Firm Boundaries and External Costs in Shale Gas Production

Matthew O'Keefe *†

October 18, 2023

CLICK HERE FOR THE LATEST VERSION

Abstract

Shale gas production is highly decentralized, complicating efforts to address the local environmental impacts of fracking. Motivated by this, I analyze optimal environmental regulation when firm boundaries are relevant to external costs. I focus on the market for fracking wastewater in Pennsylvania. In this setting, firms transport wastewater long distances by truck to avoid market exchange, exacerbating greenhouse gas emissions, air pollution, and spill risk. Exploiting regulatory data on firms' wastewater management decisions, I embed frictions at the firm boundary in an empirical model of insourcing and outsourcing. In the model, augmenting a simple uniform tax on trucking with outsourcing subsidies can reduce social costs by up to 64% of private trucking costs if frictions at the firm boundary are welfare-irrelevant internalities. Otherwise, a uniform tax is socially optimal, and results in external costs that are 14% lower. My findings highlight two distinct inference problems for a Pigouvian regulator: the problem of inferring transaction costs, and the problem of assessing their welfare-relevance.

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 of up to several years. Given the scale of

*Northwestern University. Email: MatthewOKeefe2023@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.

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 when they function properly, and may contribute to drinking water contamination and localized seismic activity when they do not (Groundwater Protection Council, 2019; EarthJustice, 2022). Moreover, wastewater is frequently transported by heavy water-hauling trucks, generating air pollution and greenhouse gas emissions while creating significant spill risk (EPA, 2016). Faced with these challenges, several states including Colorado, Oklahoma, and Texas encourage oil and gas producers to *reuse* wastewater from existing wells as a substitute for freshwater when fracking new wells, thereby reducing final disposal volumes and (in many cases) mitigating transportation-related environmental impacts, in addition to limiting freshwater usage.

Even in the absence of policy intervention, principal (or operating) firms often find it cost-effective to reuse wastewater. However, shale gas production is highly fragmented, with a large number of heterogenous operating firms ("firms") and many subcontractors (Small et al., 2014).[1] In this context, firm boundaries can inhibit efficient reuse: opportunities for reuse within a firm may be limited relative to those available when rival firms *share* wastewater. The level of external costs (i.e., negative externalities) therefore depends on the elasticity of substitution between internal reuse and sharing, and on how the external costs of reuse made possible by sharing differ those of internal reuse.

In this paper, I analyze wastewater reuse and sharing in Pennsylvania, where 90% of wastewater is reused due to high conventional disposal costs, and 10% of wastewater is shared prior to reuse. The setting provides a rare opportunity to empirically examine how firms substitute between insourcing and outsourcing (here, internal reuse and reuse via sharing), which enables me to offer a novel perspective on the role of firm structure in mediating external costs. In the first part of the paper I characterize the magnitude of *sharing frictions* and assess their significance for external costs. Then, using the same framework, I explore the implications of sharing frictions for policies intended to mitigate the external costs of wastewater management, focusing on wastewater transportation in particular.

Analyzing records maintained by the Pennsylvania Department of Environmental Protection ("DEP"), I first show that the "sharing market" resembles a spot market in key respects, but that firms' preference for internal reuse is nevertheless strong. To illustrate, I highlight the merger of EQT Corporation and Rice Energy, Inc., which created the largest natural gas

[1] Operating firms include the largest global energy firms (such as Chevron and Shell), large specialized shale firms (such as Southwestern Energy and Range Resources), and a long fringe of small independent firms (sometimes operating only a single well). Moreover, each operating firm contracts many tasks to specialized subcontractors (including well known firms such as Baker Hughes and Halliburton, among others).

producer in the United States. EQT and Rice never shared prior to merging despite clear geographic complementarities that were exploited by the merged entity immediately after consummation, suggesting that ex ante sharing frictions were large. In comparison, many less proximate pairs of firms with lower apparent gains from trade did engage in sharing. This contrast suggests that sharing frictions are heterogenous and that search and negotiation costs alone are not the sole source of frictions, pointing to a role for contracting frictions (Williamson, 1971; Gibbons and Henderson, 2012) or strategic behavior. Using a simple optimal transport benchmark, I show that sharing frictions potentially reduce industry transportation efficiency by up to 34%, but this preliminary finding conflates sharing frictions with non-transportation components of firms' private costs.

In order to estimate firms' private costs, including interfirm sharing frictions, and to draw sharper policy insights, I develop an empirical model of matching with transferable utility that captures many key features of firms' wastewater management decisions. In the model, wastewater managers minimize the cost of disposal for each new truckload of wastewater. At the same time, water sourcing managers acquire the next truckload of input water (whether wastewater or freshwater) at the least possible cost. Sharing frictions shift the joint surplus of potential matches between managers in rival firms, while the scale of latent costs shifts the relative cost of transportation within the firm. Sharing frictions are identified when firms ship wastewater greater distances to avoid sharing, the same source of identification that Atalay et al. (2019) exploit to estimate the "net benefits of ownership" in Census data.

A matching framework is well suited to the setting for a few reasons. While I observe detailed information about quantities and shipment patterns, I do not observe prices. Indeed, since the majority of shipments for reuse occur within the firm boundary, most "prices" are implicit shadow prices for which even firms themselves presumably lack data. The model enables me to impute these shadow prices (as well as explicit market prices) from quantity data alone, similar to how quantities have been used to impute prices elsewhere in empirical industrial organization (e.g., in Cosar et al. (2015)). Moreover, the model naturally captures capacity constraints inherent to the wastewater reuse market, which could otherwise lead to intractability.[2]

I estimate a parametric representation of this model from the data, leveraging the empirical frameworks developed by Choo and Siow (2006) and Galichon and Salanie (2022) for estimating matching models in marriage markets. I find that sharing frictions are three times as large as marginal transportation costs for the average sharing transaction, equiva-

[2]The amount of wastewater generated by any particular well is small, while the amount of wastewater that can be reused at one time in one location is limited by firms' (largely exogenous) drilling decisions. The underlying assumption is that the demand for disposal and the demand for water (whether freshwater or wastewater) are inelastic over short time horizons.

lent 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) and large in comparison to recent estimates of market transaction costs alone (e.g. Hodgson, 2022; MacKay, 2022).[3] The estimates give insight into the sources of sharing frictions. For instance, sharing frictions are greater when inter-operator liability concerns and risks to well productivity are greater. More generally, I find evidence that sharing frictions (whatever they might consist of) are resolvable through integration: changes in shipment patterns subsequent to the EQT-Rice merger can be rationalized by the elimination of sharing frictions. Conversely, only a few pairs of firms share at close-to-integrated rates under the status quo, suggesting that contracting alone (without integration) is typically insufficient to overcome sharing frictions. This suggests that sharing frictions can arise from difficulties in specifying or enforcing formal contracts or in establishing relational contracts. Consistent with this interpretation, I find that bilateral sharing frictions are greater when potential counterparties have better access to alternative sharing partners: when the risk of defection is greater, commitment becomes more difficult, precluding efficient trade.

Next, I consider how the presence of sharing frictions mediates external costs. I focus on the external costs associated with wastewater transportation.[4] Wastewater management in Pennsylvania requires about 500,000 truck trips each year, at external costs of around $7M. In the model, I find that firm boundaries *reduce* these costs by 13%, contrary to what might be expected. This occurs for two reasons: first, firms tend to operate in circumscribed territories. Second, firm boundaries inhibit matching on all components of private costs simultaneously (including, e.g., treatment costs), not just transportation costs. Since each firm's operations are clustered in space, matches within the firm are typically nearer in distance than those between firms. Consequently, if non-transportation costs are uncorrelated with distance, sharing frictions tend to inhibit matching on non-transportation costs at longer distances. Under the status quo, the non-randomness of firm boundaries effectively softens the adverse effects of the underlying misalignment between private and social costs. However, if transportation costs 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. This finding has two implications for local

[3] I use the term "market transaction costs" to designate the costs of searching for and bargaining with a counterparty, as distinct from "policing and enforcement costs" in Dahlman (1979)'s taxonomy of transaction costs. Policing and enforcement costs primarily encompass shadow costs arising from incomplete contracting.

[4] I choose to focus on transportation-related externalities since reuse rates are already high in Pennsylvania, and since the data is better suited to analyzing the elasticity of transportation than the elasticity of reuse in general. Nevertheless, the general lessons that I draw are applicable to consideration of freshwater withdrawal and injection disposal-related externalities. These questions can easily be addressed with my framework with better data (or stronger assumptions).

environmental policy. First, regulators face a tradeoff between the quantity and quality of sharing activity: beyond a certain point, marginal sharing transactions no longer generate external cost savings. Second, market design improvements intended to make sharing easier may bring little benefit, and could even be counterproductive, particularly if transportation is an important source of negative environmental externalities.

