Firm Boundaries and External Costs in Shale Gas Production

Matthew O'Keefe [*,†]

January 8, 2023

Please click here for the latest version.

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. I analyze how firms choose between internal reuse and sharing in Pennsylvania. To do so, I build a market-level model of wastewater management in which the extent of the sharing market is determined endogenously by firms' make-vs-buy decisions. Estimating the model, I find that transaction costs associated with sharing are large — approximately $6 per barrel on average — but heterogeneous. Variation in the estimates reveals several channels for potential policy interventions to improve sharing markets. However, increased sharing may be undesirable: because firms' operations are clustered geographically, internal reuse typically reduces transportation-related external costs. Pigouvian interventions that simultaneously address sharing market imperfections and environmental externalities can worsen local environmental harms.

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

---

[*] 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 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 regulators. Conventional disposal wells permanently remove water from 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 as a substitute for 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* — generates economies when a firm's disposal needs exceed its capacity for reuse (or vice versa) or when transportation or treatment synergies are present. 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 frequently generates joint cost savings. In this context, a reasonable concern for a regulator is whether imperfections in the *sharing market* lead to inefficiently high rates of internal reuse, which could exacerbate environmental harms. A well-functioning sharing market potentially reduces aggregate demand for final disposal — and freshwater — while lessening shipment distances and storage durations.

In this paper, 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 first part of the paper I quantify *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 data for this study consist of wastewater disposal records maintained by the Pennsylvania Department of Environmental Protection (DEP). Using these data, I first establish that the Pennsylvania sharing market resembles a spot market in key respects. At the same time, sharing frictions appear to play an important role in shaping firms' wastewater management decisions. This can be seen most 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.

Sharing occurs when joint cost savings exceed any sharing frictions. 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, intrinsic costs that are not directly related to distance (such as treatment or labor costs) appear to be important components of private costs.

In order to empirically disentangle sharing frictions from other intrinsic 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 fracked wells generate only a limited volume of wastewater in each period, while the capacity for reuse in fracking any single well is constrained by engineering considerations. Capacity constraints create tradeoffs between minimizing disposal costs and minimizing water sourcing costs. 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. Absent unobserved heterogeneity in intrinsic costs of reuse at the level of the origin-destination pair, shipments within firms' boundaries identify the intrinsic costs of sharing transactions. 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 a fully parameterized model using a maximum likelihood estimator similar to the one proposed by Galichon and Salanie (2022). The estimated sharing frictions are three times as large as distance-related transportation costs for the average sharing transaction, equivalent to a cost of about \$6 per barrel. 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 transaction costs across observably similar transactions: the standard deviation in sharing frictions across transactions is almost \$3 per barrel, due to large systematic differences in costs across firm-pairs.

Contracting frictions appear to be an important source of sharing frictions. Only a few pairs of firms share at close-to-integrated rates under the status quo, implying that firms are rarely able to circumvent sharing frictions through formal contracting or relationships. 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. The presence of a relatively thick sharing market may undermine relationship formation. These findings suggest that improvements in the contracting environment (such as clarification of liability) may be required to significantly improve the performance of sharing markets.[1,2]

In the next part of the paper, 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. In the model, I find 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. Second, firm boundaries inhibit matching on all components of the joint costs of reuse simultaneously, not just distance-related transportation costs. Because each firm's operations are clustered in space, matches within the firm are typically nearer in distance than those between firms. Consequently, sharing frictions inhibit matching at longer distances to a greater extent than

---

[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. In the Permian basin, at least one entrepreneur operates a digital platform for wastewater sharing. The same entrepreneur operated a similar platform in Pennsylvania between 2014-2016 but ultimately exited.

[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.

they inhibit matching on geographic complementarities. Under the status quo, the non-randomness of firm boundaries softens 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.

The net environmental impact of sharing is heterogenous across potential sharing transactions when transportation is a key source of environmental harms. Consequently, reductions in transaction costs do not unambiguously reduce external costs. This observation reproduces at the interfirm level a previously recognized tension between globalization and environmental outcomes (Copeland et al., 2022). An implication is that market design interventions intended to make sharing easier may have limited net benefits, and could even be counterproductive. Nevertheless, interventions to improve sharing might improve social welfare despite increasing external costs if private cost savings are sufficiently large.

The final part of the paper derives and analyzes the Pigouvian allocation, which minimizes the sum of private and external costs. If sharing frictions are Pareto-irrelevant, the social planner must effectively solve two market failures at once: firms' failure to internalize environmental costs and limited participation in the sharing market. One way to implement the optimal allocation is with targeted 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, which are not considered in this exercise, such a policy might be difficult to implement from a public choice perspective.

In practice, sharing frictions likely consist of some costs that are welfare-relevant (e.g., search and negotiation costs) and some "costs" that are not (e.g., managerial inattention).[4] These types of costs are not separately identified in my empirical framework, and — as I argue — may be difficult for a regulator to separately identify in practice. 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. This finding highlights a tension in the conventional Pigouvian analysis, which implicitly assumes that all costs are welfare-relevant. When welfare-relevance is ambiguous, a different approach is needed: I conclude by considering how Pigouvian regulation can be made robust to this type of ambiguity.

---

[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.

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) provide useful surveys discussing local environmental issues as well as broader environmental and economic considerations.[5] Groundwater Protection Council (2019) comprehensively surveys the current legal and regulatory frameworks applicable to wastewater management, including reuse within the oil and gas industry. I contribute to the existing literature by developing a structural framework that can be used to prospectively evaluate wastewater regulation.

Sharing frictions can be viewed as a particular type of transaction cost. Coase (1960)'s recognition of the significance of transaction costs when external costs are present has played a key role in the development of environmental economics, contributing to the adoption of market-based environmental regulation in many settings. In Coase (1960), transaction costs preclude efficient bargaining between polluting firms and individuals who bear the costs of pollution. In Stavins (1995) and Hahn and Stavins (2011), transaction costs are understood as the costs of trading in a cap-and-trade market, or of paying an emissions tax, or otherwise complying with regulation. In contrast, the frictions that I study in this paper pertain specifically to coordination between firms and are therefore more closely related to 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, 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.

A key reason that relatively few papers in empirical industrial organization have incorporated Coasean transaction costs is the general scarcity of data on firms' internal operations. My empirical analysis is only possible because I observe insourcing and outsourcing in the same market, which enables me to separately identify the intrinsic cost of production and

---

[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.

transaction costs.[6] Demsetz (1988) emphasizes that the make-vs-buy decisions is inherently marginal and that transaction costs may well be heterogenous across apparently similar transactions. Although my analysis is limited to one market, my approach accounts for this richness to a greater extent than the closely related approach of Atalay et al. (2019), who use Census data (which also records within-firm shipments) to quantify the "net benefits of ownership" across a wide range of manufacturing 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 across transactions within a single product market rather than across product markets.

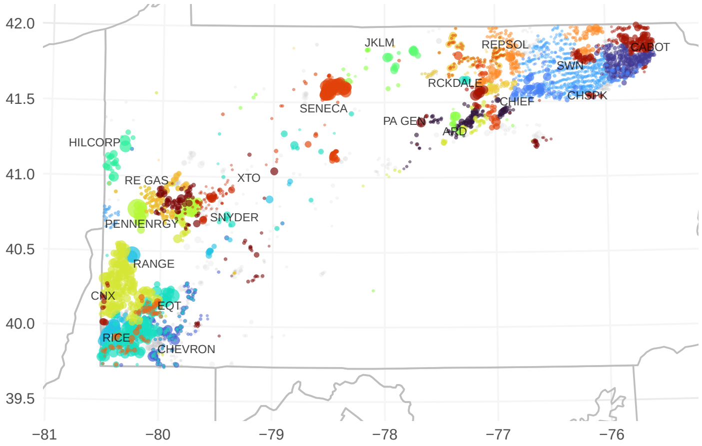
The Coasean view of the firm that I adopt is complementary to the strategic view of the firm typically encountered in industrial organization (Bresnahan and Levin, 2012). In this way, the analysis in this paper complements recent empirical work at the intersection of industrial organization and environmental economics that highlights the significance of oligopoly for external costs and regulation (e.g., Mansur, 2007; Fowlie, 2009; Leslie, 2018; Preonas, 2023). By taking a different perspective, I document a distinct but closely related class of market imperfections arising from industry structure that can pose an empirically relevant challenge for implementing environmental regulation.

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 (2022) studies search and matching frictions in subcontracting decisions. Sweeney and Covert (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 Covert and Kellogg (2023), who also analyze the environmental impacts and societal tradeoffs associated with private cost minimization in the oil and gas sector, focusing on railroad transport of crude.

---

[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: Locations of Well Pads for Top 20 Firms



2 Setting

In this section, I briefly describe the economics of wastewater reuse and sharing in Pennsylvania. Then I introduce the data and present key summary statistics on shipment patterns and participation in the *sharing market*. Finally, I quantify the environmental impacts of wastewater trucking, the focus of my policy analysis.

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 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 the largest global energy firms to small, independent firms operating only a few wells (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 fracking event 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] 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).

In Pennsylvania, wastewater *reuse* creates significant surplus for operators because conventional disposal is costly. *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 common than treatment at a CTF, but both are prevalent (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

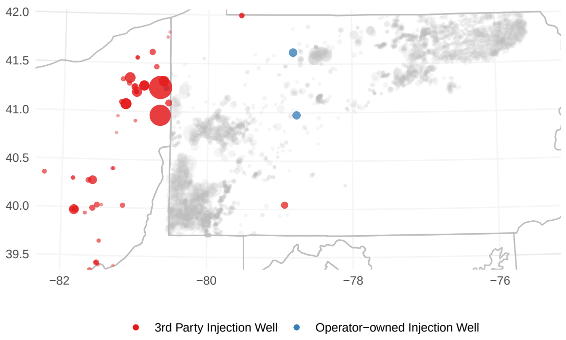
---

[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] In principle pipeline transportation is also possible, both for disposal and reuse, but this is reportedly rare in the Appalachian Basin (including Pennsylvania), 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).

