

# Firm Boundaries and External Costs in Shale Gas Production

Matthew O’Keefe <sup>\*†</sup>

October 29, 2023

Updated frequently, please [click here](#) for the latest version.

## Abstract

Wastewater reuse in the shale gas industry mitigates many of the local environmental harms associated with fracking. Most reuse occurs within the boundary of the firm, but rival operators often exchange (or “share”) wastewater prior to reuse. To quantify environmental spillovers from the tradeoff between insourcing and outsourcing, I embed frictions at the firm boundary in an empirical model of wastewater management. Estimating the model with data from Pennsylvania, I find that only 55% of insourcing volume is attributable to intrinsic cost efficiencies as opposed to sharing frictions. However, since firms’ operating acreage is typically contiguous, excessive insourcing reduces transportation-related external costs from emissions and air pollution by 13% in equilibrium. A Pigouvian regulator who views frictions at the firm boundary as welfare-irrelevant distortions optimally provides heterogeneous outsourcing subsidies in addition to corrective environmental taxes, reducing social costs by up to \$0.72 per barrel, but increasing external costs by 14% relative to a policy without subsidies. 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

---

<sup>\*</sup>Northwestern University. Email: [MatthewOKeefe2023@u.northwestern.edu](mailto:MatthewOKeefe2023@u.northwestern.edu)

<sup>†</sup>I am grateful for the generous support and guidance of Rob Porter, Mar Reguant, and Vivek Bhat-tacharya. 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.

returns to the surface as wastewater, over a period of up to several years. Given the scale of fracking activity in the United States and the long-term nature of wastewater production, the transportation and disposal of fracking wastewater pose several challenges for environmental 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](#)). Moreover, wastewater is frequently transported by heavy water-hauling trucks, generating air pollution and greenhouse gas emissions while creating significant spill risk ([EPA, 2016](#)). In light of these challenges, as well as intensive freshwater usage, state regulators typically encourage oil and gas producers to *reuse* wastewater from existing wells as a substitute for freshwater when fracking new wells, thereby reducing final disposal volumes, limiting freshwater usage, and (in many cases) mitigating transportation-related environmental impacts.

Wastewater reuse can occur either within the boundary of firm or through exchange between rival firms. Even in the absence of policy intervention, principal (or operating) firms often reuse wastewater in order to economize on wastewater disposal and freshwater acquisition costs. Exchange between rival firms – known as *sharing* – generates economies when a firms’ disposal needs exceed its capacity for reuse (or vice versa) or when transportation or treatment synergies are present. Since the volume and location of wastewater generation in the current period is determined by past drilling decisions, while new completions occur intermittently and are prone to delays, opportunities for sharing arise frequently. In this context, a reasonable concern for a regulator is whether transaction costs or other frictions in the *sharing market* lead to inefficiently high insourcing rates. A well-functioning sharing market potentially reduces aggregate demand for final disposal (and freshwater) while lessening transportation intensity and storage durations. However, the net environmental impact of sharing is heterogenous across potential transactions, and some types of sharing could worsen environmental externalities (for example, by raising average shipment distances).

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). This enables me to provide new evidence on the relative importance of Coasean transaction costs and intrinsic production costs in shaping firms’ insourcing decisions, as well as a novel perspective on the role of firm structure in mediating external costs. In the first part of the paper I quantify *sharing frictions* and assess their sources. Then, using the same framework, I explore the implications of sharing frictions for statewide external costs and for policies intended to mitigate these costs, focusing on wastewater transportation.

The data for this study consists of wastewater disposal records maintained by the Pennsylvania Department of Environmental Protection (DEP). Using this data, I first establish that the sharing market resembles a spot market in key respects, but that contracting frictions (distinctly from search or negotiation costs) appear to play an important role in shaping firms' wastewater management decisions. This can be seen most clearly in the context of the 2017 merger of EQT Corporation and Rice Energy, Inc., which created the largest natural gas producer in the United States. EQT and Rice never shared prior to merging despite clear geographic complementarities that were exploited by the merged entity immediately after consummation, suggesting that sharing frictions had been large *ex ante*. In comparison, many less proximate pairs of firms with lower apparent gains from trade did engage in sharing in the same period. This pattern would be surprising if sharing participation were primarily constrained by search or negotiation costs, pointing to a role for contracting frictions (Williamson, 1971; Gibbons and Henderson, 2012) or strategic distortions. Using a simple optimal transport benchmark, I show that sharing frictions potentially increase mean wastewater shipment distances by up to 34%, suggesting that distortions are large relative to a world without sharing frictions. However, this preliminary finding conflates sharing frictions with other costs of reuse that vary across potential shipments (for example, treatment costs, or costs related to loading and unloading wastewater from trucks).

In order to separate the intrinsic costs of reuse from sharing frictions, I develop an empirical model of matching with transferable utility that captures key features of firms' wastewater management decisions. Firms simultaneously attempt to minimize disposal costs and water sourcing costs (whether wastewater or freshwater). Previously fracked wells generate only a limited volume of wastewater, while the capacity for reuse in fracking is limited by engineering constraints, creating tradeoffs between disposal and water sourcing costs within the firm. In this context, transactions in the sharing market can crowd out opportunities for internal reuse, and vice versa. I capture these dynamics by assuming that firms make disposal and reuse decisions on a truckload-by-truckload basis in a decentralized matching environment. The firm is a coalition in a matching game; equilibrium transfers within the firm capture the shadow costs of shipments that crowd out more efficient internal shipments or profitable exchanges in the sharing market.