Finally, I consider how a Pigouvian regulator should account for the presence of sharing frictions. I argue that a central question for Pigouvian regulation is whether sharing frictions represent real economic costs (e.g., wages) or welfare-irrelevant shadow costs.[5] In the model, I show that the optimal policy in the former case (holding reuse volumes fixed) is a uniform tax on truck-miles equal in magnitude to the marginal external costs of trucking. In the latter case, the optimal policy combines the same uniform tax with heterogenous sharing subsidies, resulting in net subsidies for many shipments. The uniform tax forces firms to internalize the external costs of trucking, while the sharing subsidies externalize the internal costs of sharing. When sharing subsidies are warranted, the potential welfare gains are large: in the case that sharing frictions are entirely welfare-irrelevant, subsidies can reduce social costs by \$0.72 per barrel, or 64% of private trucking costs. However, if the cost of public funds is greater than 19.4%, the subsidy program is not cost-effective. Moreover, welfare losses from poorly calibrated policies can be substantial: if the social planner mistakenly implements subsidies when sharing frictions are welfare-relevant, social costs increase by up to \$2.43 before considering the cost of public funds. In this way, I highlight two aspects of the social planner's inference problem: the social planner not only needs to infer the level of sharing frictions, but also their welfare relevance. In reality, 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), suggesting that the optimal policy involves moderate subsidies that may be difficult for a regulator to determine. I conclude by considering how Pigouvian regulation can be made robust to this type of ambiguity.

The paper is organized as follows. The remainder of Section 1 clarifies the relationship between this paper and prior work. Section 2 describes wastewater reuse in Pennsylvania in greater detail. Section 3 presents evidence of sharing frictions. Section 4 introduces the matching model and Section 5 discusses identification and estimation. Section 6 describes the model estimates, focusing on the sharing frictions and their interpretation. Section 7 discusses the relationship between sharing frictions and external costs. Section 8 discusses Pigouvian regulation. Section 9 concludes.

[5]The distinction between welfare-relevant and welfare-irrelevant costs is most commonly encountered in the literature on switching costs in consumer markets (see, e.g., Handel and Kolstad, 2015; Allcott and Kessler, 2019). To my knowledge, this is the first paper to introduce this distinction when considering managerial decisionmaking in firms.

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.[6] 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 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. Stavins (1995) and Hahn and Stavins (2011) discuss the implications of transaction costs for the efficiency of cap-and-trade markets. In these works and related literature, 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). The make-vs-buy decision was subsequently formalized in transaction cost economics (Williamson, 1971), property rights theory (Grossman and Hart, 1986), and elsewhere. In this paper, I embed 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 pin down the *relative* costs that firms face when comparing insourcing and outsourcing (controlling for production costs).[7] My approach is closely related to the 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

[6] 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.

[7] 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 but conceptually distinct from the object of interest in my analysis (sharing frictions being the difference between market transaction costs and the governance costs of internal production). In recent work, MacKay (2022) and Hodgson (2022) estimate the costs of market transactions in comparison to the costs of inaction in long-term contracting and durable goods markets, without the possibility of insourcing.

manufacturing industries. Another related paper is Masten et al. (1991), who estimate transaction costs using a shipbuilder's component-by-component sourcing decisions. I build on these papers by documenting the extent of heterogeneity in Coasean transaction costs across similar transactions and by incorporating transaction costs into a welfare analysis.

The Coasean view of the firm that I adopt is complementary to the strategic view of the firm typically found 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 that a distinct but closely related class of market imperfections has empirical relevance for the level of external costs and for regulation.

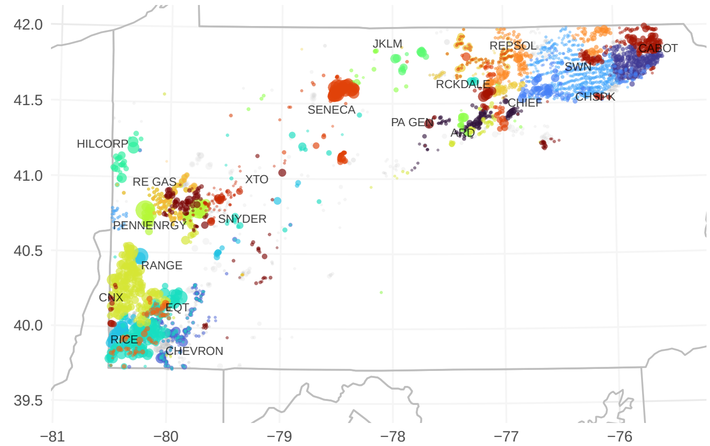
The specific policy question that I highlight relates to the more general problem of regulation under imperfect information explored in seminal work by Weitzman (1974) and Roberts and Spence (1976). This literature often emphasizes that optimal Pigouvian tax schedules can be highly complex, in contrast to textbook models with uniform taxation. In my setting, I find that a uniform tax rate has some favorable robustness properties compared to a more complex tax schedule. Empirical studies by Miravete et al. (2020) and Conlon et al. (2022) demonstrate that uniform liquor taxes have negative distributional effects, but concerns over these effects are less relevant in the oil and gas setting than in consumer markets.

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 with rig operators. Sweeney and Covert (2022) study the public disclosure of production reports in Pennsylvania and West Virginia. In contrast to these papers, I adopt a non-strategic view of the firm, instead focusing on frictions at the transaction level. I also differ in focusing on the external costs that result from market frictions. My policy analysis is most closely related to Covert and Kellogg (2023), who analyze the environmental impacts associated with railroad transport of crude (and non-pipeline hydrocarbon transport generally).

2 Setting

In this section, I briefly describe wastewater reuse in 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.

Figure 1: Locations of Well Pads for Top 20 Firms



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. Oil and gas production in Pennsylvania is conducted by numerous large firms and a fringe of small, independent competitors. 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 wastewater volumes amount to around 50% of the injected volume over the lifetime of a Marcellus well.[8] Wastewater is highly saline and may contain organic compounds, metals, and naturally occurring radioactive materials; consequently, federal regulations require careful handling and

[8]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.

specialized disposal (Groundwater Protection Council, 2019).

In Pennsylvania, wastewater *reuse* creates significant surplus for producers 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 period that I study (2017-2020). 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 (since locations at which reuse can occur are generally nearer than injection well sites) trucking costs are often much lower.[9] 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.[10]

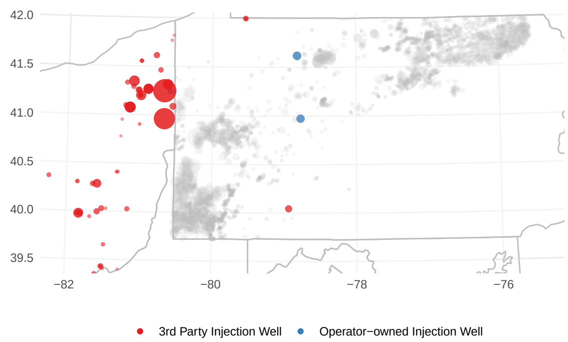
At any point in time, large firms have a stock of hundreds of completed wells producing gas and wastewater, in comparison to a handful of ongoing completions. Wells being completed by a firm may not be located near wells previously completed by that firm. Thus, a firm's demand for wastewater disposal at any particular 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 geographical complementarities (reducing transportation costs). A wide range of prices is possible. I do not observe sharing prices, but anecdotally the sending firm may pay the receiving firm a small "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 (e.g., when freshwater is scarce).[11]

[9] 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.

[10] Reuse outside the oil and gas industry is extremely limited. 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).

[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). This may reflect an effort to avoid costly

Figure 2: Injection Well Locations



2.2 Data

Wastewater disposal in Pennsylvania is regulated by the Pennsylvania Department of Environmental Protection ("DEP"). The DEP requires oil and gas operators to submit monthly reports indicating the disposal method and destination of all quantities of waste materials leaving every well pad, including each barrel of wastewater. These reports are publicly accessible on the DEP website and constitute 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 has a few limitations worth highlighting. First, I do not observe prices, contract terms, or other details of the circumstances in which a shipment was intermediated.[12] (such as whether a shipment was governed by a long-term contract, or what the terms of such a contract might have been, or which subcontractors may have been involved). This limits my ability to precisely identify the organizational mechanisms underpinning the frictions that I study. Another limitation is that I only observe the total volume of water transferred between two locations during a month, rather than the dates, modes, or volumes of particular shipments, which leads me to abstract from timing within the month, mode choice, and less-

legal disputes. For example, since wastewater includes some pre-existing groundwater, landowners might have a claim to revenues from sales of wastewater (in the case that prices are negative).

[12]For example, I do not observe whether outsourced transfers are mediated by direct interaction between two rival operators, or through a third party, but incentives might differ in each of these cases. I also do not observe the existence of long-term contracts or what the terms of such a contract might have been. 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 whether the parent or subsidiary would have retained decisionmaking authority.

than-full truckloads. I also do not observe the locations at which any treatment processes occurred, or if these occurred in different stages at different locations, which limits my ability to control for heterogeneity in treatment costs. The location of reuse is not available prior to 2017, so I focus on the period from 2017 to 2020.[13] Additional details concerning the data cleaning process are included in Appendix A.

