) "Wastewater management strategies for sustained shale gas production," *Environmental Research Letters*, 15 (2), 024001.

Montgomery, C. (2013) "Fracturing Fluid Components," *Effective and Sustainable Hydraulic Fracturing*, eds. A.P. Bunger, J. McLennan, and R. Jeffrey. IntechOpen.

Muehlenbachs, L., S. Staubli, and Z. Chu (2021) "The Accident Externality from Trucking: Evidence from Shale Gas Development," *Regional Science and Urban Economics*, 88, 103630.

Newell, R.G. and D. Raimi (2014) "Implications of Shale Gas Development for Climate Change," *Environmental Science and Technology*, 48 (15), pp. 8360–8368.

Preonas, L. (2023) "Market Power in Coal Shipping and Implications for U.S. Climate Policy," *The Review of Economic Studies*, rdad090.

Samuelson, L. and E. Stacchetti (2017) "Even up: Maintaining relationships," *Journal of Economic Theory*, 169, pp. 170–217.

Shapley, L.S. and M. Shubik (1971) "The assignment game I: The core," *International Journal of Game Theory*, 1, pp. 111–130.

Small, M.J., P.C. Stern, E. Bomberg et al. (2014) "Risks and Risk Governance in Unconventional Shale Gas Development," *Environmental Science and Technology*, 48 (15), pp. 8289–8297.

Vreugdenhil, N. (2023) "Booms, Busts, and Mismatch in Capital Markets: Evidence from the Offshore Oil and Gas Industry," Working Paper.

Walsh, J.M. (2013) "Water Management for Hydraulic Fracturing in Unconventional Resources: Part 1," https://jpt.spe.org/water-management-hydraulic-fracturing-unconventional-resourcespart-1, Magazine Article, Journal of Petroleum Technology.

Williamson, O.E. (1971) "The Vertical Integration of Production: Market Failure Considerations," *American Economic Review*, 61 (1971), pp. 112–123.

Wilson, R. (2022) "Moving to economic opportunity: the migration response to the fracking boom," *Journal of Human Resources*, 57 (3), pp. 918–955.

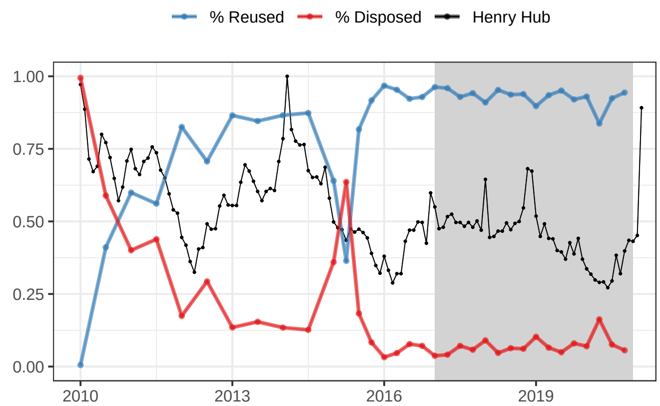
Wunz, E.C. (2015) "Summary of Shale Gas Wastewater Treatment and Disposal In Pennsylvania: 2014," Report, dated August 20.

Xu, M. and Y. Xu (2020) "Fraccidents: The impact of fracking on road traffic deaths," *Journal of Environmental Economics and Management*, 101.

10 Data preparation

The main dataset consists of Oil and Gas Well Waste Reports collected from the Pennsylvania Department of Environmental Protection web site. For the main analysis, I consider waste reports for all unconventional wells and for all production periods between January 2017 and December 2020. This choice of analysis period reflects the fact that the waste reporting format was modified in 2017 to consistently indicate the location of reuse. I choose to retain data from the Covid pandemic period. Although drilling rates in general fell during this period, the demand for disposal did not, and overall reuse rates remained relatively stable, as evidenced by Figure 8.

Figure 8: Pennsylvania wastewater reuse over time

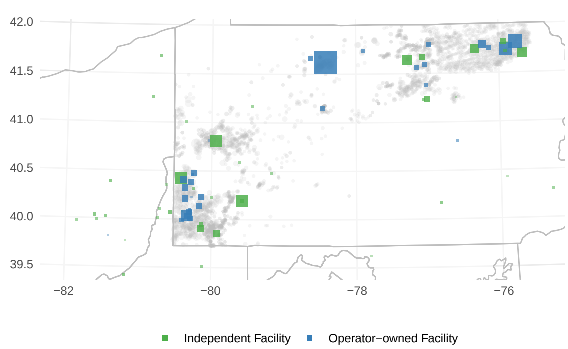


Notes: The red and blue lines indicate the share of wastewater shipments in the data for which the reported destination was a site at which only disposal could have occurred (primarily injection wells), or a site at which reuse could have occurred. The black line indicates the spot price of natural gas. The sample period for this analysis is highlighted in gray.

Operators are required to report disposal method for various waste products, including solids such as drill cuttings and shredded containment liners. I rely on the classifications from Wunz (2015) as well as internet research on the functions performed at different waste facilities (e.g., landfills vs. injection wells) to identify presumably reusable wastewater. This procedure is inevitably imperfect. Reporting errors are possible, and not all liquid waste in fact represents reusable wastewater. In particular, sludges produced as a byproduct of the treatment process are in some cases disposed via injection well. Although these volumes are presumably small in comparison to the volumes of reusable wastewater, my preferred approach to estimation avoids relying on injection well rates to avoid data contamination, as discussed in the main text.