Figure 2: Injection Well Locations



to higher standards, although these technologies are rarely used in practice.[10]

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 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, there is often no charge, and negative prices are possible (for example, if freshwater is scarce).[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 acces-

---

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

sible on the DEP website and constitute my primary data source. The data clearly indicate whether a transfer was intended for reuse and, if so, provides further information identifying the destination. I use this information to distinguish internal reuse from sharing, based on whether the destination facility was associated with the sender in related DEP databases.

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 contributing 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 consideration of less-than-full truckloads. In order to mitigate the impact of unobserved timing constraints within a month, I exclude from the main 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] Additional details concerning the data cleaning process are discussed in Appendix A.

2.2.1 Centralized treatment facilities

One limitation of the data requires special treatment throughout the analysis. Specifically, the data record 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 consid-

---

[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. Prior to the merger discussed in Section 3.1, EQT and Rice formally delegated wastewater management to midstream subsidiaries that recorded wastewater disposal revenues on a cost plus basis, but it is unclear to what extent the parent firms would have retained decisionmaking authority under such an arrangement.

[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. However, 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 10 presents the full time series of data since 2010. The figure illustrates that market conditions are remarkably stable during the sample period.

Table 1: Disposal Market Shares, 2017-2020

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

erations are not relevant for shipments to well pads, because re-transfer of wastewater from one well pad to another is prohibited by the DEP.

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, although FracFocus is known to have limitations (for example, firms may fail to submit timely or complete records). The data indicate detailed latitude and longitude information for every well pad and disposal or reuse location. 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

---

[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

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 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 frack a new 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 well pad disposed of just 3.7 truckloads, and 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 receiving 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).

---

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 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: Key Summary Statistics

	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

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 the paper.

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] Participation varies considerably over time, as illustrated in Figure 11, but no fewer than 10 distinct large firms were active in each month of the sample period. Thus, many firms share at least occasionally, and participation status is not "sticky," suggesting that sharing participation does not require large fixed cost investments (in contrast to import-export decisions, e.g. in Antras and Chor (2022)).

---

[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.

Conditional on participation, firms typically have many distinct sharing partners, suggesting that the cost of reaching an agreement with a new counterparty is not excessive. Among the twenty largest firms, firms that sent wastewater to a rival shared with of 1.7 distinct counterparties per month and 7.3 distinct counterparties 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 counterparties per month and 8.1 distinct counterparties over the course of the sample (median 7.0). Although I cannot exclude the possibility that some pairs of firms form long-term relationships like those studied empirically in Macchiavello and Morjaria (2015, 2021), 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 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 (which is prone to delays). Consistent with this mechanism, Table 6 presents probit regression coefficients obtained by regressing indicators for whether a firm sent wastewater to a rival or sourced wastewater on the sharing market on a firm's fracking rate (which I construct from fracking activity recorded in FracFocus). The estimates imply that a firm is 3.9 percentage points more likely to send wastewater to a rival when its fracking rate is one standard deviation below its mean rate, and 7.1 percentage points more likely to source wastewater on the sharing market when its fracking rate is one standard deviation above its mean rate.[21] This suggests that firms send wastewater to rivals when fracking less, and then accept wastewater from rivals when fracking more.

---

[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] The coefficient estimates are likely attenuated by measurement error due to discrepancies in reporting between the DEP data and FracFocus. Thus, the true marginal effects are likely somewhat larger.

2.4 External costs of wastewater trucking

In the policy analysis, I focus on negative environmental externalities created by wastewater trucking. I find that unpriced carbon emission and air pollution externalities are roughly 7% of private transportation costs.[22] This figure is constructed as follows. 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. 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.[23] 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.[24] 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.

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 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, with drivers perhaps being disproportionately younger, male economic migrants (Wilson, 2022) subject to looser 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

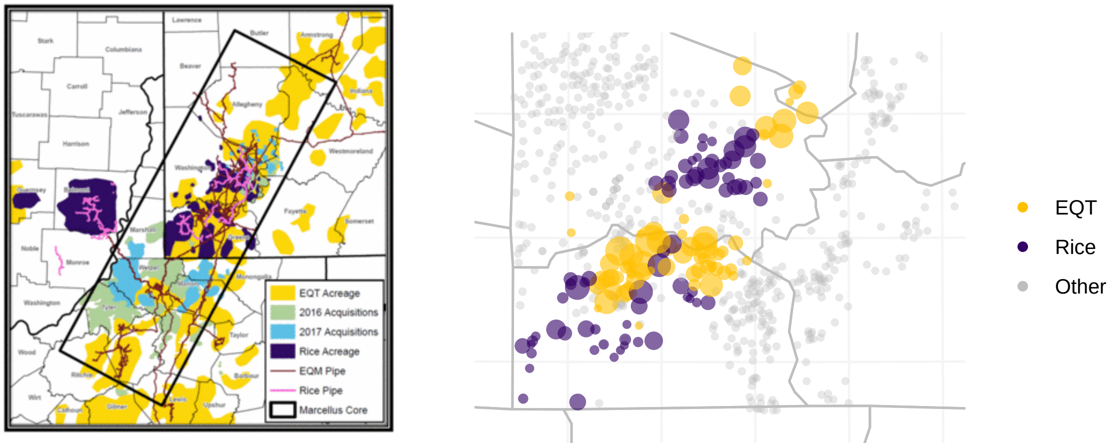
---

[22] Interestingly, this is smaller than but similar in magnitude to the 20% external damages estimate for railroad transport of crude found in (Clay et al., 2017). The difference can be rationalized by the fact that NOx and PM2.5 emissions are smaller for trucks than for locomotives, especially after model year 2010.

[23] 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.

[24] To compute air pollution, I currently 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).

Figure 3: EQT and Rice Locations, 2017



(a) Map from EQT presentation

(b) Southwest PA well pads in data

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 – real or shadow costs that are incurred if a transaction takes place between rival firms, but not if a transaction takes place within a firm. This section presents evidence that sharing frictions play an important role in shaping the allocation of wastewater for reuse.

3.1 The EQT-Rice merger

In June 2017 EQT Corporation and Rice Energy Inc. announced a merger, which was completed in November 2017. 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 created the largest natural gas producer in the United States. In this section I argue that changes in wastewater shipment patterns subsequent to the merger suggest that sharing frictions are large, and not fully attributable to search or negotiation costs.

Figure 3a reproduces a map included in an investor presentation that accompanied the merger announcement depicting EQT and Rice acreage throughout Appalachia. Figure 3b shows the locations of EQT and Rice well pads appearing in the data.[25] 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.[26]

Despite apparently large gains from trade resulting from geographic complementarity, 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 pads between January and June 2017. 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.[27]

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 market shares for former Rice- and EQT well pads from the completion of the merger in November 2017 to December 2020. 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

---

[25] As the first map indicates, EQT also had a significant presence in West Virginia, while Rice was present in Ohio.

[26] See EQT Corporation (2017). Form 8-K (0001104659-17-040193). U.S. Securities and Exchange Commission. Exhibit No. 99.2, "Investor Presentation, dated as of June 19, 2017 (furnished and not filed for purposes of Item 7.01)"

[27] 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.

Table 3: 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

of significant ex ante sharing frictions.[28]

Dahlman (1979) distinguishes between three types of transaction costs: search and information costs, bargaining and decision costs, and "policing and enforcement" costs (which might be taken to encompass problems of incomplete contracting). The above patterns can help to distinguish between which of these types of costs contribute most to sharing frictions. Suppose that search costs are declining in distance. Then search frictions are unlikely to have been the main factor inhibiting sharing between EQT and Rice, because many firms that did engage in sharing were located further apart than EQT and Rice. Similarly, suppose that negotiation costs are independent of the scale of gains of trade. Then negotiation costs are unlikely to have been the main source of sharing frictions between EQT and Rice, since many pairs of smaller and less proximate firms did engage in sharing. The merger evidence therefore suggests that "market transaction costs" (i.e., search and negotiation costs) are not obviously the main source of frictions, at least not without a nuanced model of search and negotiation costs (for example, one with decisionmaker biases). I investigate possible sources of policing and enforcement costs later in the analysis. Some examples include information frictions relating to the quality of wastewater and inter-operation liability concerns. There is also scope for moral hazard, because firms that send water to rivals do not directly bear the costs of untimely deliveries, and firms that accept wastewater do not bear the costs of unloading delays that result from facility congestion.

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

---

[28] It is also possible these patterns are the result of unobserved changes in the joint entity's completion activity or other unobserved changes in relative costs, rather than the removal of sharing frictions specifically. I revisit the EQT-Rice merger from an equilibrium perspective later in the paper.

market when gains from trade are presumably smaller, suggesting that sharing frictions are heterogenous. Thus, it is unclear what the aggregate impacts of sharing frictions might be. Transportation costs are a primary driver of wastewater management costs, so one way to understand the scale of aggregate impacts is to compare the observed allocation of wastewater to the minimum-distance allocation for the market as a whole. If sharing frictions and distance-related components of transportation costs were the only sources of variation in cost, this exercise would identify the aggregate impact of sharing frictions.

In reality, other 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 it is needed 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. In principle, there could be heterogeneity in treatment costs 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 minimize distance.

To assess the aggregate significance of sharing frictions while controlling for unobserved sources of heterogeneity in costs, I compare observed shipments for reuse within firms to two simple benchmarks. The first is random matching within each firm f, irrespective of distance. The second implements the minimum-distance allocation within each firm f.

The minimum-distance allocation can be obtained by formulating and solving a simple optimal transport problem. Consider the flow of wastewater in month t between all well pads K_f and all disposal facilities D_f operated by f. For each well pad $\kappa \in K_f$ and disposal facility $\delta \in D_f$, the data records the actual shipment volume $\hat{\mu}_{\kappa\delta}$. Suppose that all truckloads of wastewater are identical, and that all shipments within a month can be costlessly re-allocated, 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. Then the distance-minimizing allocation μ^f solves the following problem:

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

where $d_{\kappa\delta}$ is the distance between κ and δ.[29] Because this problem is a linear program, it can be solved easily. Solving this problem for each firm, I find that within-firm shipments realize 56% of possible transportation cost savings on average relative to random matching. Implementing the minimum-distance allocation within each firm would reduce industry trucking intensity by 2.3 miles per truckload relative to the status quo, or by 9% of the status quo level, suggesting that differences in distance only partially explain firm shipment decisions. These results are summarized in the first three columns of Table 4.