2.2.1 Centralized treatment facilities

One limitation of the data requires special treatment throughout the analysis. Specifically, I observe shipments to but not from centralized treatment facilities ("CTFs"). Treatment prior to reuse can occur either directly on a well pad or at a CTF. Treatment on a well pad is more common than treatment at a CTF, but both account for significant shares of the market (I provide market shares in the next section). Some CTFs are operated by oil and gas producers, and others by third parties. Producer-affiliated CTFs are often little more than semi-permanent systems of tanks or impoundments where the same treatments conducted on the well pad can be conducted at a larger scale. Third party CTFs are constructed similarly but may also have technologies that can treat water to higher standards, although these technologies are rarely used in practice. There is ultimately little difference between the standard of treatment on the well pad or at a CTF. In practice, the choice between CTF and on-pad treatment primarily turns on a tradeoff between economies of scale and transportation costs.[14]

Lack of data on shipments from CTFs creates several empirical challenges. First, I cannot perfectly distinguish internal reuse from sharing in the case that wastewater is initially transferred to a CTF. Second, I do not observe how facilities that accept wastewater substitute between direct shipments of wastewater and re-shipments from CTFs (since, for any facility, some unknown volume might have been received from CTFs). Third, since I do not observe the destination of re-shipments from CTFs, I cannot directly calculate the substantial demand for trucking implied by these re-shipments. I address these limitations depending on the question at hand, as I make clear throughout. These challenges are not relevant for shipments to well pads, because re-transfer of wastewater from one well pad to another is prohibited by the DEP.

[13] 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.

[14] Regardless of facility ownership, using a CTF may enable firms to realize some economies of scale, but can increase transportation costs since 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 requirements that factor into this decision, such as differences in bonding requirements.

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

2.2.2 Other data sources

I supplement the DEP wastewater data and related DEP databases with information from several 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 this data set is known to have limitations (for example, firms may fail to submit timely or complete records). I calculate driving distances and driving times between facilities using the Open Source Routing Machine (Luxen and Vetter, 2011) and data from OpenStreetMaps.

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 period that I study, while 8.1% was disposed in injection wells, and 3.3% was disposed by some other means.[15] 80.3% of wastewater was transferred to an internal well pad, internal CTF, or third party CTF, while 8.3% of wastewater was transferred to a rival well pad or a rival CTF.[16] 52.8% of wastewater was transferred directly to a well pad, while 35.9% was transferred to a CTF. Thus, the large

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

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

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 any 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.

Since wastewater production declines over time, the majority of well pads disposing of wastewater dispose of small volumes in comparison the 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 by truckload for each type of receiving facility. The mean shipment distance was 30.0 miles. Since firms' operations are spatially autocorrelated (shown in Figure 1), wastewater that is reused internally is typically shipped a much 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 were 75.5 miles on average. Within the firm, shipments to CTFs are shorter than shipments to well pads (since, as the name implies, these facilities are centrally located), but these shipments would imply subsequent re-shipment at some unknown distance (likely of a similar magnitude).

2.3.1 Participation in the sharing market

I define the *sharing market* to encompass any bilateral exchange of wastewater by oil and gas producers within the state of Pennsylvania for the (presumed) purpose of reuse. The data imply that sharing market volumes are substantial. In this section I briefly describe who shares wastewater, with whom, and under what circumstances. This evidence helps to motivate several key modeling assumptions later in the paper.

In any month, most large firms either send or receive wastewater through the sharing

[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

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 (Antras and Chor, 2022)). Moreover, conditional on participation, firms typically have many distinct sharing partners, suggesting that search costs are low and that any further bilateral fixed costs are likely modest. Among the same large 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 of 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 firms have long-term relationships like those studied empirically in

[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.

Macchiavello and Morjaria (2015, 2021), trade within a firm-pair tends to be short-lived or intermittent: 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 of wastewater and firms who are net demanders in a particular month. Thus, the firms that supply (accept) most of the wastewater in the sharing market do so while accepting (sending) relatively little wastewater in return, unlike what might be expected if firms fully coordinated operations in advance. Unanticipated imbalances might arise due to the inherent lumpiness and unpredictability of firms' water usage and wastewater generation, since only a few fracking events are ongoing at any point in time, fracking is prone to delays, and wastewater volumes are difficult to predict. 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 define as the Z-score of the monthly number of fracking jobs completed by a firm as reported 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 the mean, and 7.1 percentage points more likely to source wastewater on the sharing market when its fracking rate is one standard deviation above the mean.[21] This suggests that firms send wastewater to rivals when fracking less, and then accept wastewater from rivals when fracking more.

2.4 External costs of wastewater trucking

How large are (unpriced) environmental impacts due to wastewater transportation under the status quo? Taken together, the data implies that about 500,000 truckloads of wastewater leave Pennsylvania well pads each year, at an average 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 there are no backhauls, then wastewater

[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.

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.[22] Using the EASIUR air quality model (Heo et al., 2016), I value the social costs of wastewater trucking-related PM2.5 and NOx emissions at $3.3M per year.[23] Using the EPA Social Cost of Carbon for 2020, I value wastewater-related carbon emissions at $3.4M per year. Thus, the social costs of air pollution and greenhouse gas emissions sum to around $0.10 per barrel (at least). In comparison, industry trucking costs are around $5 per mile, implying average transportation costs of $1.35 per barrel. Thus, unpriced air pollution and emissions externalities are 7% of private transportation costs.[24]

External damages from spills are more difficult to assess. 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 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 the heavy truck traffic created by the oil and gas industry, such as elevated traffic fatalities (Muehlenbachs et al., 2021), road damage (Abramzon et al., 2014), and boomtown disamenities such as elevated crime (Mason et al., 2015).

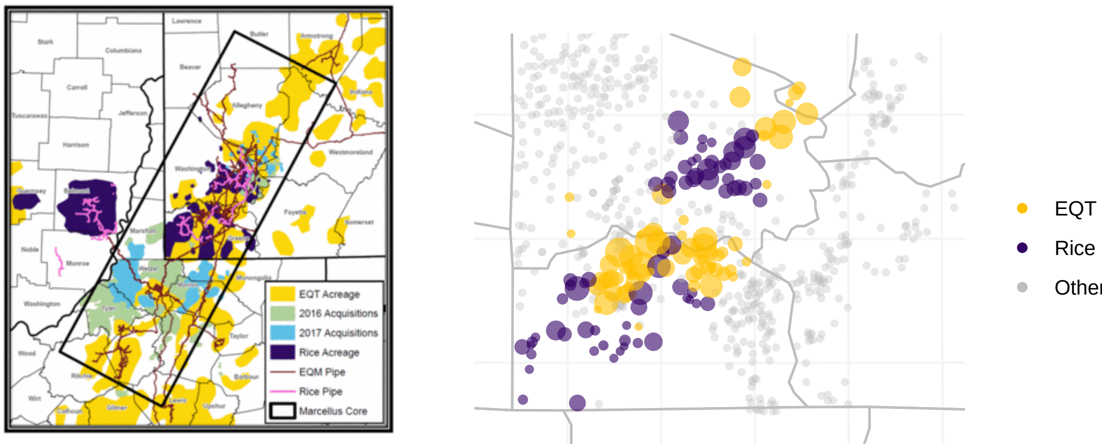
[22] I use the average emissions factors for tanker trucks from EPA SmartWay Carrier data. This data is self-reported and therefore may not be representative for the wastewater-hauling market.

[23] 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).

[24] 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.

Figure 3: EQT and Rice Locations, 2017

(a) Map from EQT presentation (b) Southwest PA well pads in data



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* are an important factor in determining in market-wide shipment patterns, and are therefore relevant to analyzing external costs.

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 are 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, while Figure 3b shows the locations of EQT and Rice well pads appearing in the data. (As the map indicates, EQT also had a significant presence in West Virginia, while Rice was present in Ohio.) 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). Indeed, 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 most importantly because of the potential to drill longer wells by amalgamating existing leases.[25]

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. 87.9% of EQTs wastewater was reused internally or sent to a third party CTF, while 12.1% was reused by 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, with the remaining 97.3% received from internal sources). In contrast, 99.9% of Rice's wastewater was reused internally or sent to a third party CTF, and 0.1% was sent to injection wells. Rice received 4.1% of wastewater from a single rival firm.[26]

After the merger, transfers between formerly-unintegrated EQT and Rice facilities increased dramatically. The second two columns of the left panel of Table 3 indicate market shares for former Rice- and EQT well pads from the completion of the merger in November 2017 to December 2020. After the merger, 23.1% of wastewater generated at EQT-linked well pads was transferred to Rice-linked facilities, and 59.9% of wastewater generated at Rice-linked facilities was transferred to EQT-linked facilities. 39.9% of wastewater received at EQT-linked facilities came from Rice-linked well pads, and 50.2% of wastewater received at Rice-linked facilities came from EQT-affiliated well pads. Thus, the removal of the firm boundary was followed by a significant increase in "sharing," consistent with the elimination of significant ex ante sharing frictions.[27]

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

[25]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)"

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

[27]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.