As described in the main text, the waste reports do not report the dates or quantities associated with specific transfer events, but rather the aggregate quantities of different types of waste transferred from a given well to a given disposal location during a specified month. Wastewater intended for reuse can be transferred either to a CTF prior to reuse or directly to another well pad for reuse. These cases appear differently in the data. In the former case, it is not possible to identify the ultimate location of reuse. However, whether the treatment facility is operated by the reporting firm or by a third party can be inferred from the reported permit information and facility names (although in some cases this requires consulting separate DEP resources). In the latter case, if the destination well pad is located in Pennsylvania, a numeric identifier associated with the destination well pad is also provided. I use this numeric identifier to determine whether a given amount of wastewater was transferred for internal or external reuse. In particular, I classify reuse locations as internal or external depending on whether the reporting firm is currently listed as an operator for any well at the destination well pad (in a separate DEP data source). If the destination well pad is located outside of Pennsylvania (primarily in West Virginia), no such identifier is provided, and I do not attempt to infer the ownership of the destination well pad. I identify firms by their DEP OGO Number (where OGO is an acronym for "Oil and Gas Operator"). I rely on press releases and changes in the data over time to account for changes in ownership over time (the Rice-EQT merger was the most significant but not the sole merger during the sample period). It is rare for multiple operators to be associated with the same well pad, but when this is the case I treat the well pad as "internal" for both parties.

Figure 9: Centralized treatment facility locations



Typically several wells are located at a single well pad, which encompasses common infrastructure such as access roads and storage tanks. Technically operators are required to report waste information on a well-by-well basis, but because wastewater is often stored in a single location on the pad most simply report well pad-level averages. Therefore I focus on the well pad rather than the well as my primary unit of analysis. I infer the number of shipments in a month by dividing the total volume by the capacity of a typical water hauling truck.[49] To mitigate the impact of data reporting errors, I winsorize shipment volumes at the 99.9%-tile (about 77,000 barrels, or 600-700 truckloads).

10.1 Wastewater quality measures

First, I link all unique fracking events in FracFocus to well pads using well API numbers. A fracking event includes "guar" if any of the listed ingredients has a chemical name that contains "guar." Likewise, a fracking event includes "acrylamide" if any of the listed ingredients has a chemical name that contains "acryl." A well pad is a "recent guar" well pad if a fracking event involving guar was completed in the previous six months. Likewise, a well pad is a "recent acryl" well pad if a fracking event involving acrylamide was completed in the previous six months. These are the indicators that I include in the regression. If δ is a CTF, I take a volume-weighted average of the indicators for well pads operated by the operator affiliated with δ.

10.2 Firm environmental record index

The PA DEP maintains facility-level compliance records for all oil and gas wells. These data include routine inspections data and incident data. When violations are found or incidents occur, the DEP assigns a detailed violation code describing the nature of the event and the relevant regulatory statutes. Using these codes, I classify violations and incidents into four categories: (1) pollution or other waste mishandling (including spills); (2) tank and impoundment failures; (3) erosion and sedimentation problems; (4) well mechanical integrity failures. I focus on the first three categories, which are most directly relevant to wastewater handling. I first tabulate the total number of (unique) violations for each firm during the sample period. Then, using the FracFocus data, I tabulate the number of fracking events for each firm during the same period. I regress the log number of violations on the log number of fracking events for each firm. A firm has a "poor environmental record" if its residual in this regression is greater than zero; otherwise, a firm has a "good environmental record." Note that I only construct this measure for firms with more than 24 observed fracking events in the sample period. I handle all other firms as a separate category when performing estimation, but for clarity of exposition these coefficients are not reported in the main text.

[49] I assume that this is 110 barrels (the modal volume), although truck capacities range from around 80 to around 130 barrels. Line items in the data are frequently reported in integer multiples of a truck capacity in this range.

10.3 External costs of wastewater trucking

In the main text, I state that unpriced carbon emission and air pollution externalities are roughly 7% of private transportation costs. This figure is constructed as follows. A full water-hauling truck weighs about 40 tons, while an empty one weighs about 20 tons. If all trucks return from each load empty, then wastewater management generates about 1.0 metric tons of PM2.5 emissions, 80,000 metric tons of carbon emissions, and 160 metric tons of NOx emissions per year.[50] Using the EASIUR air quality model (Heo et al., 2016), the social costs of wastewater trucking-related PM2.5 and NOx emissions are approximately \$3.3M per year.[51] Using the EPA Social Cost of Carbon for 2020, wastewater-related carbon emissions are approximately \$3.4M per year. Thus, the social costs of air pollution and greenhouse gas emissions sum to around \$0.10 per barrel. In comparison, industry trucking costs are around \$5 per mile, implying average transportation costs of \$1.35 per barrel.

11 Model details

11.1 Match entropy function

In general, the match entropy function is defined to be:

$$\mathcal{E}\left(\mu, \mathbf{Q}, \mathbf{C}\right) = -G^*\left(\mu, \mathbf{Q}\right) - H^*\left(\mu, \mathbf{C}\right)$$

where $G^*\left(\mu, n\right)$ is the generalized entropy of choice for disposal and $H^*\left(\mu, m\right)$ is the generalized entropy of choice for reuse. In particular,

$$G^*\left(\mu, \mathbf{Q}\right) = \sup_{U \in \mathbb{R}^{K \times D}} \left(\sum_{\kappa \in K} \sum_{\delta \in D} \mu_{\kappa\delta} U_{\kappa\delta} - \sum_{\kappa \in K} Q_\kappa E\left[\max_{\delta \in D_0} U_{\kappa\delta} + \epsilon_{i\delta}\right]\right)$$

and

$$H^*\left(\mu, \mathbf{C}\right) = \sup_{V \in \mathbb{R}^{K \times D}} \left(\sum_{\kappa \in K} \sum_{\delta \in D} \mu_{\kappa\delta} V_{\kappa\delta} - \sum_{\delta \in D} C_\delta E\left[\max_{\kappa \in K_0} V_{\kappa\delta} + \eta_{\kappa j}\right]\right)$$

[50] I use the average emissions factors for tanker trucks from EPA SmartWay Carrier data. This data is self-reported and may not be representative for the wastewater-hauling market specifically.