Next I repeat the same exercise for the market as a whole. I compute benchmark one allocation with market-wide random matching and another with the market-wide minimum-distance allocation, this time ignoring firm boundaries. These results are summarized in the last two columns of Table 4. Compared to the allocation that optimizes within-firm shipments alone, the latter allocation reduces transportation intensity by 4.0 miles per truckload, or by a further 16 percentage points relative to the status quo. Thus, changing the composition of who participates in the sharing market and how shipments are allocated within the sharing market potentially has a larger effect on overall transportation efficiency than optimizing shipments within firms, even though most reuse occurs within firms. Moreover, sharing rates are more than three times greater under the minimum-distance allocation than under the status quo. These findings suggest that sharing frictions are often large, even if they are heterogenous or if gains from trade are frequently large enough to justify trade. It is also notable that the market as a whole realizes 95% of the maximum possible transportation cost savings relative to random matching. Intuitively, because firms are geographically segregated from one another, even relatively inefficient shipments within the firm (compared to the minimum-distance benchmark) coincide with an efficient outcome when viewed from the perspective of the market as a whole. This suggests that expanding the scope of the firm or otherwise reducing barriers to sharing could result in less efficient transportation overall to the extent that extramarginal shipments would not be driven by distance-related costs.

4 Model

In this section I develop a model of wastewater management that endogenizes insourcing and outsourcing in order to disentangle 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

---

[29] 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 Benchmark

		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

finite set of facilities accepting wastewater for reuse (both well pads and operator-affiliated CTFs).[30] Each month, a unit mass of wastewater is shipped for disposal or reuse, while a unit mass of water (wastewater or freshwater) is used for completing new wells or, in the case of a CTF, processed for later use in completing new wells. For simplicity, I assume that each well pad κ and facility δ 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 (I discuss limitations of the solution concept after presenting the model).

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 (shadow) cost of reusing wastewater 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

---

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

sharing frictions incurred when they do not. Thus, sharing frictions are defined as the difference between systematic costs when firm boundaries exist and 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.

In practice, the shadow cost of reusing wastewater from κ at δ varies from truckload to truckload due to *latent cost heterogeneity.* The true shadow cost of supplying the jth truckload of source water 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 private costs of managers m_κ and m_δ, respectively, which vary from truckload to truckload. These costs capture any real or shadow 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. 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.[31]

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 unobservables, and (importantly) not 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.[32] I assume

---

[31] 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.

[32] 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.

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 $\kappa \in K_0$. Unlike the finite assignment model of Shapley and Shubik (1972), there are uncountably many truckloads of wastewater; 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 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 (and unit price elasticity is more reasonable for firms than consumers), but $\tau_{\kappa\delta}$ does not necessarily represent a cash transfer. For example, $\tau_{\kappa\delta}$ could represent a “favor,” as in Samuelson and Stacchetti (2017).

Galichon and Salanie (2022) show there is a unique equilibrium matching μ^* in the core that 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 relative to injection disposal from κ and freshwater sourcing at δ. The match entropy term $\mathcal{E}$ captures the surplus contribution of savings on latent costs. Formally, $\mathcal{E}$ is a function that depends on μ, the distributions P_K and P_D, and the marginal probability masses $\mathbf{Q}$ and $\mathbf{C}$:

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

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

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

Intuitively, $G^*\left(\mu, \mathbf{Q}\right)$ and $H^*\left(\mu, \mathbf{C}\right)$ 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) provide an extensive interpretation of these objects.

Note that in contrast to the finite assignment model of Shapley and Shubik (1972), the transfer matrix τ is unique when P_K and P_D have full support (up to the utilities of agents' outside options). In particular, Galichon and Salanie (2022) Theorem 4 establishes that $r^K_{\kappa\delta} - r^K_{\kappa 0} - \tau_{\kappa\delta} = \frac{\partial G^*_\kappa}{\partial \mu_{\kappa\delta}}$ and $r^D_{\kappa\delta} - r^D_{0\delta} + \tau_{\kappa\delta} = \frac{\partial H^*_\delta}{\partial \mu_{\kappa\delta}}$, where G^*_κ and H^*_δ are the individual choice entropies for m_κ and m_δ, respectively. Intuitively, as the number of truckloads in the model grows large, the price indeterminacy encountered in the finite assignment model vanishes. Note that I do not provide an explicit model the price formation process.

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 better 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.

An important limitation of this decentralized approach is that I exclude 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 (1972)). An alternative possibility is to consider a model in which sharing decisions are mediated by Nash-in-Nash bargaining with transaction costs.[33] One challenge with a Nash-in-Nash approach is that there is no clear notion of a network in this setting. Moreover, even if sharing networks could defined in some reasonable way, it is challenging to model network formation in the Nash-in-Nash context, significantly complicating counterfactual analysis.[34] 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.

Another limitation of my approach is that I assume 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 (Antras 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 sustained long term relationships. These issues are not central to the policy questions of interest in this paper, except insofar as they shed light on potential policy interventions that I do not consider.

5 Identification and estimation

This section discusses identification and estimation.

---

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

[34] Ho and Lee (2019) and Ghili (2022) develop models of network formation in Nash-in-Nash environments. One challenge in adapting these models to the wastewater management setting is the lack of a clear distinction between upstream and downstream firms, since 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).

5.1 Identification

Let Δr denote the $K \times D$ matrix of systematic cost savings (relative to the outside options) 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 internal costs $r^{\mathcal{I}}$ take 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 a parameter vector θ, and $u^{\mathcal{I}}_{\kappa}$ and $u^{\mathcal{I}}_{\delta}$ are additively separable unobservables.[36] Under this assumption, $r^{\mathcal{I}}$ is identified from observed shipments within each firm, while ϕ is identified by shipments between firms. Separability implies that *within the firm*, any 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. This excludes the possibility that some types of wastewater are unobservably better or worse suited for reuse at particular well pads or CTFs.

Under these assumption, $\phi_{\kappa\delta}$ is identified whenever $r^{\mathcal{I}}_{\kappa\delta}$ is identified (because Δr itself is identified, and $\phi_{\kappa\delta} = \Delta r_{\kappa\delta} - r^{\mathcal{I}}_{\kappa\delta}$). In practice, however, I do not observe a sufficient amount of data to estimate ϕ non-parametrically 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

---

[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.

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 certain pairs of firms have a comparative advantage in coordinating certain types of shipments but not others.

5.1.1 Welfare-relevance

I interpret sharing frictions as 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 at the consumer level are incorporated into structural models. Under this interpretation, the welfare-relevance of the sharing frictions is not identified: only the gross transaction cost is identified. 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. Nothing in the data separately identifies welfare-relevant and welfare-irrelevant transaction costs.

5.2 Parameterization

To estimate the model, I 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 match-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.

---

[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 $u_{\kappa}^{\mathcal{I}}$ and $u_{\delta}^{\mathcal{I}}$.

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 covariates to disentangle potential sources of sharing frictions, including measures of wastewater characteristics, facility characteristics (e.g., well pads vs. CTFs), and firm-pair characteristics (such as differences in firm reputation).

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 π_b is constant within groups introduce bias into the estimates.

Finally, I assume that the latent cost distributions $P_K\left(X_{\kappa\delta};\theta\right)$ and $P_D\left(X_{\kappa\delta};\theta\right)$ 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 implies the independence of irrelevant alternatives which —- at the cost of realism —- 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 A and additional details concerning the full specification of the model in Appendix B.

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) conditional maximum likelihood estimator for θ_0 is:

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

In Appendix B, 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 $t \in T$:

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

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

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 B, 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

---

[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.

[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.

KNITRO and obtain standard errors under maximum likelihood assumptions.[40]

5.3.1 Counterfactuals

Although estimation can be conducted without estimates of $\mu_{\kappa 0}$ and $\mu_{0\delta}$, this information is required for constructing counterfactuals in the specified model. To avoid complication, I construct all counterfactuals under the assumption that there is no outside option on either side of the market.[41] Thus, I assume that the total volume of wastewater generated and reused at each well pad K and facility D is unchanged in counterfactuals, although in practice substitution to the outside goods depends on the model primitives. This means that I am unable to consider how aggregate reuse and final disposal volumes change in counterfactuals, which are important outcomes from a policy perspective (especially when drawing policy implications for in shale basins in which reuse is less common). This limitation of the analysis reflects a conservative use of data rather than a shortcoming of the model.

6 Estimates

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

6.1 Sharing frictions

Table 5 presents key parameter estimates. On average, sharing frictions for observed shipments between rivals are equivalent to the cost of shipping a truckload of wastewater a distance of 135.1 additional miles, or about \$6.14 per barrel (assuming trucking costs are \$5 per mile, a reasonable estimate for this industry). In comparison, the mean shipment distance for the same transactions was 43.6 miles (implying distance-related costs of \$1.98 per barrel). This implies that, on average, bringing a transaction inside the firm has the same effect on private costs as a threefold decrease in shipment distance. For robustness, Table 7 presents the same estimates side-by-side with estimates from a more parsimonious specification in which ϕ is constant, which yields a similar estimate for the mean friction. The richer specification demonstrates that sharing frictions are highly heterogenous. Across

---

[40] Gradients can be constructed using the implicit function theorem. I obtain starting values by recycling point estimates from a series of increasingly complex specifications nested within the full model.

[41] 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.

the 27 firm-pairs for which I estimate a firm-pair specific sharing friction fixed effect π_b, the mean friction ranged between \$0.38 and \$11.20 per barrel (IQR [\$2.38, \$4.83]). The aggregate mean therefore reflects the selected composition of the sharing market under the status quo: if the distance-minimizing allocation were implemented for the market as a whole, realized sharing frictions would be 20% greater on average.

6.1.1 Sources of frictions

Understanding the source of sharing frictions is crucial for drawing policy implications from the model. The merger evidence, variation in sharing across firm pairs, and the estimated α parameters each give insight into the sources of sharing frictions, as I discuss in this section.