In the model, the joint costs of reuse for a given truckload of wastewater consist of distance-related costs, other intrinsic costs (whether observed or latent) such as treatment costs, and sharing frictions incurred only when reuse occurs across the firm boundary. The sum of these costs is identified conditional on the distribution of the latent costs, as demonstrated by Choo and Siow (2006) and Galichon and Salanie (2022) in the context of marriage markets. I impose functional form restrictions on the intrinsic costs of reuse and the sharing

frictions in order to obtain separate identification of each component of costs. If there are no unobserved, origin-destination specific sources of heterogeneity in the intrinsic costs of reuse, shipment patterns within firms’ boundaries identify the intrinsic costs of reuse in related sharing transactions. The difference between the intrinsic costs of sharing transactions and the shadow costs that rationalize market-wide matching patterns identify the sharing frictions. This strategy is similar to that of [Atalay et al. \(2019\)](#), who identify the “net benefits of ownership” under vertical integration from establishment-level shipment data.

I estimate a fully parameterized model using a maximum likelihood estimator similar to the one proposed by [Galichon and Salanie \(2022\)](#). The estimated sharing frictions are three times as large as distance-related transportation costs for the average sharing transaction, equivalent to a cost of about \$6 per barrel. This estimate is large, similar to prior results on the magnitude of distortions at the firm boundary ([Masten et al., 1991](#); [Atalay et al., 2019](#)) and large in comparison to recent estimates of market transaction costs alone (e.g. [Hodgson, 2022](#); [MacKay, 2022](#)).<sup>1</sup> 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, implying that contracting alone (without integration) is typically insufficient to overcome sharing frictions. This suggests that sharing frictions might arise from difficulties in specifying or enforcing formal contracts or in establishing relational contracts. Sharing frictions directly raise the cost of potential sharing transactions, and indirectly affect transfer prices by mediating market thickness, which governs the relative value of firms’ outside options. The estimates imply that 55% of insourcing volume is attributable to intrinsic costs of reuse, with the remainder attributable to the net effects of sharing frictions.

In the next part of the paper, I consider how the presence of sharing frictions mediates negative environmental externalities created by wastewater management. I focus on the external costs associated with wastewater transportation.<sup>2</sup> 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

---

<sup>1</sup>I use the term “market transaction costs” to designate the costs of search and negotiation, as distinct from “policing and enforcement costs” in [Dahlman \(1979\)](#)’s classification of transaction costs. I return to [Dahlman \(1979\)](#)’s classification of transaction costs in more detail in the body of the paper.

<sup>2</sup>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. These margins are easy to accommodate in my empirical framework with better data (or stronger assumptions).

be expected. This occurs for two reasons: first, firms tend to operate in specific, circumscribed geographies. Second, firm boundaries inhibit matching on all components of private costs simultaneously, not just distance-related 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, sharing frictions inhibit matching at longer distances to a greater extent than they inhibit matching on geographic complementarities. 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 environmental policy. First, the external costs associated with sharing transactions are heterogenous, implying that regulators face a tradeoff between the quantity and quality of sharing activity. Second, market design improvements intended to make sharing easier may bring little benefit, and could even be counterproductive, if transportation is a primary source of local environmental impacts.

The goal of a Pigouvian regulator is not to minimize external costs, but to maximize social welfare. I argue that a central question for Pigouvian regulation in this context is whether sharing frictions represent real economic costs (e.g., wages) or welfare-irrelevant shadow costs.<sup>3</sup> If sharing frictions are evidence of a market failure, then the optimal regulatory policy from the social planner’s perspective must address two market failures at once: the failure of firms to internalize external costs (e.g., from emissions), and the coincident failure of the market to externalize firms’ internalized costs of sharing. Addressing the second market failure may require policies that exacerbate the first. In the model, I show that the optimal tax and subsidy scheme can involve complex sharing subsidies in addition to corrective taxes addressing environmental externalities. The potential welfare gains from such a policy are large: depending on the interpretation of the sharing frictions, subsidies can reduce social costs by up to \$0.72 per barrel, or 64% of distance-related trucking costs for the average shipment. However, nearly all of these gains are the result of producer cost savings, while trucking-related external costs increase slightly. In reality, sharing frictions may consist of some costs that are welfare-relevant (e.g., search and negotiation costs) and some “costs” that are not (e.g., managerial inattention), implying that an optimal subsidy scheme may be difficult to calibrate in practice. Welfare losses from poorly calibrated policies can be substantial: if the social planner mistakenly implements subsidies when a simple corrective

---

<sup>3</sup>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.

tax would be optimal, social costs increase by up to \$2.43 per barrel before considering the cost of public funds. Taken together, these findings demonstrate that the presence of sharing frictions creates a unique set of challenges for the conventional Pigouvian framework. I conclude by considering how Pigouvian regulation can be made robust to one form of normative ambiguity when sharing frictions are identified but difficult to interpret.