[51] To compute air pollution, I assume that all trucking-related air pollution occurs at the well site from which the wastewater originated, rather than along the trucking route. This likely results in an underestimate of air pollution damages because well pads are often located in remote areas, whereas trucking routes pass through more populous areas (such as the Pittsburgh metropolitan area).

Intuitively, $G^*(\mu, \mathbf{Q})$ and $H^*(\mu, \mathbf{C})$ quantify the amount of latent cost heterogeneity required to rationalize a given match μ conditional on the distributions of P_K and P_D. Galichon and Salanie (2022) provides an interpretation of these objects.

12 Estimation details

12.1 Derivation of the estimator

To simplify notation, suppose we observe a single month of data. We observe a random sample of truckloads within the reuse market, $\hat{\mu}_{11}, \ldots, \hat{\mu}_{KD}$ where $\sum_{\delta \in D} \hat{\mu}_{\kappa\delta} > 0$ for all κ and $\sum_{\kappa \in K} \hat{\mu}_{\kappa\delta} > 0$ for all δ. We do not observe the shares of the outside options, $\hat{\mu}_{10}, \ldots \hat{\mu}_{K0}$ or $\hat{\mu}_{01}, \ldots \hat{\mu}_{0D}$. The conditional likelihood of observing the data is:

$$L(\theta; \mathbf{Q}, \mathbf{C}) = \sum_{\kappa\delta} \hat{\mu}_{\kappa\delta} \log \left\{ \frac{\mu_{\kappa\delta}(\theta; \mathbf{Q}, \mathbf{C})}{\sum_{\kappa\delta} \mu_{\kappa\delta}(\theta; \mathbf{Q}, \mathbf{C})} \right\}$$

Galichon and Salanie (2022) Theorem 4 implies that:

$$\Delta r_{\kappa\delta} = \frac{\partial \mathcal{E}}{\partial \mu_{\kappa\delta}}$$

where:

$$\Delta r_{\kappa\delta} = d_{\kappa\delta} + x'_{\kappa\delta}\beta + z'_{\kappa\delta}\alpha + \psi_\kappa + \psi_\delta + \pi_b$$

and:

$$\mathcal{E} \equiv \max_{U,V} \left(\begin{array}{c} \sum_{\kappa,\delta} \mu_{\kappa\delta} U_{\kappa\delta} + \sum_{\kappa,\delta} \mu_{\kappa\delta} V_{\kappa\delta} \\ - \sum_{\kappa \in K} Q_\kappa E\left[\max_{\delta \in D_0} \{U_{\kappa\delta} + \epsilon_\delta\}\right] - \sum_{\delta \in D} C_\delta E\left[\max_{\kappa \in K_0} \{V_{\kappa\delta} + \eta_\kappa\}\right] \end{array} \right)$$

For the logit, we can show that:

$$\frac{\partial \mathcal{E}}{\partial \mu_{\kappa\delta}} = -(\sigma_\kappa + \sigma_\delta) \log \mu_{\kappa\delta} + \sigma_\kappa \log \mu_{\kappa 0} + \sigma_\delta \log \mu_{0\delta}$$

and hence:

$$\mu_{\kappa\delta} = \exp\left\{ -(\sigma_\kappa + \sigma_\delta)^{-1} \left(d_{\kappa\delta} + x'_{\kappa\delta}\beta + z'_{\kappa\delta}\alpha + \psi_\kappa + \psi_\delta + \pi_b - \sigma_\kappa \log \mu_{\kappa 0} - \sigma_\delta \log \mu_{0\delta} \right) \right\}$$

which is well known (see, e.g., Graham (2011)). Moreover, $\mu_{\kappa\delta}$ must satisfy the market clearing conditions:

$$\mu_{\kappa 0} + \sum_{\delta \in D} \mu_{\kappa\delta} = Q_\kappa \quad \forall\, \kappa \in K$$
$$\mu_{0\delta} + \sum_{\kappa \in K} \mu_{\kappa\delta} = C_\delta \quad \forall\, \delta \in D$$

where the marginals Q_κ and C_δ are observed in the data:

$$Q_\kappa = \frac{\hat{\mu}_{\kappa 0} + \sum_\delta \hat{\mu}_{\kappa\delta}}{\sum_{\kappa\delta} \hat{\mu}_{\kappa\delta} + \sum_\kappa \hat{\mu}_{\kappa 0} + \sum_\delta \hat{\mu}_{0\delta}}$$
$$C_\delta = \frac{\hat{\mu}_{0\delta} + \sum_\kappa \hat{\mu}_{\kappa\delta}}{\sum_{\kappa\delta} \hat{\mu}_{\kappa\delta} + \sum_\kappa \hat{\mu}_{\kappa 0} + \sum_\delta \hat{\mu}_{0\delta}}$$

Decker et al. (2013) establish that the system of market clearing equations has a unique solution in $\mu_{\kappa 0}$ and $\mu_{0\delta}$ (see also Graham (2013)). Absent estimators $\hat{\mu}_{\kappa 0}$ and $\hat{\mu}_{0\delta}$, consider the following strategy. For any sample size n, there exists a c_n such that $c_n^{-1} \sum_{\kappa\delta} \hat{\mu}_{\kappa\delta} = 1 - S_0$ where $S_0 = \sum_{\kappa \in K} \mu_{\kappa 0} + \sum_{\delta \in D} \mu_{0\delta}$ is the population mass of the outside options. Then an alternative representation of Q_κ and C_δ is:

$$Q_\kappa = \frac{\mu_{\kappa 0} + c_n^{-1} \sum_\delta \hat{\mu}_{\kappa\delta}}{c_n^{-1} \sum_{\kappa\delta} \hat{\mu}_{\kappa\delta} + \sum_\kappa \mu_{\kappa 0} + \sum_\delta \mu_{0\delta}}$$
$$C_\delta = \frac{\mu_{0\delta} + c_n^{-1} \sum_\kappa \hat{\mu}_{\kappa\delta}}{c_n^{-1} \sum_{\kappa\delta} \hat{\mu}_{\kappa\delta} + \sum_\kappa \mu_{\kappa 0} + \sum_\delta \mu_{0\delta}}$$

which is equivalent to:

$$Q_\kappa = \mu_{\kappa 0} + c_n^{-1} \sum_\delta \hat{\mu}_{\kappa\delta}$$
$$C_\delta = \mu_{0\delta} + c_n^{-1} \sum_\kappa \hat{\mu}_{\kappa\delta}$$