Wastewater quality 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. Roughly, there are two types of fracking fluids: "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 presence of certain key chemicals.[42] I construct a similar measure for whether wastewater received at facility δ was likely to have been reused at a well that was fracked with a gel or a slickwater, aggregating across wells drilled by the operator in the case that δ is a CTF. I find that sharing frictions are more than \$5.18 per barrel greater (91% of the mean) when slickwaters are used at the sending well pad and gels are used at the receiving well pad than in the opposite case. Thus, sharing frictions appear to be greater when risks to well productivity are greater. In practice, firms may lack perfect information about the costs and benefits of accepting particular truckloads of wastewater. Uncertainty regarding match quality can make search more difficult ex ante. At the same time, the ex post productivity impact of specific barrels of wastewater may difficult to monitor after an agreement to share has been reached, because production is inherently uncertain.

Inter-operator environmental liability When a firm shares wastewater with another firm, it assumes some non-contractible risk that the receiving firm will mishandle the wastewater in some way that exposes the sending firm to civil liability, regulatory fines, or reputational damage. The most serious risk is that fracking fluid could migrate from the receiver's well or temporary storage containers into drinking water resources. Using DEP records, I

---

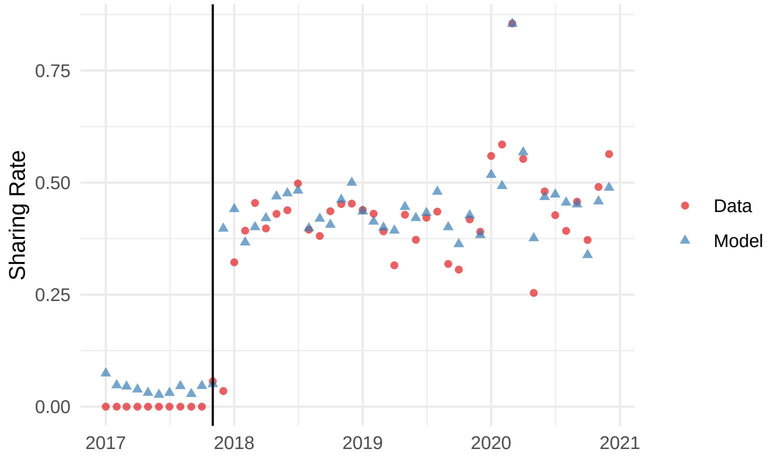
[42] Guar in the case of gels, and acrylamide in the case of slickwaters (Montgomery, 2013).

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
Sharing market cost shifters α			
rival × poor → good env record	-	-	-
rival × good → poor env record	8.5	0.110	0.39
rival × gel → slickwater	-28.6	0.103	-1.30
rival × slickwater → gel	85.3	2.996	3.88
rival × large κ → well pad	-	-	-
rival × large κ → CTF	25.2	0.044	1.15
rival × small κ → well pad	4.4	0.151	0.20
rival × small κ → CTF	29.6	0.261	1.35
Within-firm cost shifters β			
gel → slickwater	6.7	0.092	0.31
slickwater → gel	-8.7	0.046	-0.39
small κ → 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: Change in Cross-Firm Border Sharing after EQT-Rice Merger



construct a measure for the relative rates of spills and well site inspection failures among firms (controlling for firm size). If inter-operator liability concerns are significant, then sharing frictions should be greater when the firm affiliated with facility δ has a poor compliance record. Consistent with this hypothesis, I find that sharing frictions are \$0.39 per barrel greater when the sending firm has a good compliance record and the receiving firm has a poor compliance record than in the opposite case.

Non-integration Figure 4 plots the EQT-Rice "sharing rate" against the rate implied by the model before and after the merger was completed in November 2017.[43] After the merger, sharing frictions are assumed to be zero. The figure demonstrates that under this assumption, 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. Thus, the model successfully rationalizes both the absence of trade prior to the merger (despite the geographic proximity between EQT and Rice), and the level trade ex post. This suggests that integration can eliminate sharing frictions. Thus, whatever sharing frictions might consist of, it appears that they can be resolved through integration (consistent with the central but often-disputed claim of transaction cost economics, see e.g., Demsetz (1988)).

---

[43] 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).

Clarity and credibility In principle, any pair of firms can achieve optimal levels of bilateral coordination through formal or informal (i.e., relational) contracting. Although I do not observe contracts (or relationships), I am able to predict the shipment patterns that would be observed under efficient bilateral contracting. For each of the 27 firm pairs for which I obtained an estimate of π_b, I compute a counterfactual no-friction sharing rate (holding frictions for all other pairs of firms fixed). Figure 5 plots the actual sharing rate against the corresponding no-friction sharing rate in each of these counterfactuals.[44] The figure shows clearly that for most firm pairs, actual sharing volumes are far below the optimum. At the median, the actual sharing volume was 78% lower than the optimal sharing volume, and only 3 pairs of firms shared more than 75% of the optimal volume. In comparison, sharing volume between EQT and Rice-affiliated facilities was 96% of the optimal sharing volume after the merger. Thus, by comparison to the merger evidence, it is seemingly difficult to achieve efficient levels of sharing without integration.

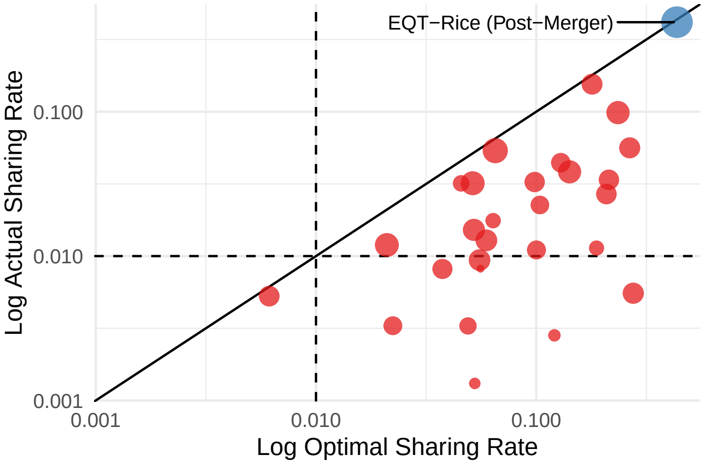
These patterns are consistent with a mechanism in which opportunistic ex post coordination is easy and surplus-maximizing ex ante coordination is difficult. Ex ante coordination can be relatively difficult for at least two reasons. First, barriers to information sharing (e.g., secrecy) can prevent firms from anticipating the gains from future coordination. Second, even if gains from trade are known, firms may not be able to credibly commit to delivery schedules that may require privately costly actions in the future when 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).

Principal-agent frictions Loosely, wastewater management for a typical firm can be divided into four different tasks. On the disposal side, managing disposal for mature wells (many locations, small volumes) is a qualitatively different task from managing disposal for recently drilled wells (few locations, large volumes); on the receiving side, managing wastewater sourcing on a well pad before a completion (inelastic demand, tight delivery windows) is qualitatively different from managing receipts at a CTF. These tasks may be delegated to different agents with different incentives. The estimated sharing frictions should be inversely related to the strength of agents' incentives to execute the lowest cost transaction. This perspective suggests a few testable hypotheses. For managers fracking new wells, water sourcing costs are a highly visible capital expenditure; in comparison, managers at operator-affiliated CTFs may have little incentive to source wastewater from rival firms even when doing so would reduce the firms' total cost, because it may be difficult for a principal to observe shar-

---

[44] For comparison, Figure 12 plots the actual sharing rate against the fitted sharing rate.

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).

ing market prices. I find that sharing frictions are \$1.15 greater for shipments to CTFs from larger well pads, and \$1.15 greater for shipments from smaller well pads.[45] I also find that sharing frictions are larger for shipments leaving smaller well pads (by \$0.20 for shipments to well pads and by \$0.20 for shipments to CTFs). This could reflect analogous differences in monitoring costs; however, there may also be relevant differences in wastewater quality in this case, because wastewater from older well pads tends to be more saline and thus may require more treatment prior to reuse.

6.2 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

---

[45] "Smaller" well pads are those that disposed of 10 or fewer truckloads in a month, encompassing about 72% of well pad-months in the sample (but only 9% of wastewater volume).

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 witihin 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 demand for resources in the sharing market. Price effects change the relative benefits of outsourcing particular transactions. To illustrate the separate impacts of each channel, Figure 13 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.

6.3 Additional results

The model also delivers estimates for other intrinsic costs within the firm, including the scale of latent cost heterogeneity $\sigma_\kappa + \sigma_\delta$. This section briefly summarizes these additional results, which contextualize the sharing frictions, as well as model fit and alternative specifications.

6.3.1 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), which have no effect on the conditional match. Thus, I only consider a small number of covariates that interact characteristics of κ with characteristics of δ. First, costs per barrel are \$0.31 greater when shipping from a well pad where a gel was used to a well pad where slickwater was likely used, and \$0.39 lower in the opposite case. This is the reverse of what one would expect, but this likely reflects (near) collinearity with the covariates included in $z_{\kappa\delta}$. I also find that costs are reduced by \$0.26 when the sending well pad is small and the receiving facility is a CTF. Together, these finding suggest that heterogeneity in wastewater quality can shift treatment costs by modest amounts, and treatment costs may differ somewhat across facility types, but these differences appear to be small in comparison to the variation in transportation costs across potential shipments. Consistent with these estimates, treatment costs of between \$0.25 and \$0.50 per barrel are are typically reported in engineering and policy reports.

6.3.2 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.[46] 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.

6.3.3 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 14 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 $u_\delta^\mathcal{I}$ (which differ for each t). The final row shows the time series of the sharing rate and aggregate mean shipment distance. This provides visual evidence that the aggregate model fit does not mask a poor model fit in specific time periods.

6.3.4 Alternative specifications

Table 9 reports the main coefficient estimate, log likelihood, and other model fit information for several alternative specifications of the model. The main results correspond to specification (2) in the table. Specification (1) is a simplified model with no cost shifters x, no

---

[46] 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).