## 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.<sup>4</sup> [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. In [Coase \(1960\)](#), transaction costs preclude efficient bargaining between polluting firms and individuals who bear the costs of pollution. In [Stavins \(1995\)](#) and [Hahn and Stavins \(2011\)](#), transaction costs are understood as the costs of trading in a cap-and-trade market, or of paying an emissions tax, or otherwise complying with regulation. In contrast, the frictions that I study in this paper pertain specifically to coordination between firms, and are therefore more closely related to the transaction costs that govern the “make-vs-buy” decision analyzed by [Coase \(1937\)](#) and subsequently formalized in transaction cost economics ([Williamson, 1971](#)), property rights theory ([Grossman and Hart, 1986](#)), and elsewhere. I build on the existing transaction costs literature by embedding the make-vs-buy decision in a structural model of industry costs.

A key reason that relatively few papers in empirical industrial organization have incorporated Coasean transaction costs is the general scarcity of data on firms’ internal operations. My empirical analysis is only possible because I observe insourcing and outsourcing in the same market, which enables me to separately identify the intrinsic cost of production and transaction costs.<sup>5</sup> [Demsetz \(1988\)](#) emphasizes that the make-vs-buy decisions is inherently

---

<sup>4</sup>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.

<sup>5</sup>Border costs in trade are closely related to sharing frictions in my framework ([Anderson and van Wincoop](#),

marginal and that transaction costs may well be heterogenous across potential transactions. Although my analysis is limited to one market, my approach accounts for this richness to a greater extent than the closely related approach of [Atalay et al. \(2019\)](#), who use Census data (which also records within-firm shipments) to quantify the “net benefits of ownership” across a wide range of manufacturing industries. I build on their work by obtaining more granular estimates of transaction costs and by incorporating transaction costs into a welfare analysis. [Masten et al. \(1991\)](#) quantify transaction costs using a selection model and relate their estimates to transaction characteristics; in comparison, I focus on heterogeneity across transactions within a single product market rather than across product markets.

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

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

## 2 Setting

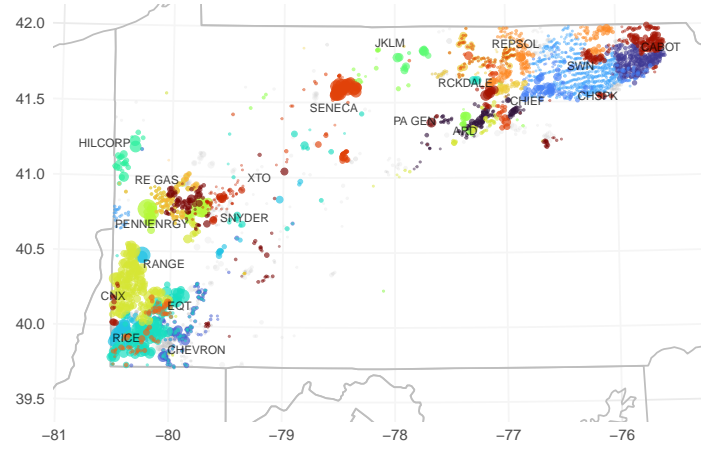
In this section, I briefly describe the economics of wastewater reuse and sharing in Pennsylvania. Then I introduce the data and present key summary statistics on shipment patterns

---

[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.



Figure 1: Locations of Well Pads for Top 20 Firms



and participation in the *sharing market*. Finally, I quantify the environmental impacts of wastewater trucking, the focus of my policy analysis.

## 2.1 Wastewater reuse and sharing in Pennsylvania

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

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



injected volume over the lifetime of a Marcellus well.<sup>6</sup> Wastewater is highly saline and may contain organic compounds, metals, and naturally occurring radioactive materials; consequently, federal regulations require careful handling and specialized disposal ([Groundwater Protection Council, 2019](#)).

In Pennsylvania, wastewater *reuse* creates significant surplus for operators because conventional disposal is costly. *Injection disposal*, which involves using specialized wells to inject wastewater deep below the surface of the earth, is the conventional method of wastewater disposal in the oil and gas industry. In Pennsylvania and West Virginia the underlying geology is not well suited to drilling injection wells ([McCurdy, 2011](#)). Injection wells are common in Ohio, but the distance between Pennsylvania gas wells and Ohio injection wells can be significant. This is illustrated in Figure 2, which shows the location of active injection wells relative to active gas wells in the sample period (discussed below). The costs of injection disposal can be substantial: trucking costs to Ohio disposal wells are \$2-3 per barrel for producing wells in southwestern Pennsylvania and \$10-11 per barrel for producing wells in northeastern Pennsylvania, before disposal fees of \$2-4 per barrel ([Menefee and Ellis, 2020](#)). The costs of wastewater reuse are small by comparison: only a limited amount of chemical treatment and filtering is needed, at a cost of around \$0.25-0.50 per barrel or less, and trucking costs are often much lower.<sup>7</sup> Reuse also reduces the need for freshwater, which would otherwise need to be acquired and transported from local sources at a typical cost of around \$2 per barrel. Note that reuse outside the oil and gas industry is extremely limited.<sup>8</sup>

Treatment prior to reuse can occur either directly on a well pad or at a centralized treatment facility (CTF). Treatment on a well pad is more common than treatment at a CTF, but both are prevalent (I provide market shares in Section 2.3). Some CTFs are operated by oil and gas producers, and others by third party treatment firms. Producer-affiliated CTFs are often little more than semi-permanent systems of tanks or impoundments where the same treatments conducted on a well pad can be conducted at a larger scale. Third party CTFs are constructed similarly but may also have technologies that can treat water to higher standards, although these technologies are rarely used in practice.<sup>9</sup>