Substituting these expressions into the market clearing conditions gives:

$$\sum_{\delta \in D} \mu_{\kappa\delta} = c_n^{-1} \sum_\delta \hat{\mu}_{\kappa\delta}$$
$$\sum_{\kappa \in K} \mu_{\kappa\delta} = c_n^{-1} \sum_\delta \hat{\mu}_{\kappa\delta}$$

Expanding terms and re-arranging gives:

$$\sum_{\delta\in D} \exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\psi_\kappa+\psi_\delta+\pi_b-\sigma_\kappa\{\log c_n\mu_{\kappa 0}\}-\sigma_\delta\{\log c_n\mu_{0\delta}\})\right\}=\sum_\delta \hat{\mu}_{\kappa\delta}$$
$$\sum_{\kappa\in K} \exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\psi_\kappa+\psi_\delta+\pi_b-\sigma_\kappa\{\log c_n\mu_{\kappa 0}\}-\sigma_\delta\{\log c_n\mu_{0\delta}\})\right\}=\sum_\delta \hat{\mu}_{\kappa\delta}$$

Now consider the system of equations:

$$\sum_{\delta\in D} \exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\pi_b)-\tilde{u}_\kappa-\tilde{v}_\delta\right\}=\sum_\delta \hat{\mu}_{\kappa\delta} \tag{12}$$
$$\sum_{\kappa\in K} \exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\pi_b)-\tilde{u}_\kappa-\tilde{v}_\delta\right\}=\sum_\kappa \hat{\mu}_{\kappa\delta}$$

Clearly, $\tilde{u}_\kappa=(\sigma_\kappa+\sigma_\delta)^{-1}\psi_\kappa-(\sigma_\kappa+\sigma_\delta)^{-1}\sigma_\kappa\{\log c_n+\log\mu_{\kappa 0}\}$ for all κ and $\tilde{v}_\delta=(\sigma_\kappa+\sigma_\delta)^{-1}\psi_\delta-(\sigma_\kappa+\sigma_\delta)^{-1}\sigma_\delta\{\log c_n+\log\mu_{0\delta}\}$ for all δ is a solution to this system. Moreover it is easy to see that $\tilde{u}_\kappa=\alpha+(\sigma_\kappa+\sigma_\delta)^{-1}\psi_\kappa-(\sigma_\kappa+\sigma_\delta)^{-1}\sigma_\kappa\{\log c_n+\log\mu_{\kappa 0}\}$ for all κ and $\tilde{v}_\delta=-\alpha+(\sigma_\kappa+\sigma_\delta)^{-1}\psi_\delta-(\sigma_\kappa+\sigma_\delta)^{-1}\sigma_\delta\{\log c_n+\log\mu_{0\delta}\}$ all δ is also a solution for any $\alpha\in\mathbb{R}$. Indeed, provided that $\sum_\delta\hat{\mu}_{\kappa\delta}>0$ for all κ and $\sum_\kappa\hat{\mu}_{\kappa\delta}>0$ for all δ, $\tilde{u}$ and $\tilde{v}$ satisfying (12) exist and are unique up to scale, implying that all solutions take this form. This follows from results closely related to Sinkhorn's Theorem. In particular, Theorem 3.1 in Idel (2016) implies the existence of $\tilde{u}$ and $\tilde{v}$ satisfying these equations; and moreover, that $\tilde{u}$ and $\tilde{v}$ are unique up to scale.[52]

Finally, observe that:

$$\log\left(\frac{\mu_{\kappa\delta}(\theta;\mathbf{Q},\mathbf{C})}{\sum_{\kappa\delta}\mu_{\kappa\delta}(\theta;\mathbf{Q},\mathbf{C})}\right)=\log\left(\frac{\exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\pi_b)-\tilde{u}_\kappa-\tilde{v}_\delta\right\}}{\sum_{\kappa\delta}\exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\pi_b)-\tilde{u}_\kappa-\tilde{v}_\delta\right\}}\right)$$

where, by construction $\sum_{\kappa\delta}\exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\pi_b)-\tilde{u}_\kappa-\tilde{v}_\delta\right\}=\sum_{\kappa\delta}\hat{\mu}_{\kappa\delta}$. The expression in (7) is obtained by simplifying this last expression.

12.2 Model specification

As in the case of gravity models, κ and δ-specific covariates have no effect on the equilibrium match under the maintained assumptions.[53] Therefore, it is only necessary to include covari-

[52]To apply the theorem, note that $\exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\pi_b)\right\}>0$ for all κ and δ. The matrix $B=(\iota'\tilde{Q})^{-1}\tilde{Q}\tilde{C}'$ has all positive entries and row and column sums $\tilde{Q}$ and $\tilde{C}$, respectively, where $\tilde{Q}_\kappa=\sum_\delta\hat{\mu}_{\kappa\delta}$ and $\tilde{C}_\delta=\sum_\kappa\hat{\mu}_{\kappa\delta}$.

[53]To illustrate, suppose reuse at δ incurs an additional cost of c_δ per truckload. If this cost is the same for truckloads from all origins κ, then the magnitude of c_δ has no effect on the relative probability that trucks from κ or κ' are matched to δ.

ates for economically relevant interactions between κ and δ-specific covariates. In the main specification of the model, only the covariates indicated in Table 7 are included in x: (1) guar- and acrylamide indicators, as discussed in Appendix 10; (2) an indicator for shipments from small well pads to CTFs. (2) is intended to capture the interaction between the nature of disposal at well pads generating little wastewater in comparison to well pads generating more wastewater and the nature of water sourcing at a CTF in comparison to at a well pad. In addition to the covariates listed in Table 7, z includes additional liability-related dummy variables pertaining to firms for which there was insufficient data to perform the classification described in Appendix 10.