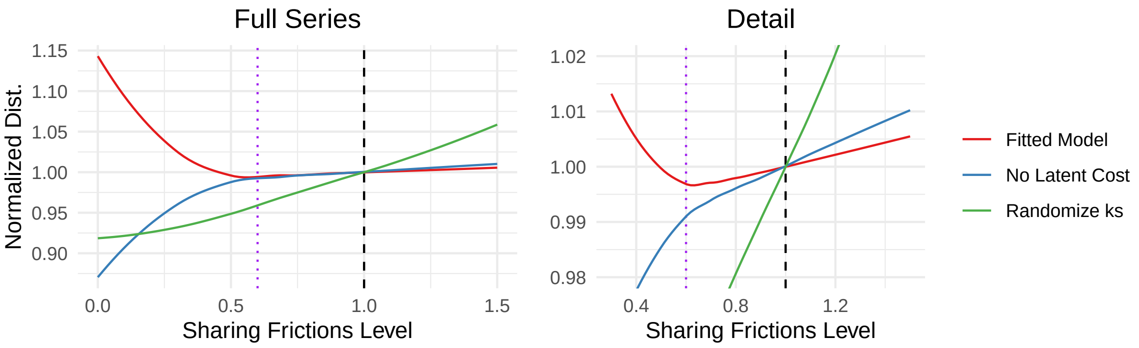
firm-pair fixed effects π, and only a constant included in the friction shifter z. Relative to this model, my preferred specification significantly improves model fit, as evidenced by improvements in log likelihood and the other reported model fit statistics. In the main specification $d_{\kappa\delta}$ 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). Using driving time in hours slightly improves model fit, while the log and non-linear specifications reduce the quality of model fit. Because the difference in model fit between over-the-road distance and over-the-road driving time is small, I report results for the distance-based specification to simplify the policy analysis, which focuses on externalities that scale with truck-miles. Finally, it is natural to suppose that sharing frictions differ with distance. I do not include distance in z 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 on the other hand (especially the latent costs ϵ and η). 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

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.

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, extramarginal matches created when sharing frictions fall increasingly involve non-transportation synergies at greater and greater distances.

The intuition for this finding is illustrated in Figure 15. In the left panel, sharing frictions inhibit many opportunities for sharing at reduced distance. In the right panel, sharing frictions only inhibit sharing that involves long shipment distances. The actual geographic distribution of firms falls between these two extremes: unrealized geographic complementarities are present in some areas but not others. Thus, reducing sharing frictions leads to reductions in shipment distances in some areas, and increases in shipment distances in others, with the net effect market-wide being positive at the estimated parameters.

In the case of the EQT-Rice merger, unrealized ex ante geographic complementarities were presumably large, but I find that the elimination of sharing frictions in particular generated only a modest improvement in transportation efficiency. Table 10 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.[47] 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 wastewater generation and reuse, which are taken to be exogenous.[48]

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 (Pigouvian) 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. A Pigouvian regulator attempts to correct for both (perceived) imperfections at once, either through taxes and subsidies or through other policy interventions (such as quantity restrictions).

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 ϕ. This means that for each $\kappa\delta$, the social planner should internalize $s\phi_{\kappa\delta}$ but not $(1-s)\,\phi_{\kappa\delta}$.[49] In this section, I assume that the social planner observes the estimated cost parameters, the marginal external cost of trucking γ, and the

---

[47] 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.

[48] 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."

[49] Of course, s could differ across $\kappa\delta$. I abstract from this possibility.

welfare relevance parameter s. Holding fixed the overall level of wastewater generation and reuse, the Pigouvian match μ_s^* in state s is the solution to:

$$\min_{\mu \in \mathcal{M}(Q,C)} \Gamma(\mu) + C_s(\mu) \tag{10}$$

where $\Gamma(\mu)$ represents the external costs of trucking under shipment plan μ:

$$\Gamma(\mu) = \gamma \times \left\{ \sum_{\kappa\delta} \mu_{\kappa\delta} d_{\kappa\delta} \right\}$$

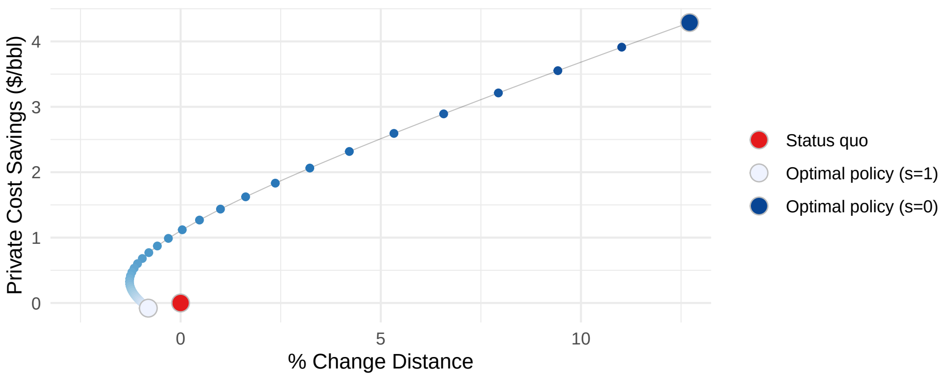
and $C_s(\mu)$ represents the welfare-relevant component of private costs under μ in state s. In Appendix C, I show that the optimal match μ_s^* is implemented by Pigouvian tax on truck-miles where the optimal tax (or subsidy) on shipments between κ and δ in state s is:

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

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 (for example, if sharing frictions reflect some managerial biases), 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 (14) implies that subsidies are greater when sharing frictions are greater, and when shipment distances are smaller. I calibrate $\gamma = 0.07$ based on the calculation in Section 2.4, such that $tax_{\kappa\delta}^{(1)} = 0.07$. In the case that $s = 0$, the expected tax is -2.79 (a large subsidy).

The Pigouvian tax and subsidy scheme (11) minimizes the sum of the external costs of trucking and firms' private costs (the objective in (10)). This sum is weakly lower under the optimal policy relative to the status quo. However, depending on the value of s, one of either external costs or firms' private costs can increase. 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%. Thus, the Pigouvian policy may exacerbate one market imperfection to address another. Firms' private costs are monotonically decreasing in s, but the relationship between s and mean shipment distances is non-monotonic: for any $s \in (0.30, 1.00)$, the optimal policy results in lower shipment distance than under a uniform tax alone. On the other hand, for any $s < 0.30$,

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.

large subsidies are implied by (11), increasing transportation intensity.

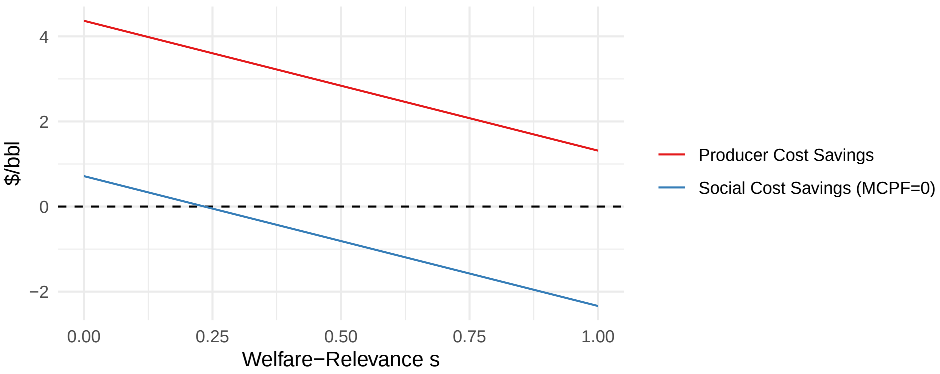
8.1.1 Marginal impact of trucking taxes

The optimal policy (11) consists of a uniform tax on trucking and a set of sharing subsidies. The incidence of the uniform tax specifically is sensitive to the effective level of frictions after subsidies. When sharing frictions are present, the uniform tax reduces shipment distances by 0.2 miles (or 0.8% of the mean). When sharing frictions are absent (for example, under full sharing subsidies), the uniform tax reduces shipment distances by 0.7 miles (or 2.4% of the mean). Thus, the impact of a uniform tax on trucking is muted in the presence of sharing frictions; conversely, sharing subsidies amplify the impact of a uniform tax on trucking.

8.1.2 Program cost and net social welfare impacts

Figure 16 shows the sum of producer and external cost savings under $tax_{\kappa\delta}^{(s)}$ against program cost. When s is close to one, tax revenue from the uniform tax exceeds the cost of subsidies. For smaller values of s, however, the optimal policy (11) may be welfare negative depending on the cost of public funds. The blue curves indicate the cost of implementing $tax_{\kappa\delta}^{(s)}$ under different assumptions regarding the marginal cost of public funds λ. If the marginal cost of public funds is calibrated to a reasonable level ($\lambda = 0.3$), a tax and subsidy scheme like $tax_{\kappa\delta}^{(s)}$ is unlikely to be justified for small s, even if gross tax impacts are large.

Figure 8: Full Sharing Subsidies vs. Uniform Tax Under Misspecification



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.

8.2 Pigouvian taxation under ambiguity

It may be difficult for the social planner to obtain the information needed to specify the optimal policy, even before considering the practical challenges of implementing a complex subsidy scheme. In principle, a firm should be able to compute the welfare-relevant component of sharing frictions $s\phi_{\kappa\delta}$, because $s\phi_{\kappa\delta}$ encompasses real costs that are known or knowable (for example, search costs). Since the social planner can infer $\phi_{\kappa\delta}$, eliciting this information would be sufficient to identify s. However, the form of (11) implies that firms have an incentive to overstate s in order to earn larger subsidy payments. In this context, incorrect inference of s can lead to large policy errors.

To illustrate the incentive problem and the potential for policy regret, Figure 8 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 ($\lambda = 0$). When sharing frictions are not welfare-relevant ($s = 0$), the full subsidy delivers social cost savings of \$0.71 per barrel relative to the uniform tax alone (63% of private trucking costs, which are \$1.13 per barrel). However, producer cost savings are above zero for any value of s. Thus, producers always have an incentive to overstate s. If sharing frictions are in fact welfare-relevant ($s = 1$) but the full subsidy is implemented anyway, social costs increase by \$2.34 per barrel even for $\lambda = 0$. These findings imply that misspecification of the social welfare function is plausible even when ϕ is identified and can lead the social planner to implement policies that result in significant regret.