---

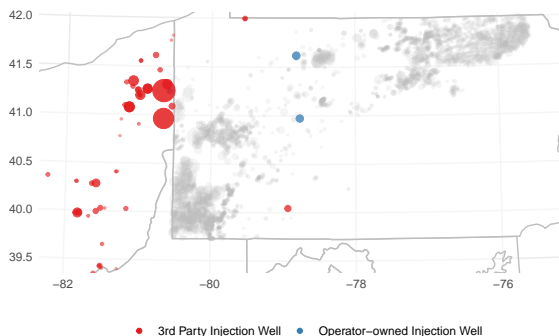
<sup>6</sup>The Marcellus and Utica shales (the main formations underlying Appalachia) are considered “dry” in the sense that relatively little water returns to the surface. In other regions, wastewater generation can be an order of magnitude larger ([Kondash et al., 2018](#)), substantially changing the economics of reuse.

<sup>7</sup>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.

<sup>8</sup>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).

<sup>9</sup>In practice, the choice between CTF and on-pad treatment primarily turns on a tradeoff between economies of scale and transportation costs. Regardless of ownership, the use of CTFs can increase transportation costs since wastewater must be transported twice – once to the CTF, and then again to a location

Figure 2: Injection Well Locations



At any point in time, large firms have a stock of hundreds of completed wells producing gas and wastewater, in comparison to a handful of ongoing completions. Wells being completed by a firm may not be located near wells previously completed by that firm. Thus, a firm’s demand for wastewater disposal at a particular point in time might be larger or smaller than its capacity for reuse, and the transportation costs associated with reuse inside the firm might be significant. In this context, wastewater *sharing* enables firms to reuse wastewater more efficiently, by resolving temporary supply and demand imbalances and exploiting transportation synergies. A wide range of prices is possible. I do not observe sharing prices, but anecdotally the sending firm may pay the receiving firm a “tipping fee” on the order of \$1-3 per barrel in addition to paying for transportation. However, there is often no charge, and negative prices are possible (for example, if freshwater is scarce).<sup>10</sup>

## 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

---

where it can be reused. There are also differences in regulatory compliance costs that factor into this decision, such as differences in bonding requirements.

<sup>10</sup>One explanation for the prevalence of barter is that firms may seek to avoid a commercial designation for wastewater exchanges (Groundwater Protection Council, 2019).

whether the destination facility was associated with the sender in related DEP databases.

The data has a few limitations worth highlighting. First, the data does not include prices, contract terms, or other details of the circumstances in which a shipment was intermediated.<sup>11</sup> This limits my ability to identify contractual mechanisms contributing to sharing frictions. Furthermore, only the total volume of water transferred between two locations during a month is recorded, rather than the dates, modes, or volumes of particular shipments, which leads me to abstract from timing within the month, mode choice, and consideration of less-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. The location of reuse is not available prior to 2017, so I focus on the period from 2017 to 2020.<sup>12</sup> Details concerning the data cleaning process are discussed in Appendix A.

### 2.2.1 Centralized treatment facilities

One limitation of the data requires special treatment throughout the analysis. Specifically, the data records shipments to but not from CTFs. Lack of data on shipments from CTFs creates several empirical challenges. First, it is not possible to perfectly distinguish internal reuse from sharing in the case that wastewater is initially transferred to a CTF. Second, the data does not reveal how firms 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, I cannot directly calculate the substantial demand for trucking implied by re-shipments of wastewater from CTFs. Addressing these limitations requires ad hoc treatment, as I make clear throughout.<sup>13</sup>

### 2.2.2 Other data sources

I supplement the DEP wastewater data and related DEP databases with information from other sources. The DEP requires operators to file fracking records on the public FracFocus database maintained by the Ground Water Protection Council and the Interstate Oil and Gas Compact Commission. I rely on this database to understand the timing of fracking

---

<sup>11</sup>For example, I do not observe whether outsourced transfers are mediated by direct interaction between two rival operators, or through a third party. Incentives might differ in each of these cases. Prior to the merger discussed in Section 3.1, EQT and Rice formally delegated wastewater management to midstream subsidiaries that recorded wastewater disposal revenues on a cost plus basis, but it is unclear to what extent the parent firms would have retained decisionmaking authority under such an arrangement.

<sup>12</sup>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.

<sup>13</sup>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.

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

events and fracking fluid composition, although FracFocus is known to have limitations (for example, firms may fail to submit timely or complete records). The data indicates detailed latitude and longitude information for every well pad and disposal or reuse location. I calculate driving distances and driving times between facilities using the Open Source Routing Machine (Luxen and Vetter, 2011) and data from OpenStreetMaps. One limitation of this approach is that I do not account for roadway-specific vehicle weight restrictions that could alter optimal shipment routes for trucks in comparison to passenger vehicles.

## 2.3 Description of the data

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

Table 1 presents market shares for each of the disposal and reuse methods described previously. I assume that all wastewater transferred to a CTF is ultimately reused. In total, 88.6% of wastewater was reused in the analysis period, while 8.1% was disposed in injection wells, and 3.3% was disposed by some other means.<sup>14</sup> 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.<sup>15</sup> 52.8% of wastewater was transferred directly to a well pad, while 35.9% was transferred to a CTF. Thus, the large majority of wastewater is reused, often but not always after being treated on the well pad, and a significant percentage of reuse occurs across the firm boundary.