12.3 Computation

I solve (9) using a coordinate descent procedure similar to the iterated projection fitting procedure described in Galichon and Salanie (2022). Observe that we can re-write (9) as:

$$\sum_{\delta \in D} z_{\kappa\delta} U_\kappa V_\delta = \tilde{Q}_\kappa \quad \forall\, \kappa \in K$$
$$\sum_{\kappa \in K} z_{\kappa\delta} U_\kappa V_\delta = \tilde{C}_\delta \quad \forall\, \delta \in D$$

Even more compactly, this is:

$$U \circ ZV = \tilde{Q}$$
$$V \circ Z'U = \tilde{C}$$

where $U = (U_1, \ldots, U_K)$ and $V = (V_1, \ldots, V_D)$, and Z is the $K \times D$ matrix of $z_{\kappa\delta}$ values, and $\circ$ denotes the Hadamard product. Initialize a positive $U^{(0)}$ and $V^{(0)}$. I perform the following iteration:

$$U^{(s+1)} = \tilde{Q} \circ \left(ZV^{(s)}\right)^{-1}$$
$$V^{(s+1)} = \tilde{C} \circ \left(ZU^{(s+1)}\right)^{-1}$$

Following Conlon and Gortmaker (2020), I stop the iteration when the absolute error is less than 10^{-12}, where the absolute error is:

$$\max\left\{\left\|\tilde{Q} - U \circ ZV\right\|_\infty, \left\|\tilde{C} - V \circ Z'U\right\|_\infty\right\}$$

13 Policy analysis details

13.1 Derivation of tax schedules

The cost function under s is:

$$C_s(\mu) \propto \begin{cases} \sum_{\kappa\delta} \mu_{\kappa\delta} r_{\kappa\delta} - \sum_{\kappa\delta} \mu_{\kappa\delta} \phi_{\kappa\delta} - \mathcal{E}(\mu; Q, C) & \text{if } s = 0 \\ \sum_{\kappa\delta} \mu_{\kappa\delta} r_{\kappa\delta} - \sum_{\kappa\delta} \mu_{\kappa\delta} (1-s) \phi_{\kappa\delta} - \mathcal{E}(\mu; Q, C) & \text{if } s \in (0,1) \\ \sum_{\kappa\delta} \mu_{\kappa\delta} r_{\kappa\delta} - \mathcal{E}(\mu; Q, C) & \text{if } s = 1 \end{cases}$$

(Note that $\mathcal{E}$ does not depend on ϕ.) Under the assumptions of the model and the restriction that volumes in the sharing market are fixed, it can be shown that the optimal shipment plan μ_s^* satisfies:

$$\mu_{\kappa\delta}^{(s)} \propto \begin{cases} \exp\left\{-(\sigma_\kappa + \sigma_\delta)^{-1} \left(\gamma d_{\kappa\delta} + r_{\kappa\delta} - \phi_{\kappa\delta} - \tilde{u}_\kappa^{(1)} - \tilde{v}_\delta^{(1)}\right)\right\} & \text{if } s = 0 \\ \exp\left\{-(\sigma_\kappa + \sigma_\delta)^{-1} \left(\gamma d_{\kappa\delta} + r_{\kappa\delta} - (1-s)\phi_{\kappa\delta} - \tilde{u}_\kappa^{(1)} - \tilde{v}_\delta^{(1)}\right)\right\} & \text{if } s \in (0,1) \\ \exp\left\{-(\sigma_\kappa + \sigma_\delta)^{-1} \left(\gamma d_{\kappa\delta} + r_{\kappa\delta} - \tilde{u}_\kappa^{(0)} - \tilde{v}_\delta^{(0)}\right)\right\} & \text{if } s = 1 \end{cases} \tag{13}$$

where $\tilde{u}^{(s)}$ and $\tilde{v}^{(s)}$ satisfy the conditional market clearing conditions:

$$\sum_{\delta \in D} \mu_{\kappa\delta}^{(s)} = \sum_{\delta \in D} \hat{\mu}_{\kappa\delta} \quad \forall\, \kappa \in K$$
$$\sum_{\kappa \in K} \mu_{\kappa\delta}^{(s)} = \sum_{\kappa \in K} \hat{\mu}_{\kappa\delta} \quad \forall\, \delta \in D$$

Inspection of (13) shows that the Pigouvian tax on truck-miles between κ and δ is:

$$tax_{\kappa\delta}^{(s)} = \begin{cases} \gamma - d_{\kappa\delta}^{-1} \phi_{\kappa\delta} & \text{if } s = 0 \\ \gamma - (1-s)\, d_{\kappa\delta}^{-1} \phi_{\kappa\delta} & \text{if } s \in (0,1) \\ \gamma & \text{if } s = 1 \end{cases} \tag{14}$$

13.2 Robustness of uniform tax within $\mathcal{S}$

Consider $\mathcal{S} = \{\alpha : \alpha \in [0,1]\}$, where the state $s = \alpha$ implies that a fraction α of the friction $\phi_{\kappa\delta}$ is welfare-relevant and a fraction $1-\alpha$ is not. Under this assumption, we can write a private cost function C_α for state $s = \alpha$, which takes the form:

$$C_\alpha(\mu) \propto \sum_{\kappa\delta} \mu_{\kappa\delta} r_{\kappa\delta} - (1-\alpha) \sum_{\kappa\delta} \mu_{\kappa\delta} \phi_{\kappa\delta} - \mathcal{E}(\mu; Q, C)$$