8.2.1 Robust taxation

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:

$$\max_{\mu \in \mathcal{M}(Q,C)} \min_{s \in \mathcal{S}} - \Gamma(\mu) - C_s(\mu)$$

At the estimated parameters, the uniform tax $tax^{(1)}$ is robust in the sense that it implements the solution $\mu^*_{\mathcal{S}}$. Under this restriction, robustness holds for any parameter vector such that social costs are strictly decreasing in s. In practice, I verify this condition numerically at the estimated model parameters. A proof of the analytical result is provided in Appendix C. The proof follows from an application of the minimax theorem, exploiting the concavity of $\mathcal{E}$. In this sense, the uniform tax $tax^{(1)}$ is more robust than the adjusted tax $tax^{(s)}$ or similar tax rules. Moreover, a uniform tax is simpler to implement than an adjusted 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.

9 Conclusion

Addressing the local environmental impacts of fracking is an important policy priority in oil and gas producing states; addressing these challenges may become yet more important if shale gas production continues to grow in the United States and elsewhere.[50] Fragmentation and decentralization in the shale gas industry create challenges for addressing these impacts. In the particular setting that I study, I find that sharing frictions primarily serve to inhibit socially inefficient trade at long distances, but the magnitudes of the estimated frictions are large enough to suggest that sharing frictions could exacerbate external costs in other environments. Although wastewater management is only one component of the shale gas value chain, it is plausible that similar frictions exist elsewhere (for instance, with respect to gas gathering pipelines, a major source of methane emissions).

The theoretical significance of interfirm transaction costs has long been recognized in economics (since Coase (1937)), but transaction costs are typically not incorporated into empirical analysis in modern industrial organization and environmental economics, largely due to a lack of data. I show that the presence of meaningful firm boundaries can mute or

---

[50] Indeed, the largest shale reservoirs are the Vaca Muerta in Argentina and the Sichuan Basin China, where development has only recently begun to accelerate.

amplify external costs in a similar manner to strategic incentives under oligopoly; in each case, market imperfection leads to behavior that may not be anticipated in the simplest models of environmental regulation in which production is frictionless and exogenous.

Accounting for firm structure can therefore help regulators design better environmental policies. My results suggest that practical interventions concerning quality disclosures, liability rules, and pre-registration of drilling activities are likely to have larger effects on sharing activity than market design interventions that target search frictions alone, such as the establishment of a digital platform for exchanging wastewater. Nevertheless, the benefits of any interventions are likely modest at best, and could be counterproductive, due to the geographic configuration of firms and the misalignment between private and social costs.

Economists typically view market imperfections as problems to be solved. I use a Pigouvian framework to formalize the tradeoff between private and social welfare in the presence of market imperfections. This exercise raises an important question for welfare analysis: when do transaction costs represent real, economic costs of production, and when do they represent internalities, from the perspective of the social planner? In general, this question may prove difficult to answer, just as it has proven difficult to answer in the context of a variety of consumer markets (Handel and Schwartzstein, 2018; Goldin and Reck, 2018). I show by example that formalizing and incorporating normative ambiguity into the social planner's problem, rather than ignoring it, may be one productive path forward.

The policy analysis in this paper can be extended in a few ways. First, with improved data (or stronger assumptions) 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 is natural to consider whether quantity regulations have any advantages over corrective taxation in the presence of sharing frictions, although the more general question of welfare-relevance is not specific to tax-and-subsidy schemes. Finally, 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 hope to address these questions in future work.

References

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

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

Antras and Chor (2022) “Global Value Chains,” *Handbook of International Economics.*

Atalay, Hortacsu, Syverson, and Li (2019) “How Wide is the Firm Border?” *Quarterly Journal of Economics.*

Berry, Levinsohn, and Pakes (1995) “Automobile Prices in Market Equilibrium,” *Econometrica*, 63 (4), 841–890.

Black, Boslett, Hill, Ma, and McCoy (2021) “Economic, Environmental, and Health Impacts of the Fracking Boom,” *Annual Review of Resource Economics.*

Bonnet, Galichon, Hsieh, O’Hara, and Shum (2022) “Yogurts Choose Consumers? Estimation of Random-Utility Models via Two-Sided Matching,” *The Review of Economic Studies*, 89 (6), 3085–3114.

Bresnahan and Levin (2012) “Vertical Integration and Market Structure,” *NBER Working Paper.*

Buchanan and Stubblebine (1962) “Externality,” *Economica*, 29 (116), 371–384.

Carlton (2020) “Transaction costs and competition policy,” *International Journal of Industrial Organization.*

Chassang (2010) “Building Routines: Learning, Cooperation, and the Dynamics of Incomplete Relational Contracts,” *American Economic Review*, 100 (1), 448–65.

Chiong, Galichon, and Shum (2016) “Duality in dynamic discrete-choice models,” *Quantitative Economics*, 7, 83–115.

Choo and Siow (2006) “Who Marries Whom and Why,” *Journal of Political Economy*, 114 (1), 175–201.

Clay, Jha, Muller, and Walsh (2017) “The External Costs of Transporting Petroleum Products by Pipelines and Rail: Evidence From Shipments of Crude Oil from North Dakota,” *NBER Working Paper.*

Coase (1937) “The Nature of the Firm,” *Economica.*

——— (1960) “The Problem of Social Cost,” *The Journal of Law & Economics.*

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

Copeland, Shapiro, and Taylor (2022) “Globalization and the Environment,” *Handbook of International Economics*, V, 61–146.

Covert (2015) “Experiential and Social Learning in Firms: The Case of Hydraulic Fracturing in the Bakken Shale,” *Working Paper.*

Covert and Kellogg (2023) “Environmental Consequences of Hydrocarbon Infrastructure Policy,” *Working Paper.*

Dahlman (1979) “The Problem of Externality,” *The Journal of Law and Economics*, 22 (1), 141–162.

Decker, Lieb, McCann, and Stephens (2013) “Unique equilibria and substitution effectsin a stochastic model of the marriage market,” *Journal of Economic Theory*, 148, 773–792.

Demsetz (1988) “The Theory of the Firm Revisited,” *Journal of Law, Economics, & Organization*, 4 (1), 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).”

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

Galichon and Salanie (2022) “Cupid’s Invisible Hand: Social Surplus and Identification in Matching Models,” *Review of Economic Studies.*

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

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

Goldin and Reck (2018) “Rationalizations and Mistakes: Optimal Policy with Normative Ambiguity,” *AEA Papers and Proceedings*, 108, 98–102.

Graham (2011) “Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers,” *Handbook of Social Economics*, 1B.

——— (2013) “Comparative static and computational methods for an empirical one-to-one transferable utility matching model,” *Mimeo.*

Graham, Irving, Tang et al. (2015) “Increased traffic accident rates associated with shale gas drilling in Pennsylvania,” *Accident Analysis and Prevention.*

Gretsky, Ostroy, and Zame (1992) “The Nonatomic Assignment Model,” *Economic Theory*, 2 (1), 103–127.

Grossman and Hart (1986) “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration,” *Journal of Political Economy*, 94 (1986), 691–719.

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

Hahn and Stavins (2011) “The Effect of Allowance Allocations on Cap-and-Trade System Performance,” *Journal of Law and Economics*, 54 (4), S267–S294.

Handel and Kolstad (2015) “Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare,” *American Economic Review*, 105 (8), 2449–2500.

Handel and Schwartzstein (2018) “Frictions or Mental Gaps: What’s Behind the Information We (Don’t) Use and When Do We Care?” *Journal of Economic Perspectives*, 32 (1), 155–78.

Harris and Nguyen (2023) “Long-Term Relationships and the Spot Market: Evidence from US Trucking,” *Working Paper.*

Hausman and Kellogg (2015) “Welfare and Distributional Implications of Shale Gas,” *Brookings Papers on Economic Activity*, 71–125.

Head and Meyer (2014) “Gravity Equations: Workhorse, Toolkit, and Cookbook,” *Handbook of International Economics*, 131–195.

Heo, Adams, and Gao (2016) “Reduced-form modeling of public health impacts of inorganic PM2.5 and precursor emissions,” *Atmospheric Environment.*

Ho and Lee (2019) “Equilibrium Provider Networks: Bargaining and Exclusion in Health Care Markets,” *American Economic Review*, 109 (2), 473–522.

Hodgson (2022) “Trade-ins and Transaction Costs in the Market for Used Business Jets.”

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

Idel (2016) “A review of matrix scaling and Sinkhorn’s normal form for matrices and positive maps,” *arXiv: Rings and Algebras*, arXiv:1609.06349.

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

Kondash, Lauer, and Vengosh (2018) “The intensification of the water footprint of hydraulic fracturing,” *Science Advances.*

Leslie (2018) “Tax induced emissions? Estimating short-run emission impacts from carbon taxation under different market structures,” *Journal of Public Economics.*

Luxen and Vetter (2011) “Real-time routing with OpenStreetMap data,” *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems.*

Macchiavello and Morjaria (2015) “The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports,” *American Economic Review*, 105 (9), 2911–45.

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

MacKay (2022) “Contract Duration and the Costs of Market Transactions,” *American Economic Journal: Microeconomics*, 14 (3).

Maloney, Baruch-Mordo, 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.*

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

Masten, Meehan, and Snyder (1991) “The Costs of Organization,” *Journal of Law, Economics, & Organization*, 7 (1), 937– 976.

McCurdy (2011) “Underground Injection Wells For Produced Water Disposal.”

Menefee and Ellis (2020) “Wastewater management strategies for sustained shale gas production,” *Environmental Research Letters.*

Montgomery (2013) “Fracturing Fluid Components,” *E!ective and Sustainable Hydraulic Fracturing.*

Muehlenbachs, Staubli, and Chu (2021) “The Accident Externality from Trucking: Evidence from Shale Gas Development,” *Regional Science and Urban Economics.*

Newell and Raimi (2014) “Implications of Shale Gas Development for Climate Change,” *Environmental Science and Technology*, 48 (15), 8360–8368.

Preonas, Louis (2023) “Market Power in Coal Shipping and Implications for U.S. Climate Policy,” *Mimeo.*

Samuelson and Stacchetti (2017) “Even up: Maintaining relationships,” *Journal of Economic Theory*, 169, 170–217.