In any month, there are many more well pads generating wastewater than facilities receiv-

<sup>14</sup>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.

<sup>15</sup>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.

ing wastewater for reuse. The first section in Table 2 shows the distribution of the number of well pads reporting wastewater transfers and the number of well pads and CTFs appearing as destinations each month. In the average month 1,712.6 distinct well pads reported wastewater transfers, encompassing transfers to 51.6 destination well pads, 11.0 producer-affiliated CTFs and 10.9 independent CTFs. Since wastewater production declines over time, the majority of well pads disposing of wastewater dispose of small volumes in comparison the volume of wastewater required to frack a new well. The second section of Table 2 shows the distributions of monthly volumes by facility type. The mean volume of wastewater per well pad was 23.8 truckloads, but the median well pad disposed of just 3.7 truckloads, and 34% of well pads disposed of fewer than two full truckloads.<sup>16</sup> In comparison, a well pad that received transfers in a given month received an average of 430.8 truckloads (median 29.5). Producer-affiliated and third party CTFs received 907.6 and 465.4 truckloads per month, respectively.

The last section of Table 2 shows the distribution of shipment distance for each type of receiving facility. The mean shipment distance was 30.0 miles. Since firms’ operations are spatially autocorrelated (shown in Figure 1), wastewater that is reused internally is typically shipped a shorter distance than wastewater that is shared. Internal shipments were 22.5 miles on average, while shipments to rivals were 45.0 miles on average. In comparison, shipments to injection wells (primarily from southwestern Pennsylvania) were 75.5 miles on average. Within the firm, shipments to CTFs are shorter than shipments to well pads (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 operating firms within the state of Pennsylvania for the 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. I rely on this evidence to justify key modeling assumptions later in the paper.

In any month, most large firms either send or receive wastewater through the sharing market. Among the twenty largest firms, all but five were active on both sides of the sharing market at some point during the sample, and only one never participated at all.<sup>17</sup> In the

---

<sup>16</sup>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).

<sup>17</sup>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

Table 2: Key Summary Statistics

	Mean	Std	5%	25%	50%	75%	95%
<i>Facility count per month</i>							
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
<i>Truckloads sent or received by facility-month</i>							
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
<i>Miles per truckload by destination type</i>							
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

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.<sup>18</sup> Participation varies considerably over time, as illustrated in Figure 11, but no fewer than 10 distinct large firms were active in each month of the sample period. Thus, many firms share at least occasionally, and participation status is not “sticky,” suggesting that sharing participation does not require large fixed cost investments (in contrast to import-export decisions, e.g. in [Antras and Chor \(2022\)](#)).

Conditional on participation, firms typically have many distinct sharing partners, suggesting that the cost of reaching an agreement with a new counterparty is not excessive. Among the twenty largest firms, firms that sent wastewater to a rival shared with of 1.7 distinct counterparties per month and 7.3 distinct counterparties over the course of the sample (median 7.0), implying that sharing partners frequently change. Likewise, firms that received 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 pairs of firms form long-term relationships like those studied empirically in [Macchiavello and Morjaria \(2015, 2021\)](#), trade within a firm-pair tends to be short-lived or intermittent, which would be surprising if long-term relationships were an important source of surplus: among pairs of firms that ever shared, sharing occurred in only 13.1 of 48 months

the largest firms in terms of wastewater disposal volumes (rather than gas production volumes).

<sup>18</sup>Among all firms, 18.8 firms sent wastewater to a rival, 8.2 accepted wastewater, and 3.5 did both in the average month.

on average, and only 9 pairs of firms shared in more than 24 months.<sup>19</sup>

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 that are net demanders in a particular month. Thus, the firms that supply most of the wastewater in the sharing market do so while accepting relatively little wastewater in return, and vice versa, unlike what might be expected if firms fully coordinated wastewater management in advance. Unanticipated imbalances might arise due to the inherent lumpiness and unpredictability of wastewater generation and fracking activity (which is prone to delays). Consistent with this mechanism, Table 6 presents probit regression coefficients obtained by regressing indicators for whether a firm sent wastewater to a rival or sourced wastewater on the sharing market on a firm’s fracking rate (which I construct from fracking activity recorded in FracFocus). The estimates imply that a firm is 3.9 percentage points more likely to send wastewater to a rival when its fracking rate is one standard deviation below its mean rate, and 7.1 percentage points more likely to source wastewater on the sharing market when its fracking rate is one standard deviation above its mean rate.<sup>20</sup> 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

In the policy analysis, I focus on negative environmental externalities created by wastewater trucking. I find that unpriced carbon emission and air pollution externalities are roughly 7% of private transportation costs.<sup>21</sup> This figure is constructed as follows. 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, including independent 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

---

<sup>19</sup>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.

<sup>20</sup>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.

<sup>21</sup>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.



management generates about 1.0 metric tons of PM<sub>2.5</sub> emissions, 80,000 metric tons of carbon emissions, and 160 metric tons of NO<sub>x</sub> emissions per year.<sup>22</sup> Using the EASIUR air quality model (Heo et al., 2016), I value the social costs of wastewater trucking-related PM<sub>2.5</sub> and NO<sub>x</sub> emissions at \$3.3M per year.<sup>23</sup> 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.