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

help to distinguish between which of these types of costs contribute most to sharing frictions. Suppose that search costs are diminishing in distance. Then search frictions are unlikely to have been the main factor inhibiting sharing between EQT and Rice, since many firms that did engage in sharing were located much further from one another. Similarly, suppose that negotiation costs are independent of the scale of gains of trade. Then negotiation costs are likewise 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 less significant than policing and enforcement costs, or that more nuanced models of search and negotiation costs required (for example, with decisionmaker biases).

3.2 Heterogeneity and non-transportation costs

The EQT-Rice merger suggests that sharing frictions can be significant. However, the data indicate many firms frequently participate in the sharing market when gains from trade are presumably smaller, suggesting that sharing frictions might be heterogenous. Thus, it is unclear what the aggregate impacts of sharing frictions might be. One way to understand the scale of aggregate impacts is to compare observed shipment patterns to the minimum-distance allocation for the market as a whole. However, this exercise would conflate two distinct forces that contribute to apparently inefficient transportation patterns: sharing frictions on the one hand, and all other components of private costs on the other. 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. There are many other factors that may 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 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).

To understand the scale of non-transportation costs within the firm, I first compare observed shipments for reuse within the firm to two simple benchmarks. The first is random matching, irrespective of distance. The second is the minimum-distance shipment plan within the firm. This plan 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 a particular firm 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_κ at well pad κ and the total volume received C_δ at facility δ. Then the distance-minimizing allocation μ_f solves the following problem:

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

where $d_{\kappa\delta}$ is the distance between κ and δ.[28] Since this problem is a linear program, it can be solved easily. Solving this problem, I find that on average within-firm shipments realize 56% of the maximum possible transportation cost savings relative to random matching. Implementing the minimum-distance shipment plan 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 non-transportation components of cost are economically significant. These results are summarized in the first three columns of Table 4.

Next I repeat the same exercise for the market as a whole. That is, I compute a market-wide random shipment plan and the market-wide optimal shipment plan, this time ignoring any firm boundaries. Compared to the allocation that optimized within-firm shipments alone, the market-wide optimal shipment plan reduces transportation intensity by 4.0 miles per truckload, or by a further 16 percentage points relative to the status quo. These results

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

Table 4: Minimum-distance Benchmark

	Data	Within-firm		Market-wide	
		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	29.1	16.0
Rival	44.0	44.0	44.0	169.4	24.6
Sharing rate	10.2	10.2	10.2	84.1	31.1

are summarized in the last two columns of Table 4. 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 with firms, even though most reuse occurs within firms under the status quo. This suggests that sharing frictions are large relative to non-transportation costs. Moreover, under this allocation, the sharing rate is 31.1% rather than 10.2%, suggesting that sharing rates under the status quo may be inefficiently low. At the same time, it is notable that the market as a whole realizes 95% of the maximum possible transportation cost savings relative to random matching. Intuitively, since firms are geographically isolated from one another, and most shipments for reuse occur within the firm, even relatively inefficient shipments within the firm (from a transportation cost perspective) coincide with an efficient outcome when viewed from the perspective of the market as a whole. Since shipments within the firm are less efficient than the market as a whole by this measure, expanding the scope of the firm or otherwise reducing barriers to sharing could result in less efficient transportation overall to the extent that shipments decisions are driven by non-transportation costs.

4 Model

In this section I develop a model of the wastewater reuse market in order to disentangle transportation costs, sharing frictions, scarcity rents (i.e., markups), and other sources of private cost heterogeneity. Each month, operators report the shipment destination of each barrel of wastewater leaving each well pad. I view this data as the realization of a transferrable utility matching market in which individual truckloads of reusable wastewater are matched to well pads and CTFs at which reuse can occur. Specifically, I adapt the matching framework of Choo and Siow (2006) and Galichon and Salanie (2022) to the setting of the Pennsylvania wastewater market.

Suppose a unit mass of wastewater is generated each month, while a unit mass of water (either wastewater or freshwater) is used for completing new wells or, in the case of a CTF, stored for future use in completing new wells. Let K denote the finite set of all well pads generating wastewater, and D the finite set of all facilities accepting wastewater for reuse (both well pads and operator-affiliated CTFs).[29]

I model the firm as a collection of independent 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 makes decisions truckload-by-truckload, shipping 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 source $\kappa \in K$ or freshwater (the outside option, $\{0\}$), on a truckload-by-truckload basis.

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 *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 vary from truckload to truckload due to *latent private cost*s. The true shadow cost of supplying the jth truckload of demand at δ with truckload i from κ is:

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

where $\epsilon_{i\delta}$ and $\eta_{j\delta}$ represent latent private costs of managers m_κ and m_δ, respectively, which vary from truckload to truckload. These costs capture any real costs or shadow costs that

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

are relevant for manager's decisions regarding specific truckloads of wastewater, but distinct from the systematic costs affecting all shipments between κ and δ. Observe that 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.[30]

If i is not reused, it can be sent to an injection disposal well at cost $\tilde{r}^K_{i0} = r^K_{\kappa 0} - \epsilon_{i0}$, where $r^K_{\kappa 0}$ is the systematic cost of injection disposal from κ. Likewise, freshwater can be obtained for j at cost $\tilde{r}_{0j} = r_{0\delta} - \eta_{0j}$, where $r_{0\delta}$ is the systematic cost of obtaining freshwater at δ.

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 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.[31] I assume that P_K and P_D have full support and finite expectations.

Equilibrium is defined as a stable matching μ of $\mathbf{Q} = (Q_1, \ldots, Q_K)$ to $\mathbf{C} = (C_1, \ldots, C_D)$ describing the mass of wastewater shipped from each $\kappa \in K$ to each $\delta \in D$ when managers allocate each i and each j optimally. Formally, 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 costs of reuse incurred by the sender and $\tau_{\kappa\delta}$ is a (possibly negative) transfer of utility to m_δ ($D_0 = D \cup \{0\}$ encompasses both reuse and the outside option, injection disposal). Note that $\tau_{\kappa\delta}$ is determined in equilibrium, and that utility is perfectly transferrable. This assumption is reasonable since firms can readily exchange cash, 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)). Symmetrically, each manager m_δ solves a discrete

[30] These implications are reasonable empirically: 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.

[31] Relaxing this assumption amounts to allowing for the possibility that sharing frictions differ from truckload to truckload. 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 exist at the level of the facility-month.

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, $\tau_{\kappa\delta}$ is the utility transfer received from m_κ, and $K_0 = K \cup \{0\}$. 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 in the reuse market. Equilibrium is characterized by a matrix of utility transfers τ ensuring that 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). A matching 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.

The existence of a unique equilibrium follows from Galichon and Salanie (2022) Theorem 3.[32] In particular, there is a unique equilibrium matching μ^* that maximizes a social welfare function:

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

where $\mathcal{M}\left(\mathbf{Q},\mathbf{C}\right)$ is the set of feasible matches, and (once again) $\mu_{\kappa\delta}$ is the mass of shipments from κ to δ.[33] In this expression, the first term captures all systematic cost savings from reuse. Intuitively, the match entropy $\mathcal{E}$ captures the surplus contribution of savings on latent adaptation 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,

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

[32] See also Roth and Sotomayor (1990) and Gretsky et al. (1992).

[33] The set of feasible matches $\mu \in \mathcal{M}\left(\mathbf{Q},\mathbf{C}\right)$ consists of all matches 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 $\kappa\delta\in K\times D$.

and

$$H^*(\mu, \mathbf{C}) = \sup_{V \in \mathbb{R}^{K \times D}} \left(\sum_{\kappa \in K} \sum_{\delta \in D} \mu_{\kappa\delta} V_{\kappa\delta} - \sum_{\delta \in D} C_\delta E \left[\max_{\delta \in D_0} V_{\kappa\delta} + \eta_{\kappa j} \right] \right)$$

Intuitively, $G^*(\mu, \mathbf{Q})$ and $H^*(\mu, \mathbf{C})$ quantify the amount of noise 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 (Shapley and Shubik, 1971), 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 $r^K_{\kappa\delta} - \tau_{\kappa\delta} = \frac{\partial G^*_\kappa}{\partial \mu_{\kappa\delta}}$ and $r^D_{\kappa\delta} + \tau_{\kappa\delta} = \frac{\partial H^*_\delta}{\partial \mu_{\kappa\delta}}$, 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. However, an important limitation of the decentralized approach is that I exclude other forms of strategic behavior by firms. For example, a firm cannot earn more surplus by threatening to abstain from sharing. An alternative possibility is to consider a model in which sharing decisions are mediated by Nash-in-Nash bargaining at the level of the firm. However, a Nash-in-Nash bargaining model would be difficult to reconcile with the observed data, insofar as bilateral sharing occurs sporadically and at low volumes, which would be surprising in a model that predicts full surplus extraction conditional on trade.[34]

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, the 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

[34] Carlton (2020) discusses the absence of transaction costs in Nash-in-Nash bargaining models.

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 since 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.

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. 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 and CTFs.