Then a robust Pigouvian allocation $\mu^*_{\mathcal{S}}$ is the solution to:

$$\max_{\mu \in \mathcal{M}(Q,C)} \min_{\alpha \in [0,1]} f(\mu, \alpha)$$

where:

$$f(\mu, \alpha) = -\gamma \times \left\{ \sum_{\kappa\delta} \mu_{\kappa\delta} d_{\kappa\delta} \right\} - C_\alpha(\mu)$$

Note that $\mathcal{M}(Q,C)$ is a compact, convex set. Observe that $f(\cdot, \alpha)$ is concave in μ by the concavity of $\mathcal{E}$, while $f(\mu, \cdot)$ is linear in α and therefore convex. Hence, by the minimax theorem, $\mu^*_{\mathcal{S}}$ is the solution to:

$$\min_{\alpha \in [0,1]} \max_{\mu \in \mathcal{M}(Q,C)} f(\mu, \alpha)$$

If $f(\mu, \alpha)$ is strictly increasing in α, it follows that $f(\mu, \alpha)$ obtains a minimum at $\alpha = 0$.

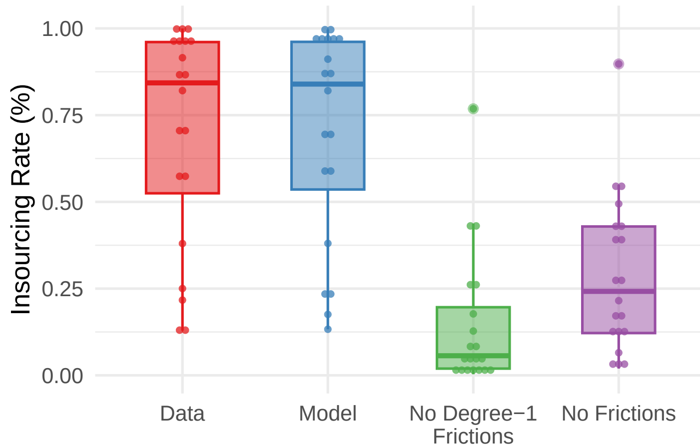
14 Supplemental material

14.1 Substitution between insourcing and outsourcing

To further contextualize the sharing frictions, I analyze the extent to which aggregate insourcing rates are driven by sharing frictions as opposed to transaction-specific differences in the intrinsic cost of reuse. Because each firm's operations are geographically concentrated (depicted in Figure 1), the intrinsic costs of reuse within the firm are often lower than the intrinsic costs of potential sharing transactions. Thus, qualitatively large sharing frictions might contribute relatively little to observed insourcing rates: put differently, large frictions might coincide with relatively small distortions (as recognized by Atalay et al. (2019)).

I find that the effects of sharing frictions on insourcing rates are modestly large. Absent sharing frictions, only 49% of reuse would occur within firms. The observed insourcing rate of 91% reflects the net effect of two potentially conflicting forces. Sharing frictions directly raise the cost of potential sharing transactions, and indirectly mediate price levels in the sharing market. The removal of sharing frictions simultaneously increases the supply of and

Figure 10: Internal reuse (insourcing) rates for twenty largest firms

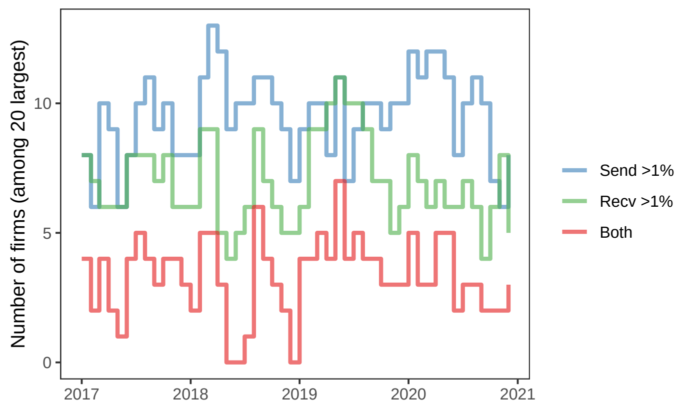


demand for resources in the sharing market. Price effects change the relative benefits of outsourcing particular transactions.

Figure 10 presents the distribution of insourcing rates for the twenty largest firms in the data, the fitted model, and in two counterfactuals. Among these firms, the average firm reused 70% of wastewater internally under the status quo. The purple series shows that in a world without sharing frictions, 29% of wastewater would be reused internally on average. But if only the frictions directly affecting potential transactions to which firm f is a counterparty were eliminated (i.e., f's degree-one frictions), f's insourcing rate would fall to 15% on average (shown in green), implying that price effects tend to mute the direct effects of the removal of sharing frictions in this setting.

15 Additional tables and figures

Figure 11: Sharing market participation among twenty largest firms



Notes: Indicates the number of firms among the twenty largest firms (by wastewater disposal volume) that sent more than 1% of wastewater to a rival, received more than 1% of wastewater from a rival (among observed shipments), or both, on a monthly basis.

Figure 12: Pairwise fitted sharing rates vs. data (in logs)