External damages from spills are more difficult to quantify. Spills pose serious ecological risks and can threaten drinking water resources (EPA, 2016). Maloney et al. (2017) find that transportation and storage contribute to 50 wastewater spills per year in Pennsylvania, but it is unclear how many of these spills occur on the road rather than at a well site (for example, during loading or unloading). Statistics from EPA (2016) suggest that the crash rate for tanker trucks is on the order of 100 per 100M truck-miles and the spill rate conditional on crashing is about 5-10%, implying a total of 1-2 spills a year. However, oil and gas-related trucking may differ from trucking in other industries, with drivers perhaps being disproportionately younger, male economic migrants (Wilson, 2022) subject to looser hours of service regulations (Muehlenbachs et al., 2021). Several papers have found evidence that accident rates increase during shale development (e.g., Graham et al., 2015; Xu and Xu, 2020), but Muehlenbachs et al. (2021) fail to find causal evidence of elevated crash rates for trucks (primarily water-hauling trucks) in Pennsylvania specifically. While I focus on environmental externalities, prior work has examined other unpriced externalities associated with heavy truck traffic created by the shale boom, such as elevated traffic fatalities (Muehlenbachs et al., 2021) and road damage (Abramzon et al., 2014).

### 3 Sharing frictions

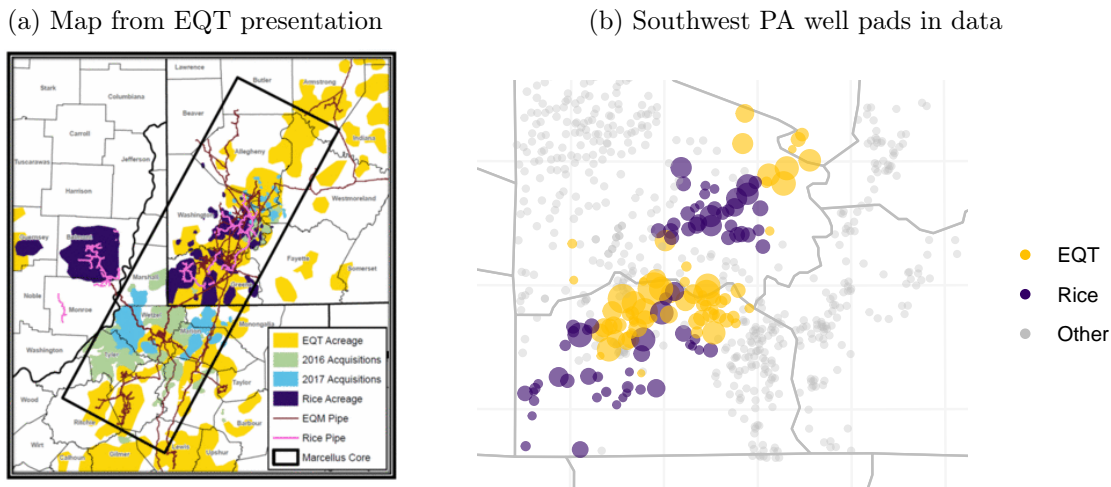
I define *sharing frictions* as transaction costs specific to sharing transactions – real or shadow costs that are incurred if a transaction takes place between rival firms, but not if a transaction takes place within a firm. This section presents evidence that sharing frictions play an important role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

---

<sup>22</sup>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.

<sup>23</sup>To compute air pollution, I currently assume that all trucking-related air pollution occurs at the well site from which the wastewater originated, rather than along the trucking route. This likely results in an underestimate of air pollution damages because well pads are often located in remote areas, whereas trucking routes pass through more populous areas (such as the Pittsburgh metropolitan area).

Figure 3: EQT and Rice Locations, 2017



### 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. Figure 3b shows the locations of EQT and Rice well pads appearing in the data.<sup>24</sup> In the six months leading up to the merger announcement, 99% of Rice’s wastewater volume originated at well pads within 22.5 miles of an EQT facility that accepted wastewater, and 68% of EQT’s wastewater volume originated at well pads within 22.5 miles of a Rice facility that accepted wastewater (22.5 miles being the mean distance of internal shipments in the data). The exceptional geographic proximity between EQT and Rice was the primary stated rationale for the merger, in part because of potential synergies in “rig allocation, pad sites, water, access roads, etc.” and pipeline access, but most importantly because of the potential to drill longer wells by amalgamating existing leases.<sup>25</sup>

<sup>24</sup>As the first map indicates, EQT also had a significant presence in West Virginia, while Rice was present in Ohio.