Under these assumption, $\phi_{\kappa\delta}$ is identified whenever $r^{\mathcal{I}}_{\kappa\delta}$ is identified (since 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 for every $\kappa\delta$. I therefore assume that sharing frictions ϕ take the

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

following functional form:

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

where $h_{\kappa\delta}\left(X_{\kappa\delta};\theta\right)$ is a parametric function of observables and π_b is a bilateral fixed effect for sharing between a pair of rivals firms b (the operators of κ and δ). This structure allows for firm pair-specific unobserved heterogeneity (for example, the presence of a relationship), but in a restricted form: sharing frictions are constant within firm pairs, except insofar as they differ with observable match characteristics. Thus, there are no unobserved source 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 include could 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, henceforth "transportation 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 transportation costs, which I assume are known. I focus on the case of linear (rather than log) over-the-road distance since trucking contracts in similar industries often have a per mile component, and since linear distance appears to deliver a good model fit.[36] $x_{\kappa\delta}$ includes controls for differences between sending well pads and receiving facilities that could (possibly) shift the cost of reuse. Specifically, $x_{\kappa\delta}$ includes proxies intended to capture potential differences in fracking fluid formulation and wastewater composition.

Similarly, I also assume that the observed 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 various other covariates to disentangle potential sources of sharing frictions, including measures of wastewater characteristics, facility characteristics (e.g., well pads vs. CTFs), and firm characteristics (such as environmental compliance rates).

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 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 reason. 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. The partial aggregation strategy that I adopt ensures the existence of an estimate for each firm pair fixed effect. The cost of this assumption is that the 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 faster 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.

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

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 this 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 reveals $\mathbf{Q}$ and $\mathbf{C}$. This assumption is analogous to the conventional assumption that market shares are observed without error in demand models. In my setting, this approach is unattractive because $\mu_{\kappa 0}$ and $\mu_{0\delta}$ are observed with noise in the data.[37] 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:

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

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

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

where $\tilde{u} \in \mathbb{R}^K$ and $\tilde{v} \in \mathbb{R}^D$ 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 Thoerem, $\tilde{u}$ 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. 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

[37] 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.

"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).[38] Standard errors can be obtained under the usual maximum likelihood assumptions.

5.3.1 Counterfactuals

Although estimation can be conducted without estimates of $\mu_{\kappa 0}$ and $\mu_{0\delta}$, knowledge of $\mathbf{Q}$ and $\mathbf{C}$ is required for constructing counterfactuals in the specified model. Given this limitation, I construct all counterfactuals under the assumption that there is no outside option on either side of the market.[39] 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 in other shale basins where reuse is less common). Nevertheless, these margins are easily incorporated into the analysis when the outside options are observed (or when ad hoc assumptions to estimate them are justified).

6 Estimates

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

6.1 Sharing frictions

Table 5 presents several key parameter estimates. On average, sharing frictions for shipments between rivals observed in the data are equivalent to the cost of shipping a truckload of wastewater a distance of 135.1 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 (or \$1.98 per barrel), implying that realized sharing frictions are three times as large as realized transportation costs on average. However,

[38] Bonnet et al. (2022) formalize the equivalence between demand inversion and equilibrium computation in two-sided matching models; in the context of inter-firm trade, it is reasonable to assume a price elasticity of one (as I do here). When this is not the case, demand inversion can be re-cast as the problem of solving a two-sided matching market with imperfectly transferrable utility.

[39] This specification 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 estimated sharing frictions are highly heterogenous. As discussed above, I estimate firm-pair specific sharing friction fixed effects π_b for 27 different firm pairs (encompassing 50% of all sharing volume). For most firm-pairs, the expected friction across all shipments (averaging over observable sharing frictions $z'_{\kappa\delta}\alpha$) is between \$3 and \$7 per barrel, but seven firm-pairs had expected frictions below \$3 per barrel and four had expected frictions above \$7 per barrel. The aggregate mean 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, firms would incur mean sharing frictions equivalent to a cost of \$7.38 per barrel.

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 friction-shifting α parameters each give insight into the sources of sharing frictions. I discuss all three perspectives in this section. I discuss covariate construction in Appendix A and additional details concerning the full specification of the model in Appendix B.

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 be fracked with a gel or a slickwater-based fracking fluid, which I infer from the presence of certain key chemicals.[40] 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 \$5.12 per barrel greater (83% 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. This suggests that sharing frictions are greater when firms lack perfect information about the costs and benefits of accepting particular truckloads of wastewater, either because uncertainty regarding quality makes search more difficult ex ante or because the quality of wastewater and the productivity-impact of specific barrels of wastewater is difficult to monitor ex post.

[40] 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	135.1	5.928	6.14
weighted by benchmark	162.3	3.487	7.38
Sharing market cost shifters α			
rival $\times$ gel $\rightarrow$ slickwater	-27.5	0.102	-1.25
rival $\times$ slickwater $\rightarrow$ gel	85.0	2.985	3.87
rival $\times$ low $\rightarrow$ low liability	-	-	-
rival $\times$ low $\rightarrow$ high liability	27.0	0.164	1.23
rival $\times$ high $\rightarrow$ low liability	-15.2	0.250	-0.69
rival $\times$ high $\rightarrow$ high liability	-23.6	0.246	-1.07
rival $\times$ large κ $\rightarrow$ well pad	-	-	-
rival $\times$ small κ $\rightarrow$ well pad	9.4	0.152	0.43
rival $\times$ small κ $\rightarrow$ CTF	25.1	0.260	1.14
rival $\times$ large κ $\rightarrow$ CTF	24.6	0.044	1.12
Within-firm cost shifters β			
gel $\rightarrow$ slickwater	6.8	0.091	0.31
slickwater $\rightarrow$ gel	-8.6	0.046	-0.39
small κ $\rightarrow$ CTF	-5.7	0.129	-0.26
$\sigma_\kappa + \sigma_\delta$	22.5	0.006	1.02
Mean distance (sharing market)			
weighted by data	43.6	-	1.98
weighted by benchmark	14.1	-	0.64

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. The reported mean distances differ from those reported in Section 3 due to the omission of CTFs (explained in the text).

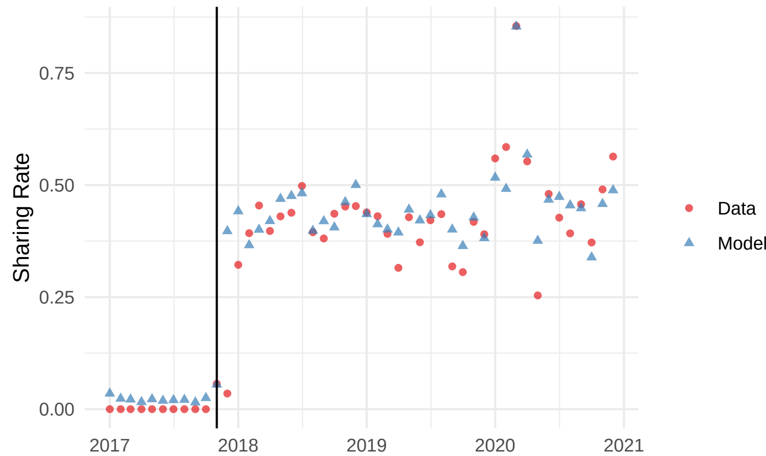
Inter-operator environmental liability When a firm shares wastewater with another firm, it assumes some non-contractible risk that the receiving firm will mishandle the wastewater in some way that exposes the sending firm to civil liability, regulatory fines, or reputational damage. The most serious risk is that fracking fluid could migrate from the receiver's well or temporary storage containers into drinking water resources, which has happened on multiple occasions. Using DEP records, I construct a measure for the relative rates of spills and well site inspection failures among firms (on a per well basis). If inter-operator liability concerns are significant, then sharing frictions should be greater when the firm affiliated with facility δ has a worse compliance record than the firm affiliated with well pad κ (but not necessarily in the reverse case). Consistent with this hypothesis, I find that sharing frictions are \$1.23 per barrel greater when the sending firm has a good compliance record and the receiving firm has a poor compliance record than when both have good compliance records.

Non-integration Figure 4 plots the "sharing rate" across the pre-merger firm border against the rate implied by the model before and after the EQT-Rice merger was completed in November 2017.[41] After the merger, sharing frictions are assumed to be zero. The figure demonstrates that under this assumption, the model successfully rationalizes both the absence of trade prior to the merger (despite the geographic proximity between EQT and Rice), and the ex post level trade between EQT and Rice-affiliated facilities. The mean absolute error in sharing rates is 1.5% for the pre-period and 4.6% for the post-period. 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)).

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 sharing rate in this no-friction counterfactual. The figure shows clearly that for most firm pairs, actual sharing volumes are far below the optimum. At the me-

[41] 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, I cannot estimate a firm-pair specific fixed effect π_b 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).