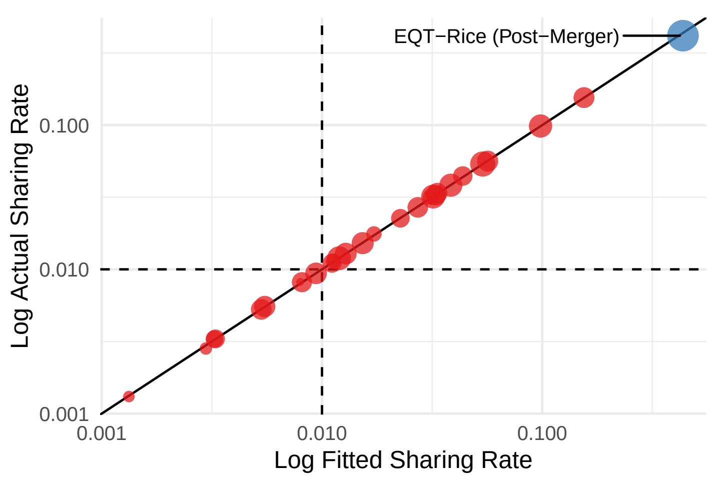
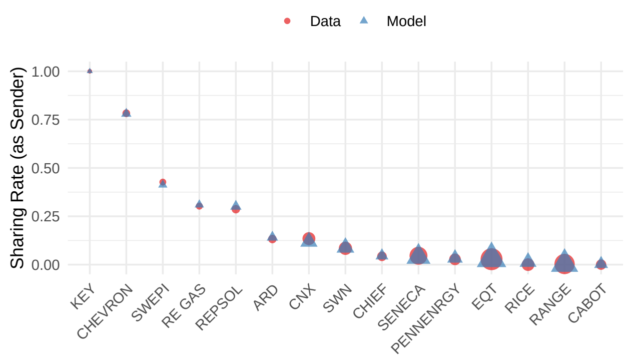


Figure 13: Model fit diagnostics

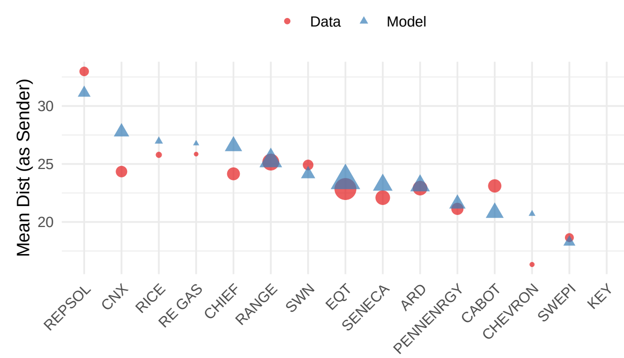
(a) Sharing Rate (as sender, by firm)



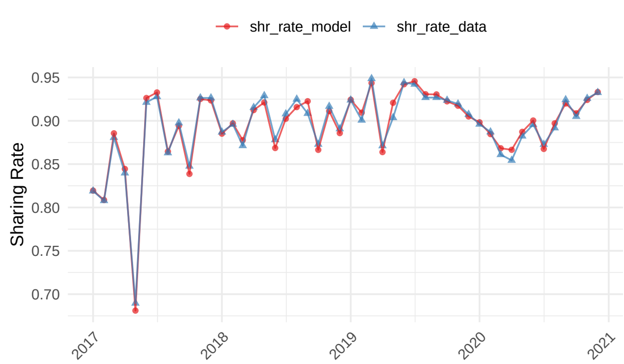
(b) Mean Dist (as sender, by firm)

(c) Sharing Rate (as receiver, by firm)

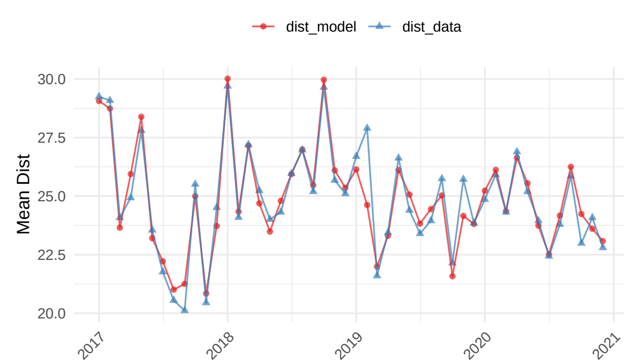
(d) Mean Dist (as receiver, by firm)



(e) Sharing Rate (by month)



(f) Mean Dist (by month)



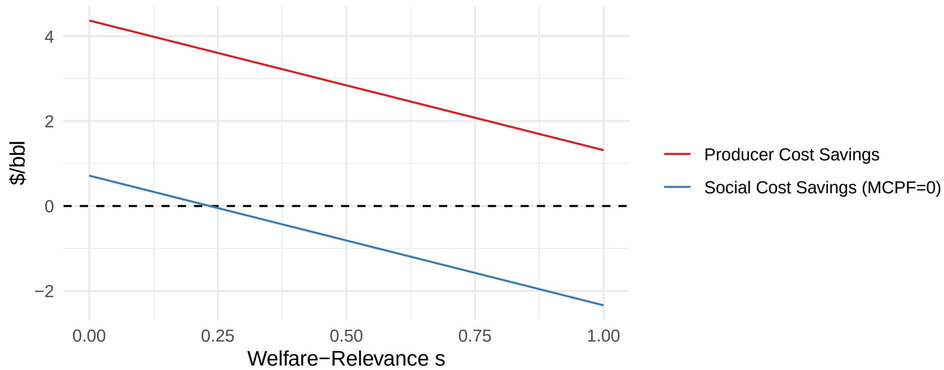
Notes: In the first two rows: figures include the fifteen largest firms by disposal volume (I use fifteen rather than twenty for legibility); firms are sorted by the indicated variable; the size of each dot corresponds to the relevant volume for each firm (volume as sender, or volume as receiver, respectively). In the last row, aggregate statistics at the monthly level are plotted.