Shapley and Shubik (1972) “The assignment game I: The core,” *International Journal of Game Theory.*

Small, Stern, Bomberg et al. (2014) "Risks and Risk Governance in Unconventional Shale Gas Development," *Environmental Science and Technology.*

Stavins (1995) "Transaction Costs and Tradeable Permits," *Journal of Environmental Economics and Management.*

Sweeney and Covert (2022) "Secrecy Rules and Exploratory Investment: Theory and Evidence from the Shale Boom."

Vreugdenhil (2022) "Booms, Busts, and Mismatch in Capital Markets: Evidence from the Offshore Oil and Gas Industry."

Walsh (2013) "Water Management for Hydraulic Fracturing in Unconventional Resources:Part 1," *Journal of Petroleum Technology.*

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

Wilson (2022) "Moving to economic opportunity: the migration response to the fracking boom," *Journal of Human Resources.*

Wunz Associates (2014) "Summary of Shale Gas Wastewater Treatment and Disposal In Pennsylvania," *Report.*

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

Appendix

A 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 10.

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 Associates (2014) 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

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.[51] 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).

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.

Liability risk measures 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. I classify a firm as "high liability risk" if its residual in this regression is greater than zero; otherwise, I classify a firm as "low liability risk." 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.

---

[51] 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.

B Estimation Details

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:

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

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:

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

For the logit, we can show that:

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

and hence:

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

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:

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

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:

$$
\begin{aligned}
\log\left(\frac{\mu_{\kappa\delta}(\theta;\mathbf{Q},\mathbf{C})}{\sum_{\kappa\delta}\mu_{\kappa\delta}(\theta;\mathbf{Q},\mathbf{C})}\right) &= \log\left(\frac{\exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\pi_b-\sigma_\kappa\log\mu_{\kappa 0}-\sigma_\delta\log\mu_{0\delta})\right\}}{\sum_{\kappa\delta}\exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\pi_b-\sigma_\kappa\log\mu_{\kappa 0}-\sigma_\delta\log\mu_{0\delta})\right\}}\right)\\
&= \log\left(\frac{\exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\pi_b-\sigma_\kappa\log\mu_{\kappa 0}-\sigma_\delta\log\mu_{0\delta})+\log c_n\right\}}{\sum_{\kappa\delta}\exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\pi_b-\sigma_\kappa\log\mu_{\kappa 0}-\sigma_\delta\log\mu_{0\delta})+\log c_n\right\}}\right)\\
&= \log\left(\frac{\exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\pi_b)-\tilde{u}_\kappa-\tilde{v}_\delta\right\}}{\sum_{\kappa\delta}\exp\left\{-(\sigma_\kappa+\sigma_\delta)^{-1}(d_{\kappa\delta}+x'_{\kappa\delta}\beta+z'_{\kappa\delta}\alpha+\pi_b)-\tilde{u}_\kappa-\tilde{v}_\delta\right\}}\right)
\end{aligned}
$$

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

Model specification details

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 covariates 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 acrylimide indicators, as discussed in Appendix A; (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 comparsion 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 A.

---

[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.

Computational Details

I solve (9) using a coordinate descent procedure similar to the IPFP 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:

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

C Policy Analysis Details

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:

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

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:

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

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:

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

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

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

where:

$$f\left(\mu,\alpha\right)=-\Gamma\left(\mu\right)-C_\alpha\left(\mu\right)$$

Note that $\mathcal{M}\left(Q,C\right)$ is a compact, convex set. Observe that $f\left(\cdot,\alpha\right)$ is concave in μ by the concavity of $\mathcal{E}$, while $f\left(\mu,\cdot\right)$ 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 $\alpha = 0$.

D Additional Tables and Figures

Figure 9: Centralized Treatment Facility Locations

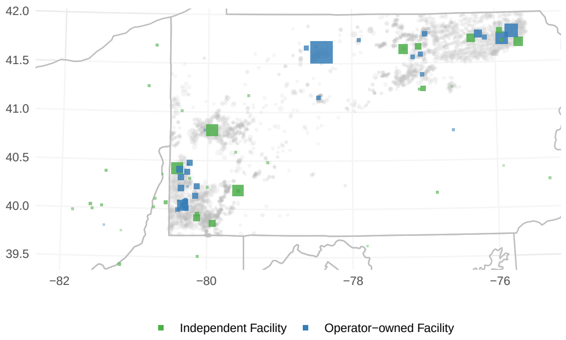
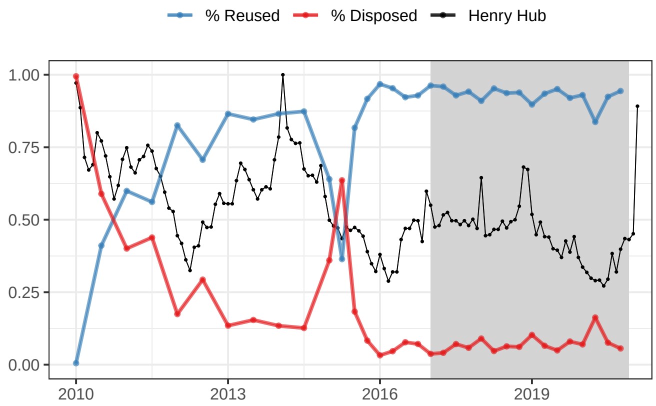
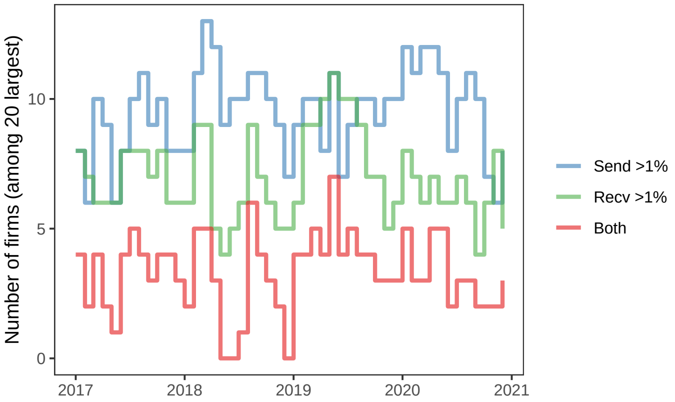


Figure 10: Wastewater Reuse Over Time in PA



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.

Figure 11: Large Firm Sharing Market Participation, 2017-2020



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.

Table 6: Probit Regression of Large Firm Sharing Market Participation on Frac Rate

	Dependent variable:			
	1{Sender}		1{Receiver}	
	(1)	(2)	(3)	(4)
Fracking rate (Z-score)	-0.100^{*} (0.053)	-0.181^{**} (0.071)	0.178^{***} (0.054)	0.293^{***} (0.067)
Firm FE?	No	Yes	No	Yes
Observations	580	580	580	580
Log Likelihood	−398.277	−207.416	−395.872	−248.078
Note:				*p<0.1; **p<0.05; ***p<0.01

Notes: The unit of observation is a firm-month. I restrict attention to the twenty largest firms by disposal volume. I first exclude any months before the first and after the last observed fracking event for each firm. For this exercise only, I also exclude March 2020 and subsequent months due to the Covid pandemic; see Appendix A for further discussion. Columns (1) and (2) report probit coefficients where the dependent variable is a binary indicator for whether a firm sent more than 1% of wastewater to a rival in month. Columns (3) and (4) report probit coefficients where the dependent variable is a binary indicator for whether a firm accepted more than 1% of wastewater from rivals (among observed shipments) in a month. Firm fixed effects capture systematic differences in firms likelihood of participating in the sharing market.

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

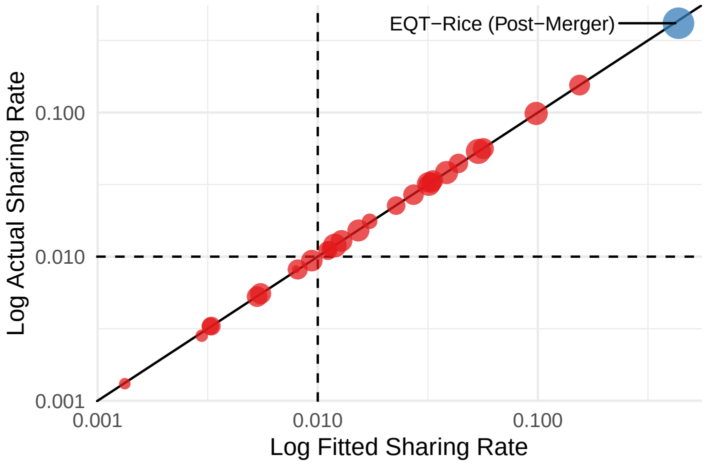


Table 7: Constant Only vs. Full Specification Results (in miles)

	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 × constant	127.8	0.044	5.81			
rival × poor → good env record				-	-	-
rival × good → poor env record				8.5	0.110	0.39
rival × gel → slickwater				-28.6	0.103	-1.30
rival × slickwater → gel				85.3	2.996	3.88
rival × large κ → well pad				-	-	-
rival × large κ → CTF				25.2	0.044	1.15
rival × small κ → well pad				4.4	0.151	0.20
rival × small κ → CTF				29.6	0.261	1.35
Within-firm cost shifters β						
gel → slickwater	-6.7	0.041	-0.31	6.7	0.092	0.31
slickwater → gel	-7.0	0.044	-0.32	-8.7	0.046	-0.39
small κ → 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 $u^{\mathcal{I}}_{\kappa}$ and $u^{\mathcal{I}}_{\delta}$ are not reported, and that κ- and δ-specific covariates in x (such as small well pad or CTF indicators) are not separately identitified from $u^{\mathcal{I}}_{\kappa}$ and $u^{\mathcal{I}}_{\delta}$.