Figure 4: Change in Cross-Firm Border Sharing after EQT-Rice Merger

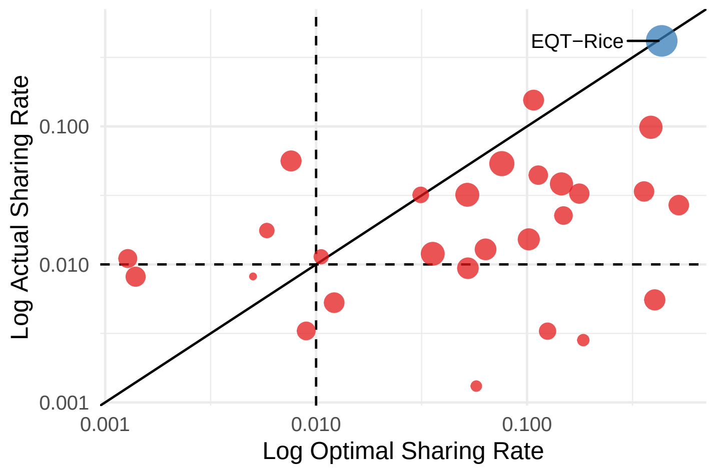


dian, the actual sharing volume was 66% lower than the optimal sharing volume, and only 8 pairs of firms shared more than 90% of the optimal volume. Thus, by comparison to the merger evidence, it is difficult to achieve efficient levels of sharing without integration. This suggests that firms may fail to anticipate the gains from full coordination (for example, if future drilling plans are kept secret), or that the risk of defection is high. 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). To investigate the second problem, I construct a measure of the thinness of the nearby sharing market if firms were to defect from the optimal bilateral sharing agreement.[42] The Spearman correlation of this measure with the expected friction is -0.50, implying that sharing frictions are smaller when the local sharing market is thinner, consistent with challenges in establishing credibility (or policing and enforcement costs, in Dahlman (1979)'s terminology).

Managerial biases 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. Loosely, sharing frictions should be inversely

[42] First, I determine the volumes of wastewater that would be shared under optimal bilateral sharing. Then, I compute the minimum possible shipment distance if these volumes were shipped to other rivals instead. Formally, this is the Wasserstein distance from the hypothetical sharing volumes to "the sharing market." When the nearby sharing market is thicker, this measure is smaller.

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



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. I find that sharing frictions are \$1.14 greater for shipments to CTFs from larger well pads, and \$0.69 greater for shipments from smaller well pads.[43] I also find that sharing frictions are larger for shipments leaving smaller well pads (by \$0.43 for shipments to well pads and by \$0.02 for shipments to CTFs). This could reflect analogous differences in monitoring; however, there may also be relevant differences in wastewater quality in this case, since wastewater from older well pads tends to be more saline and thus may require more treatment prior to reuse.

6.2 Additional results

In addition to the sharing frictions, the model delivers estimates for non-transportation costs within the firm and the scale of latent heterogeneity $\sigma_\kappa + \sigma_\delta$. This section briefly summarizes these additional results, which contextualize the sharing frictions, and discusses model fit.

6.2.1 Non-transportation costs within the firm

Table 5 also 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

[43]"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).

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, I find that costs 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 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 small amounts, and treatment costs may differ somewhat by location, but these differences appear to be small in comparison to the variation in transportation costs across potential shipments. Consistent with these estimates, treatment costs are typically reported as being between \$0.25 and \$0.50 per barrel in engineering and policy reports.

6.2.2 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. 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). Nevertheless, two simple counterfactuals provide a sense for the empirical significance of the estimated value of $\sigma_K + \sigma_D$. Table 7 indicates that the expected shipment distance for the next truckload of wastewater is 15% greater at the estimated value of $\sigma_K + \sigma_D$ than when $\sigma_K + \sigma_D \to 0$. Thus, heterogeneity in latent costs has the same effect on transportation efficiency as increasing all distances by 15%, implying that latent cost heterogeneity is an important component of firms' private costs. In comparison, the expected shipment distance is more than five times greater when $\sigma_K + \sigma_D \to \infty$, suggesting that latent cost heterogeneity generates significantly less match surplus than systematic costs $r_{\kappa\delta}$ in the fitted model (otherwise there would be much less sorting on distance).

6.2.3 Model fit

The first two rows of Table 7 indicate that the model provides a close fit for the aggregate sharing rate and the mean shipment distance observed in the data. Figure 12 presents several additional model fit diagnostics at the firm and monthly level. The first two rows indicate the predicted and observed sharing market participation rates and mean shipment distances for the largest firms, both as sender and receiver. This provides visual evidence that the model successfully captures firm-level sharing market participation as well as firm level mean shipment distances. 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 particular time periods.

6.2.4 Alternative specifications

Table 8 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 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 between κ and δ in miles. Since 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. Since 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 exposition. Finally, it is natural to suppose that sharing frictions might 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 0.5% 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 non-transportation costs 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 the mean shipment distance. The blue line in the figure indicates the mean shipment distance in the model at the estimated parameters if the scale of latent costs $\sigma_K + \sigma_D$ were assumed to be zero, while the green line indicates the mean shipment distance if the ownership of each well pad κ were randomly re-assigned. Setting sharing frictions to zero reduces the mean shipment distance by 13% in the first case and 9% in the second. Under the status quo, eliminating sharing frictions increases matching on transportation and non-transportation cost simultaneously. Because firms are geographically clustered, extra-marginal matches increasingly involve non-transportation synergies at greater and greater distances. This would not necessarily be the case if firms were geographically dispersed or if non-transportation costs were relatively less important.

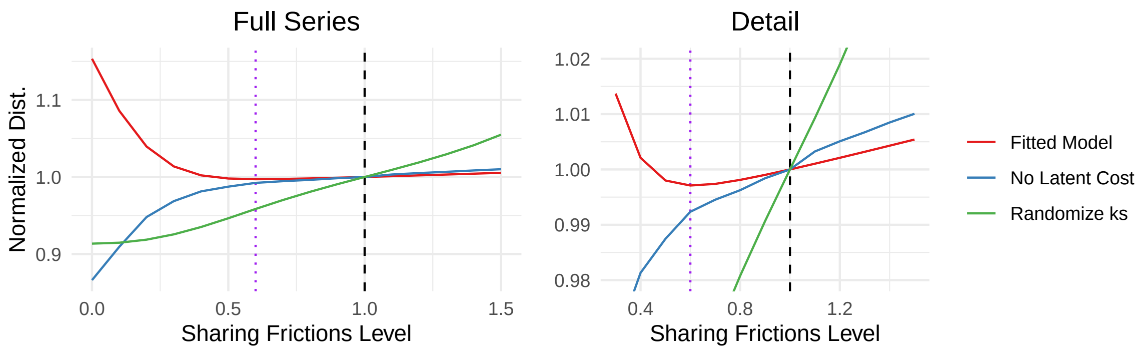
The intuition for this finding is illustrated in Figure 13. In the left panel, sharing frictions inhibit many opportunities for sharing at reduced distance. In the right panel, sharing frictions only inhibit sharing that involves long shipment distances. The actual geographic distribution of firms falls 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 depending on the scale of the reduction.

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 9 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 3.1% after the merger.[44] In the counterfactual with no merger, shipment distances would have fallen by 2.1%. Thus, the elimination of sharing frictions alone explains only 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.[45]

[44]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.

[45]I do not predict how Rice and EQT's drilling activity would have evolved but for 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

Figure 6: Shipment Distance as a Function of Sharing Frictions



These results have a few immediate policy implications. Most immediately, 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, extramarginal sharing could exacerbate external costs.

8 Optimal regulation

The goal of a Pigouvian regulator is not to minimize external costs, but to maximize social welfare. In this section, I derive and analyze the optimal allocation of wastewater, emphasizing how this allocation depends on the specification of firms' private costs.

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 $(1-s)\,\phi_{\kappa\delta}$.[46] In this section, I assume that the social planner observes the estimated cost parameters, the marginal external cost of trucking γ, and the welfare relevance parameter s. Holding fixed the overall level of wastewater generation and

frictions should not be interpreted as a "merger-to-monopoly."

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

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 cost of the next truckload of wastewater shipped under shipment plan μ:

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

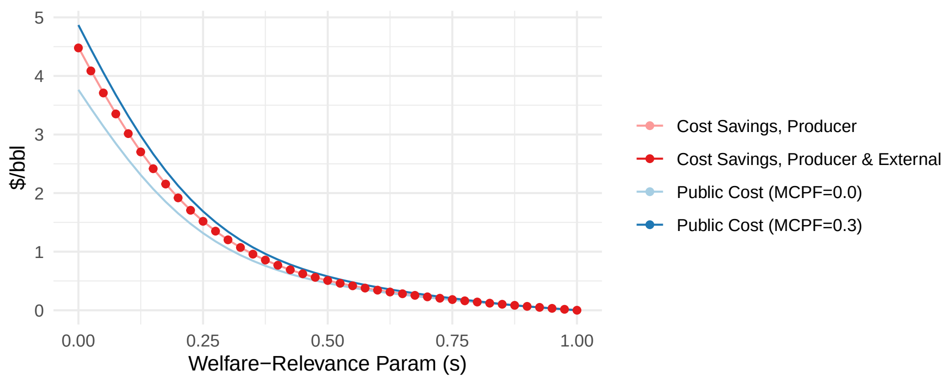
and $C_s(\mu)$ represents the welfare-relevant component of the private cost of shipping the next truckload of wastewater under μ in state s. In Appendix C, I show that the optimal match μ_s^* is implemented by Pigouvian tax on truck-miles where the optimal tax on shipments between κ and δ in state s 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 smaller 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 $tax_{\kappa\delta}^{(1)} = \gamma$. In the case that $s = 0$, the expected tax on the next truckload is $tax_{\kappa\delta}^{(0)} = -2.89$ (a large subsidy).