Table 6: Alternative specifications

	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
Units	Miles, Linear	Miles, Linear	Miles, Linear	Miles, Linear	Hours, Linear	Miles, Log	Miles, Non-linear	Miles, Non-linear
Mean ϕ	127.6	135.1	132.3	131.7	4.1	6.4	291.7	250.2
α coefficients								
wastewater quality	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
liability	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
facility types	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
distance	No	No	Linear	Non-linear	No	No	No	Non-linear
β parameters	No	Yes	Yes	Yes	Yes	Yes	Yes	Yes
$\sigma_K + \sigma_D$	22.0	22.5	22.0	22.0	0.6	0.9	50.4	43.8
Log likelihood	3.8776	3.9099	3.9100	3.9103	3.9107	3.8787	3.8843	3.8851
Model fit (median abs. err.)								
monthly mean distance (mi)	0.46	0.40	0.39	0.38	0.40	2.67	1.17	1.17
firm-month mean distance (mi)	0.99	0.95	0.97	0.97	1.01	1.70	1.35	1.36
monthly share %	0.0053	0.0042	0.0042	0.0041	0.0042	0.0049	0.0043	0.0041
firm-month share %	0.0064	0.0031	0.0032	0.0031	0.0032	0.0052	0.0036	0.0037

Notes: The baseline specification is (2). In this specification, $d_{\kappa\delta}$ corresponds to linear miles. In (5), $d_{\kappa\delta}$ corresponds to linear drive time (in hours). In (6), $d_{\kappa\delta}$ corresponds to the log of the distance between κ and δ in miles. In (7) and (8), distance is represented non-linearly with indicators for 30 mile increments; $d_{\kappa\delta}$ is an indicator that equals one when the shipment between κ and δ is less than 30 miles. Likewise, when distance is included non-linearly in $z_{\kappa\delta}$ in (4) and (8), it is represented with indicators for 30-mile increments. The log-likelihood is the negative of the objective function in equation (7) in the body of the text. The model fit statistics report the median deviation between the observed and fitted expected distance and sharing probability across the indicated category (months, or firm-months).

Figure 14: Costs differences under full sharing subsidies vs. uniform tax



Notes: Shows difference cost savings when implementing (11) under the assumption that $s = 0$ (full sharing subsidies) vs. the assumption $s = 1$ (uniform tax), conditional on the true but unknown value of s.

Table 7: Constant only vs. full specification estimates (in miles)

	Constant Only			Full Spec.		
	Est	SE	\$/bbl	Est	SE	\$/bbl
Mean $\phi_{\kappa\delta}$						
weighted by data	127.8	0.044	5.81	125.7	0.072	5.71
weighted by benchmark	127.8	0.044	5.81	154.2	0.081	7.01
Sharing market cost shifters α						
rival $\times$ constant	127.8	0.044	5.81			
rival $\times$ poor $\rightarrow$ good env record				-	-	-
rival $\times$ good $\rightarrow$ poor env record				8.5	0.110	0.39
rival $\times$ gel $\rightarrow$ slickwater				-28.6	0.103	-1.30
rival $\times$ slickwater $\rightarrow$ gel				85.3	2.996	3.88
rival $\times$ large $\kappa \rightarrow$ well pad				-	-	-
rival $\times$ large $\kappa \rightarrow$ CTF				25.2	0.044	1.15
rival $\times$ small $\kappa \rightarrow$ well pad				4.4	0.151	0.20
rival $\times$ small $\kappa \rightarrow$ CTF				29.6	0.261	1.35
Within-firm cost shifters β						
gel $\rightarrow$ slickwater	-6.7	0.041	-0.31	6.7	0.092	0.31
slickwater $\rightarrow$ gel	-7.0	0.044	-0.32	-8.7	0.046	-0.39
small $\kappa \rightarrow$ CTF	-3.2	0.078	-0.15	-5.7	0.129	-0.26
$\sigma_\kappa + \sigma_\delta$	22.0	0.006	1.00	22.5	0.006	1.02
Mean distance (sharing market)						
weighted by data	43.6	-	1.98	43.6	-	1.98
weighted by benchmark	24.5	-	1.12	24.5	-	1.12

Notes: SE indicates the MLE standard error. Point estimates are converted into dollars per barrel (\$/bbl) under the assumption that marginal transportation costs are \$5/mile and that each water-hauling truck holds a full capacity of 110 barrels. The "benchmark" refers to distance-minimizing allocation. Note that $u_\kappa^{\mathcal{I}}$ and $u_\delta^{\mathcal{I}}$ are not reported, and that κ- and δ-specific covariates in x (such as small well pad or CTF indicators) are not separately identified from $u_\kappa^{\mathcal{I}}$ and $u_\delta^{\mathcal{I}}$.

Table 8: Selected counterfactuals

	Full Sample		Pre-merger		Post-merger	
	Dist	Shr%	Dist	Shr%	Dist	Shr%
Parameter Sensitivity						
Data	24.86	10.60	-	-	-	-
Model	24.85	10.58	-	-	-	-
$\phi_{\kappa\delta}$ Counterfactuals						
optimal scaling ($\approx 0.60 \times \phi$)	24.78	13.08	-	-	-	-
high friction ($10 \times \phi$)	28.00	9.49	-	-	-	-
no friction	28.74	51.37	-	-	-	-
$\sigma_K + \sigma_D$ Counterfactuals						
$\sigma_K + \sigma_D \to 0$	21.63	9.71	-	-	-	-
$\sigma_K + \sigma_D \to \infty$	146.99	84.37	-	-	-	-
Within EQT-Rice						
Data	-	-	22.62	0.00	21.67	42.58
Model	-	-	22.52	4.29	21.68	44.47
Merger Counterfactuals						
never merged	-	-	22.52	4.29	21.90	20.35
always merged	-	-	21.68	34.11	21.68	44.47

Notes: First panel: the ϕ counterfactuals report the mean shipment distance and sharing rate when scaling all estimated sharing frictions by a common factor: respectively, the level that minimizes mean shipment distance (approximately 0.60), ten, and zero. The $\sigma_K + \sigma_D \to 0$ counterfactual is constructed by solving a linear program where the cost function is the estimated systematic cost $r_{\kappa\delta}$. The $\sigma_K + \sigma_D \to \infty$ is constructed by assuming that each truckload i is sent to each facility δ with equal probability, and likewise that each delivery slot j is allocated to a truckload from κ with equal probability. Second panel: distance indicates the mean distance in miles for shipments within the EQT-Rice joint entity. Share % indicates the percentage of truckloads crossing the pre-merger firm boundary. Pre-merger refers to the period from 2017–01 to 2017-06; post-merger refers to the period from 2017-12 to 2020-12.