Figure 13: Insourcing Rates for 20 Largest Firms

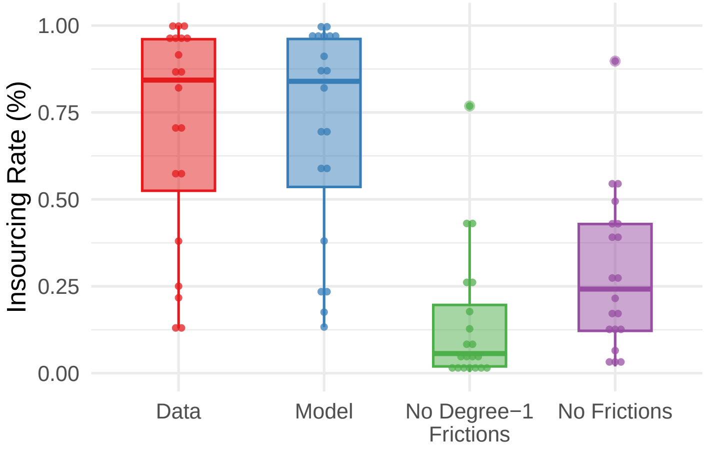
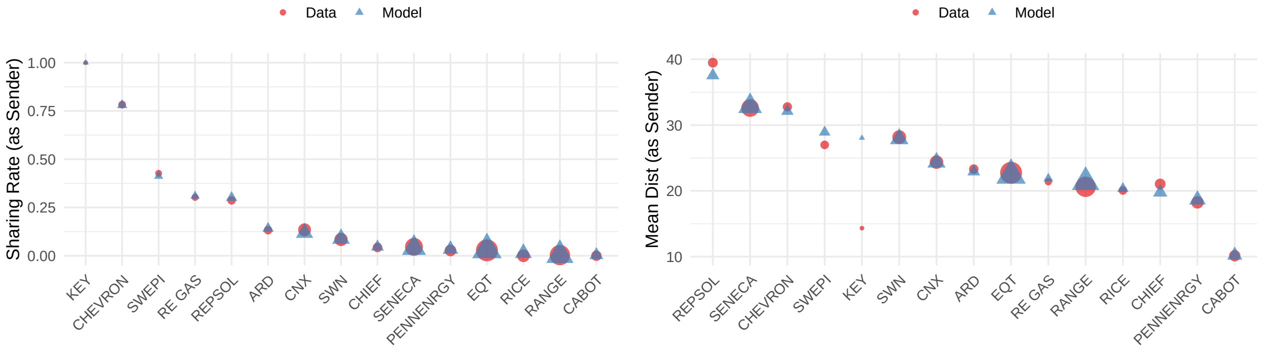


Table 8: Model Fit & Key Counterfactuals

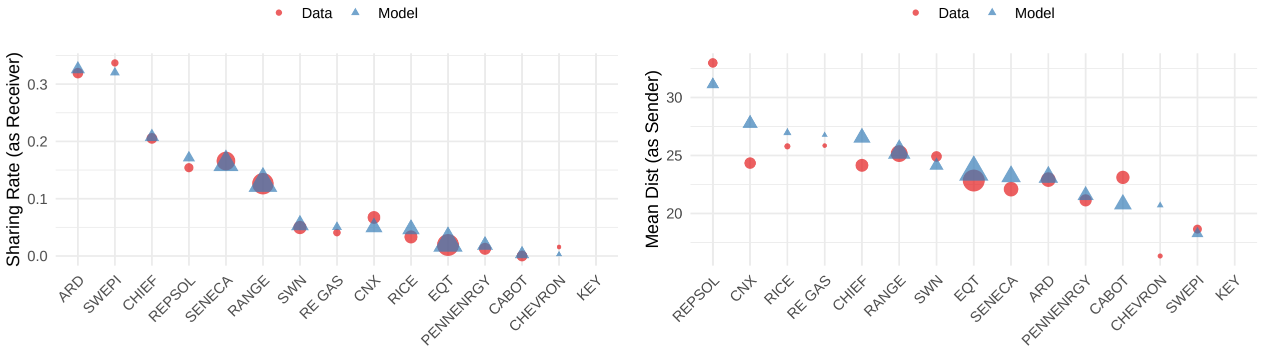
	Mean Dist (mi)	Share %
Data	24.86	10.60
Fitted model	24.85	10.58
$\phi_{\kappa\delta}$ Counterfactuals		
optimal scaling ($\approx 0.60 \times \phi$)	24.78	13.08
high friction ($10 \times \phi$)	28.00	9.49
no friction	28.74	51.37
$\sigma_K + \sigma_D$ Counterfactuals		
$\sigma_K + \sigma_D \to 0$	21.63	9.71
$\sigma_K + \sigma_D \to \infty$	146.99	84.37

Notes: 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.

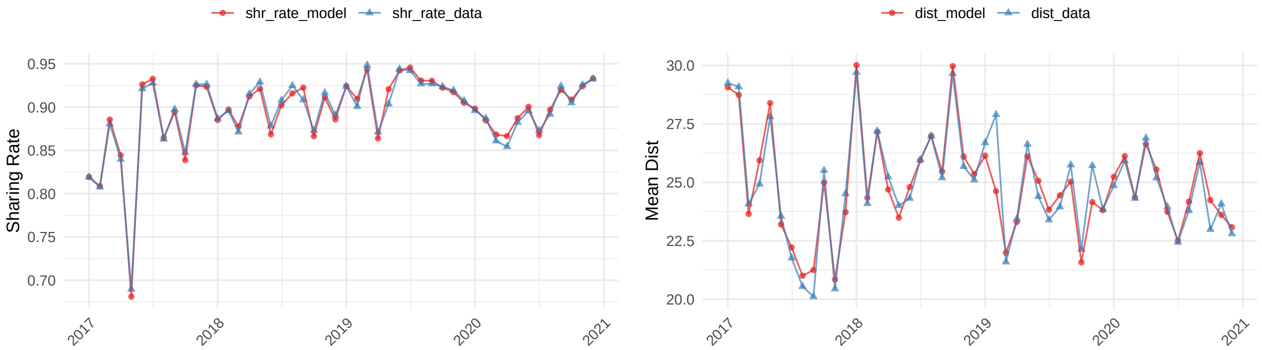
Figure 14: Model Fit Diagnostics


(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)

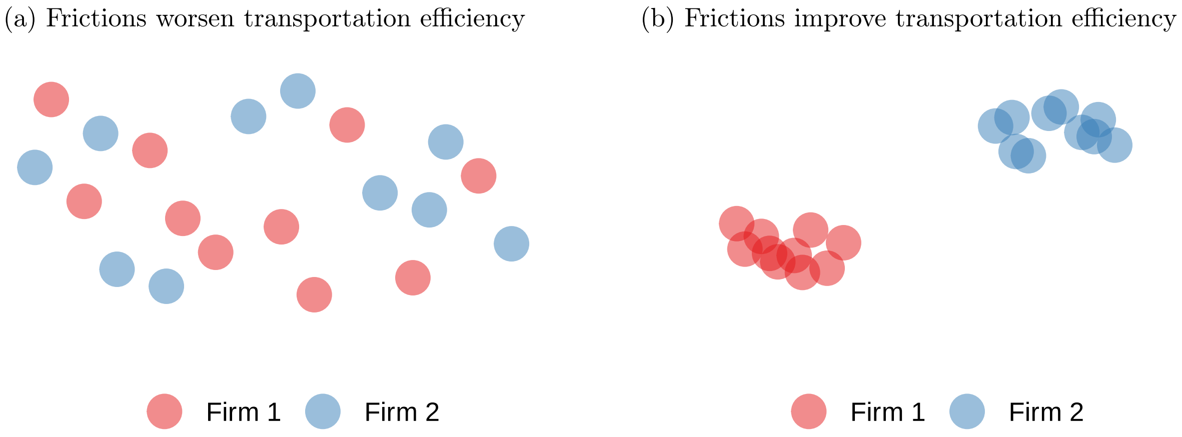
Notes: In the first two rows: figures include the fifteen largest firms by disposal volume (I use fifteen rather than twenty for legibility); firms are sorted by the indicated variable; the size of each dot corresponds to the relevant volume for each firm (volume as sender, or volume as receiver, respectively).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

Table 9: 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 15: Firm Locations and Transportation Efficiency (Illustration)


(a) Frictions worsen transportation efficiency

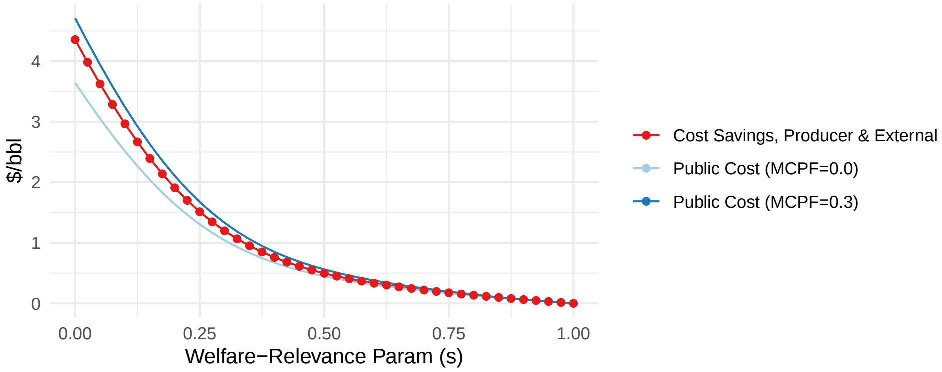
(b) Frictions improve transportation efficiency

Table 10: Counterfactual Shipment Distances & Sharing Rate Within EQT-Rice

	Pre-merger		Post-merger	
	Dist (mi)	Share %	Dist (mi)	Share %
Data	22.62	0.00	21.67	42.58
Fitted model	22.52	4.29	21.68	44.47
EQT-Rice Merger CF				
never merged	22.52	4.29	21.90	20.35
always merged	21.68	34.11	21.68	44.47

Notes: 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 January 2017 to June 2017 (when the merger was initially announced). Post-merger refers to the period from December 2017 (after the merger was completed) to December 2020.

Figure 16: Gross Cost Savings vs Program Costs



Notes: Shows marginal cost savings per barrel vs. public costs under different assumptison regarding the cost of public funds. Note that for sufficiently large s, net subsidies are negative (uniform tax revenues exceed expenditures on subsidies), and the blue lines coincide.