In Figure 7, the dark red dots indicate the sum of producer cost and external cost savings from implementing $tax_{\kappa\delta}^{(s)}$ rather than the uniform tax $tax_{\kappa\delta}^{(1)}$ in state s. The light red curve indicates the producer costs alone. Although implementing the optimal tax generates significant social cost savings when s is small, this is almost entirely attributable to producer cost savings. Indeed, when $s = 0$, the optimal tax *increases* external costs by 13.5% relative to the uniform tax. The Pigouvian policy exacerbates one market failure (firms' failure to internalize external costs) to address another (firms' failure to externalize internal costs). However, note that (10) omits the cost of public funds. The blue curves indicate the cost of implementing $tax_{\kappa\delta}^{(s)}$ rather than $tax_{\kappa\delta}^{(1)}$ 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.

Figure 7: Private Savings & Public Costs vs. Uniform Tax

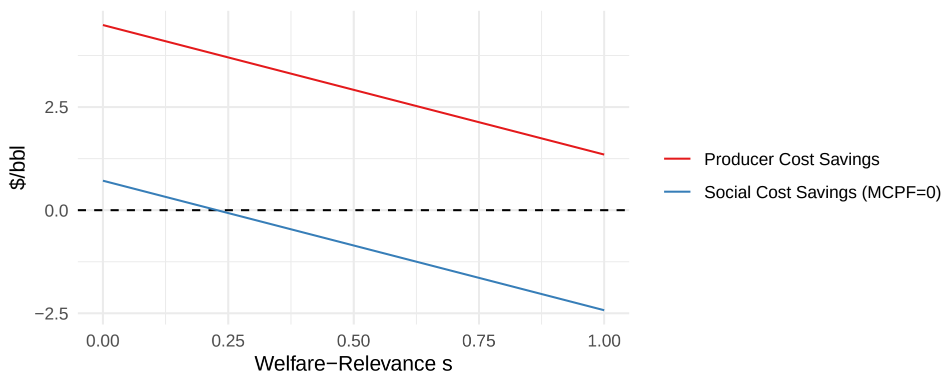


8.2 Pigouvian taxation under ambiguity

Even if the Pigouvian allocation raises the external costs associated with wastewater trucking, correcting market failures in the sharing market may nevertheless be justified from a social welfare perspective. In practice, however, it may be difficult for the social planner to infer s. In principle, a firm should be able to compute the welfare-relevant component of sharing frictions $s\phi_{\kappa\delta}$, since $s\phi_{\kappa\delta}$ 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 $\phi_{\kappa\delta}$ and s. However, the form of (11) implies that firms have an incentive to overstate s in order to earn larger subsidy payments. Thus, even though the social planner can infer $\phi_{\kappa\delta}$ in this setting, it may be difficult for the social planner to obtain the additional information needed to specify the optimal policy, even before considering the practical challenges of implementing a complex subsidy scheme.

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) under the full subsidy $tax^{(1)}_{\kappa\delta}$ relative to uniform tax $tax^{(0)}_{\kappa\delta}$, 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 \$0.72 per barrel relative to the uniform tax alone (64% 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 \$2.43 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 regulations that result in significant policy regret.

Figure 8: Uniform Tax vs. Full Subsidy Under Misspecification



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 outside the United States (indeed, the largest shale reservoirs are the Vaca Muerta in Argentina and the Sichuan Basin

China, where development has only recently begun to accelerate). Fragmentation and decentralization in the US shale 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 (with respect to transportation), 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 (albeit an important one), 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 of this kind 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 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.

Taking firm structure seriously can therefore help regulators to design better 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. Nevertheless, the benefits of market design improvements are modest at best, and could be counterproductive, due to the geographic configuration of firms and the misalignment between private and social costs. Accounting for frictions at the firm boundary can potentially lead to large welfare gains under Pigouvian regulation (albeit at the cost of elevated external costs). However, 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 assuming it away, may be one productive path forward.

The policy analysis that I present can be extended in a few ways. First, with improved data (or more aggressive 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 my framework 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 insights that I draw do not depend on the specific policy instrument used to implement the Pigouvian allocation. Finally, it would be valuable to link sharing frictions to firms' drilling decisions in order to understand how sharing frictions might mediate the broader environmental impacts of shale gas production. 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).

Allcott and Kessler (2019) "The Welfare Effects of Nudges: A Case Study of Energy Use Social Comparisons," *American Economic Journal: Applied Economics*, 11 (1), 236–76.

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.*

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.

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.

Conlon, Rao, and Wang (2022) "Who Pays Sin Taxes? Understanding the Overlapping Burdens of Corrective Taxes," *The Review of Economics and Statistics.*

Cosar, Greico, and Tintelnot (2015) "Borders, Geography, and Oligopoly: Evidence from the Wind Turbine Industry," *The Review of Economics and Statistics*, 97 (3), 623–637.

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.

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

EarthJustice (2022) "Petition," *Petition before the Administrator of the United States Environmental Protection Agency.*

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.*

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, 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.*

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.

Mason, Muehlenbachs, and Olmstead (2015) "The Economics of Shale Gas Development," *Annual Review of Resource Economics*, 7, 269–89.

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*.

Miravete, Seim, and Thurk (2020) "One Markup to Rule Them All: Taxation by Liquor Pricing Regulation," *American Economic Journal: Microeconomics*, 12 (1), 1–41.

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.*

Roberts and Spence (1976) “Effluent charges and licenses under uncertainty,” *Journal of Public Economics*, 5 (3-4), 193–208.

Roth and Sotomayor (1990) “Two-sided Matching,” *Monograph.*

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

Shapley and Shubik (1971) “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.*

Weitzman (1974) “Prices vs. Quantities,” *The Review of Economic Studies*, 41 (4), 477–491.

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 since 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.[47] 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

[47] 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.

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.

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:

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

Galichon and Salanie (2022) Theorem 4 implies that:

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

where:

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

and:

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

For the logit, we can show that:

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

and hence:

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

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

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

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

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

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:

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

which is equivalent to:

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

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:

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

Now consider the system of equations:

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

Clearly, $\tilde{u}_\kappa=\left(\sigma_\kappa+\sigma_\delta\right)^{-1}\psi_\kappa-\left(\sigma_\kappa+\sigma_\delta\right)^{-1}\sigma_\kappa\left\{\log c_n+\log\mu_{\kappa 0}\right\}$ for all κ and $\tilde{v}_\delta=\left(\sigma_\kappa+\sigma_\delta\right)^{-1}\psi_\delta-\left(\sigma_\kappa+\sigma_\delta\right)^{-1}\sigma_\delta\left\{\log c_n+\log\mu_{0\delta}\right\}$ for all δ is a solution to this system. Moreover it is easy to see that $\tilde{u}_\kappa=\alpha-\left(\sigma_\kappa+\sigma_\delta\right)^{-1}\sigma_\kappa\left\{\log c_n+\log\mu_{\kappa 0}\right\}$ for all κ and $\tilde{v}_\delta=-\alpha-\left(\sigma_\kappa+\sigma_\delta\right)^{-1}\sigma_\delta\left\{\log c_n+\log\mu_{0\delta}\right\}$ 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.[48]

Finally, observe that:

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

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

---

[48] To apply the theorem, note that $\exp\left\{-\left(\sigma_\kappa+\sigma_\delta\right)^{-1}\left(d_{\kappa\delta}+x_{\kappa\delta}'\beta+z_{\kappa\delta}'\alpha+\pi_b\right)\right\}>0$ 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 $\tilde{C}_\delta=\sum_\kappa\hat{\mu}_{\kappa\delta}$.

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.[49] 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 5 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 5, z includes additional liability-related dummy variables pertaining to firms for which there was insufficient data to perform the classification described in Appendix A.

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:

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

[49] 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.

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:

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

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

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

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

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

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

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

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

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

where:

$$f(\mu, \alpha) = -\Gamma(\mu) - C_{\alpha}(\mu)$$

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

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

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

D Additional Tables and Figures

Figure 9: Centralized Treatment Facility Locations

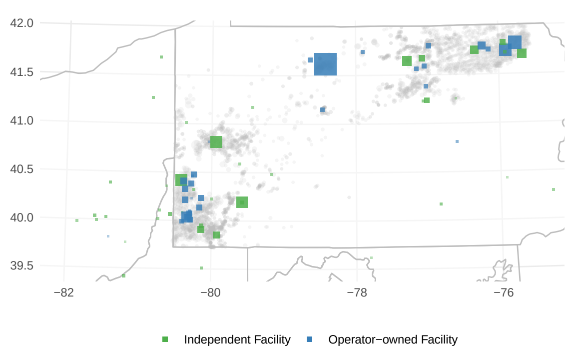
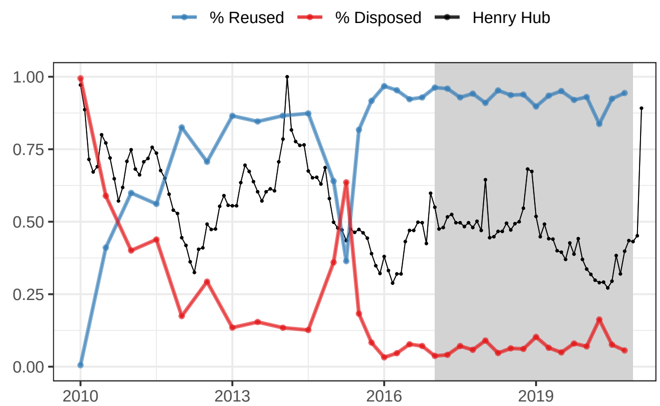
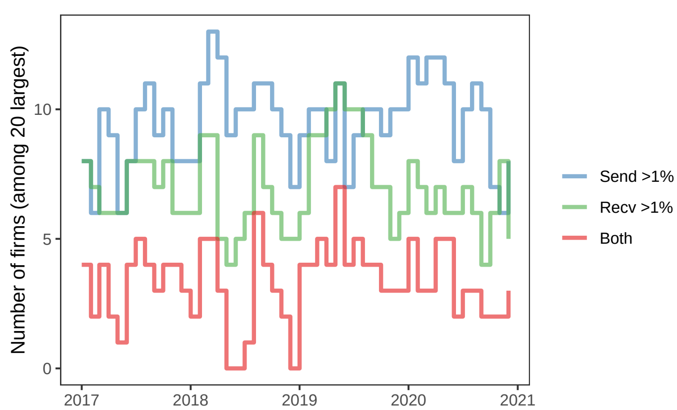


Figure 10: Wastewater Reuse Over Time in PA



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^{*}	-0.181^{**}	0.178^{***}	0.293^{***}
	(0.053)	(0.071)	(0.054)	(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.

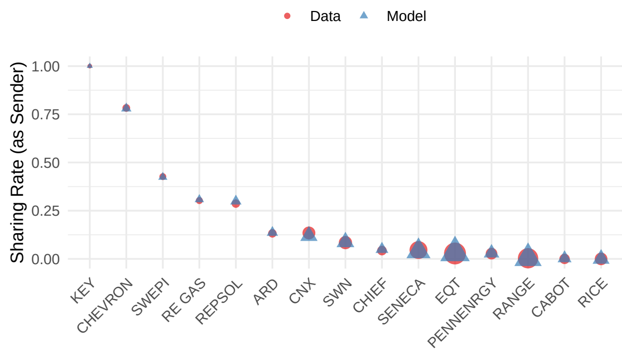
Table 7: Model Fit & Key Counterfactuals

	Mean Dist (mi)	Share %
Data	24.86	10.60
Fitted model	24.86	10.58
$\phi_{\kappa\delta}$ Counterfactuals		
optimal scaling ($\approx 0.60 \times \phi$)	24.79	13.01
high friction ($10 \times \phi$)	28.38	9.51
no friction	28.71	51.34
$\sigma_K + \sigma_D$ Counterfactuals		
$\sigma_K + \sigma_D \to 0$	21.61	9.72
$\sigma_K + \sigma_D \to \infty$	146.99	84.37

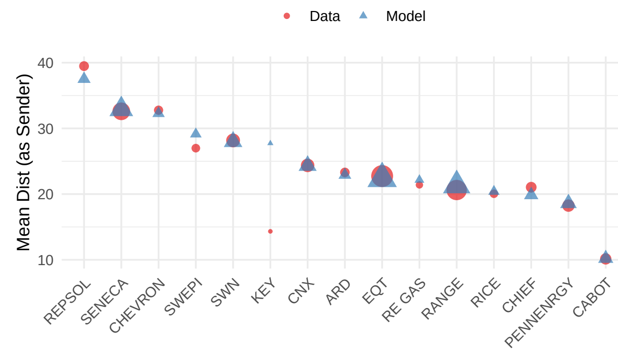
Notes: The ϕ counterfactuals report the mean shipment distance and sharing rate when scaling all estimated sharing frictions by a common factor: respectively, the level that minimizes mean shipment distance (approximately 0.60), ten, and zero. The $\sigma_K + \sigma_D \to 0$ counterfactual is constructed 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 12: Model Fit Diagnostics

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

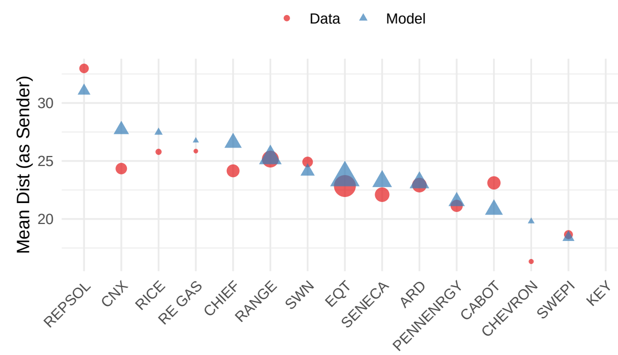


(b) Mean Dist (as sender, by firm)

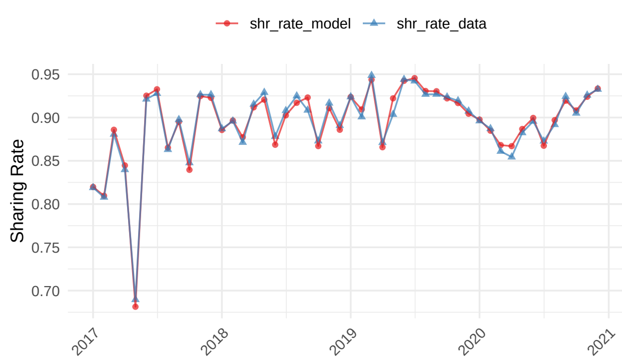


(c) Sharing Rate (as receiver, by firm)

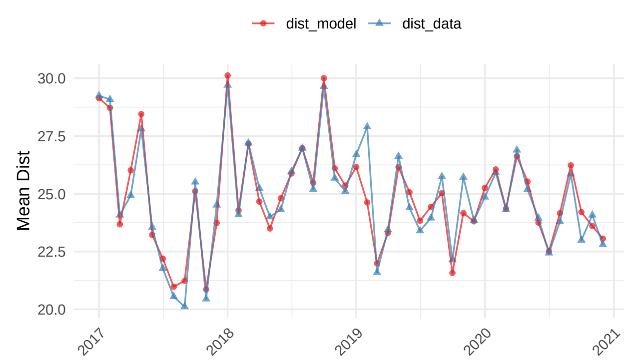
(d) Mean Dist (as receiver, by firm)



(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 8: 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 13: Firm Locations and Transportation Efficiency (Illustration)

(a) Frictions worsen transportation efficiency (b) Frictions improve transportation efficiency

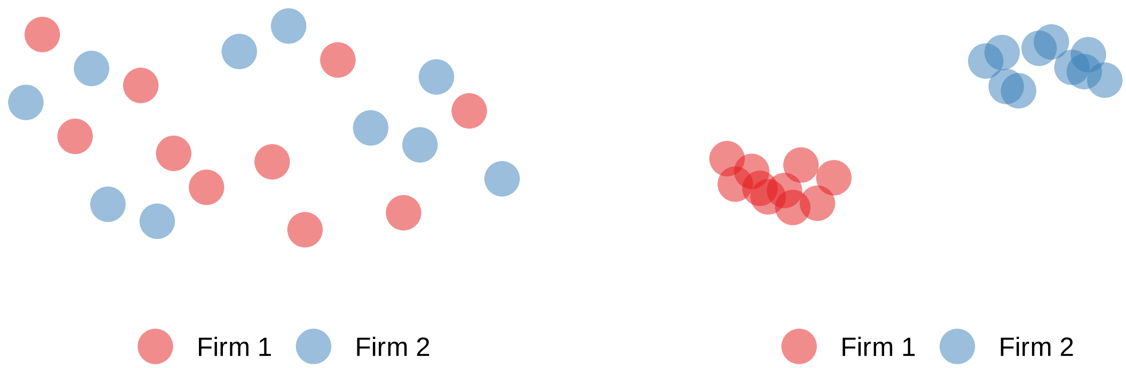


Table 9: Counterfactual Shipment Distances & Sharing Rate Within EQT-Rice

	Pre-merger		Post-merger	
	Dist (mi)	Share %	Dist (mi)	Share %
Data	22.62	0.00	21.67	42.58
Fitted model	22.40	1.66	21.69	44.45
EQT-Rice Merger CF				
never merged	22.40	1.66	21.94	19.50
always merged	21.61	34.55	21.69	44.45

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.