<sup>25</sup>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

Table 3: Pre- and Post-Merger Market Shares

Share of Wastewater Leaving Well Pad					Share of Wastewater Received				
Destination	Pre-merger		Post-merger		Source	Pre-merger		Post-merger	
	EQT	Rice	EQT	Rice		EQT	Rice	EQT	Rice
EQT pad	83.4	0.0	65.1	62.4	EQT pad	97.3	0.0	59.1	50.5
Rice pad	0.0	70.6	22.5	31.0	Rice pad	0.0	95.9	40.1	49.1
Other rival	8.9	0.0	2.7	0.6	Other rival	2.7	4.1	0.8	0.4
3rd party CTF	7.7	29.3	8.5	4.6					
Injection well	0.0	0.0	1.2	1.4					

Despite apparently large gains from trade resulting from geographic complementarity, EQT and Rice never shared prior to the merger. The first two columns of the left panel of Table 3 summarize disposal market shares for EQT and Rice well pads between January and June 2017. 91.1% of EQTs wastewater was reused internally or sent to a third party CTF, while 8.9% was sent to rivals other than Rice. The right panel indicates that during this period EQT received some wastewater from rivals other than Rice (about 2.7% of total wastewater received). In contrast, all of Rice’s wastewater was reused internally or sent to a third party CTF. Rice received 4.1% of wastewater from a single rival firm.<sup>26</sup>

After the merger, transfers between formerly-unintegrated EQT and Rice facilities increased dramatically. The second two columns of the left panel of Table 3 indicate market shares for former Rice- and EQT well pads from the completion of the merger in November 2017 to December 2020. After the merger, 22.5% of wastewater generated at EQT-linked well pads was transferred to Rice-linked facilities, and 62.4% of wastewater generated at Rice-linked facilities was transferred to EQT-linked facilities. 40.1% of wastewater received at EQT-linked facilities came from Rice-linked well pads, and 49.1% of wastewater received at Rice-linked facilities came from EQT-affiliated well pads. Thus, the removal of the firm boundary was followed by a significant increase in “sharing,” consistent with the elimination of significant ex ante sharing frictions.<sup>27</sup>

Dahlman (1979) distinguishes between three types of transaction costs: search and in-

---

purposes of Item 7.01)"

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

<sup>27</sup>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.

formation costs, bargaining and decision costs, and “policing and enforcement” costs (which might be taken to encompass problems of incomplete contracting). The above patterns can help to distinguish between which of these types of costs contribute most to sharing frictions. Suppose that search costs are declining in distance. Then search frictions are unlikely to have been the main factor inhibiting sharing between EQT and Rice, since many firms that did engage in sharing were located much further apart than EQT and Rice. Similarly, suppose that negotiation costs are independent of the scale of gains of trade. Then negotiation costs are unlikely to have been the main source of sharing frictions between EQT and Rice, since many pairs of smaller and less proximate firms did engage in sharing. The merger evidence therefore suggests that “market transaction costs” (i.e., search and negotiation costs) are not obviously the main source of frictions, at least not without a nuanced model of search and negotiation costs (for example, one with decisionmaker biases). I investigate possible sources of policing and enforcement costs later in the analysis. Some examples include information frictions relating to the quality of wastewater, inter-operation liability concerns, or shadow costs arising from the potential for holdup as discussed in transaction cost economics.

### 3.2 Heterogeneity and other intrinsic costs

The EQT-Rice merger suggests that sharing frictions can have significant effects on wastewater allocations. However, the data indicate many firms frequently participate in the sharing market when gains from trade are presumably smaller, suggesting that sharing frictions might be heterogenous. Thus, it is unclear what the aggregate impacts of sharing frictions might be. Transportation costs are a primary driver of wastewater management costs, so one way to understand the scale of aggregate impacts is to compare the observed allocation of wastewater to the minimum-distance allocation for the market as a whole. If sharing frictions and distance-related components of transportation costs were the only sources of variation in cost, this exercise would identify the aggregate impact of sharing frictions.

In reality, other factors can shift the real or shadow costs of potential shipments, even within the firm. For instance, the timing of when wastewater must be removed from one well pad might not align well with the timing of when it is needed at another. Transportation costs unrelated to distance might differ unobservably depending on the structure of trucking contracts, loading and unloading times, or opportunities for backhauls. In principle, there could be heterogeneity in treatment costs depending on wastewater composition and the particular characteristics of a completion (fracking fluid formulation, well construction, target formation). Thus, even if all well pads  $K$  and disposal facilities  $D$  were controlled by a single firm, the observed shipment plan would not necessarily minimize distance.

To assess the aggregate impacts of sharing frictions while controlling for unobserved sources of heterogeneity in costs, I compare observed shipments for reuse within the firm to two simple benchmarks. The first is random matching within each firm  $f$ , irrespective of distance. The second implements the minimum-distance allocation within each firm  $f$ . This allocation can be obtained by formulating and solving a simple optimal transport problem for  $f$ . Consider the flow of wastewater in month  $t$  between all well pads  $K_f$  and all disposal facilities  $D_f$  operated by  $f$ . For each well pad  $\kappa \in K_f$  and disposal facility  $\delta \in D_f$ , the data records the actual shipment volume  $\hat{\mu}_{\kappa\delta}$ . Suppose that all truckloads of wastewater are identical, and that all shipments within a month can be costlessly re-allocated, holding fixed the total disposal volume  $Q_\kappa = \sum_{\delta \in D} \hat{\mu}_{\kappa\delta}$  at well pad  $\kappa$  and the total volume received  $C_\delta = \sum_{\kappa \in K} \hat{\mu}_{\kappa\delta}$  at facility  $\delta$ . Then the distance-minimizing allocation  $\mu_f$  solves the following problem:

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

where  $d_{\kappa\delta}$  is the distance between  $\kappa$  and  $\delta$ .<sup>28</sup> 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 allocation within each firm would reduce industry trucking intensity by 2.3 miles per truckload relative to the status quo, or by 9% of the status quo level, suggesting that intrinsic costs of reuse not related to distance 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. I compute benchmark allocations with market-wide random matching and the market-wide minimum-distance allocation, this time ignoring firm boundaries. These results are summarized in the last two columns of Table 4. Compared to the allocation that optimizes within-firm shipments alone, the latter allocation reduces transportation intensity by 4.0 miles per truckload, or by a further 16 percentage points relative to the status quo. Thus, changing the composition of who participates in the sharing market and how shipments are allocated within the sharing market potentially has a larger effect on overall transportation efficiency than optimizing shipments

---

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

Table 4: Minimum-distance Benchmark

		Within-firm		Market-wide	
	Data	Random	Optimal	Random	Optimal
<i>Miles per truckload</i>					
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
<i>Sharing rate</i>					
	10.2	10.2	10.2	84.1	31.1

within firms, even though most reuse occurs within firms. Moreover, sharing rates are more than three times greater under this allocation than under the status quo. These findings suggest that sharing frictions are often large, even if they are heterogeneous or if gains from trade are frequently large enough to justify trade. It is also notable that the market as a whole realizes 95% of the maximum possible transportation cost savings relative to random matching. Intuitively, since firms are geographically segregated from one another, even relatively inefficient shipments within the firm (compared to the minimum-distance benchmark) coincide with an efficient outcome when viewed from the perspective of the market as a whole. This suggests that expanding the scope of the firm or otherwise reducing barriers to sharing could result in less efficient transportation overall to the extent that shipments decisions under integration are not driven by distance-related costs.

## 4 Model

In this section I develop a model of wastewater management that endogenizes insourcing and outsourcing in order to disentangle sharing frictions and other intrinsic costs of reuse. To do so, I adapt the matching framework of [Choo and Siow \(2006\)](#) and [Galichon and Salanie \(2022\)](#) to the setting of the Pennsylvania wastewater market.

Let  $K$  denote the finite set of well pads generating wastewater in month  $t$ , and  $D$  the finite set of facilities accepting wastewater for reuse (both well pads and operator-affiliated CTFs).<sup>29</sup> Each month, a unit mass of wastewater is shipped for disposal or reuse, while a unit mass of water (wastewater or freshwater) is used for completing new wells or, in the case of a CTF, processed for later use in completing new wells.

The firm is a coalition of managers who make decisions independently from one another. Each wellpad  $\kappa \in K$  is controlled by a manager  $m_\kappa$ , while each facility  $\delta \in D$  is controlled by a manager  $m_\delta$ .  $m_\kappa$  ships each truckload of wastewater  $i$  to the least cost destination,

<sup>29</sup>For the purpose of estimation, I exclude shipments to third party CTFs from the sample.

whether reuse at a facility  $\delta \in D$  or injection disposal (the outside option,  $\{0\}$ ). At the same time,  $m_\delta$  sources each truckload of completion water  $j$  from the least cost source, whether wastewater from a well pad  $\kappa \in K$  or freshwater (the outside option,  $\{0\}$ ). In the core of the matching game that I describe, this representation of the firm is without loss, since a coalition of managers cannot achieve lower costs than a collection of managers acting independently (I discuss limitations of the solution concept after presenting the model).

A mass of  $Q_\kappa$  truckloads of wastewater are generated at  $\kappa \in K$ , while a mass of  $C_\delta$  truckloads are needed at  $\delta \in D$ . Let  $r_{\kappa\delta}$  denote the systematic (shadow) cost of reusing wastewater from  $\kappa$  at  $\delta$ , 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  $\kappa$  and  $\delta$ . The form of  $r_{\kappa\delta}$  differs depending on whether  $\kappa$  and  $\delta$  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:

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

where  $r_{\kappa\delta}^{\mathcal{I}}$  is the systematic cost of reuse if  $\kappa$  and  $\delta$  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.  $r_{\kappa\delta}^{\mathcal{I}}$  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  $\kappa$  and  $\delta$  belong to the same firm.

In practice, the shadow cost of reusing wastewater from  $\kappa$  at  $\delta$  vary from truckload to truckload due to *latent cost heterogeneity*. The true shadow cost of supplying the  $j$ th truckload of demand at  $\delta$  with truckload  $i$  from  $\kappa$  is:

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

where  $\epsilon_{i\delta}$  and  $\eta_{\kappa j}$  represent latent private costs of managers  $m_\kappa$  and  $m_\delta$ , respectively, which vary from truckload to truckload. These costs capture any real or shadow costs that are relevant for manager's decisions regarding specific truckloads  $i$  and  $j$ , but distinct from the systematic costs affecting all shipments between  $\kappa$  and  $\delta$ . Latent costs are additively separable across  $i$  and  $j$ . This implies that  $m_\kappa$  is indifferent as to whether  $i$  is used as input  $j$  or  $j'$  at  $\delta$ , and that  $m_\delta$  is indifferent as to whether truckload  $i$  or  $i'$  is received from  $\kappa$ .<sup>30</sup>

---

<sup>30</sup>These implications are empirically reasonable: the composition of wastewater changes little on short time horizons at  $\kappa$ , and wastewater is unlikely to be segregated at  $\delta$ . Separability does not exclude “matching on unobservables” altogether:  $m_\kappa$ 's cost of shipping to  $\delta$  or might vary from  $i$  to  $i'$ ; and symmetrically,  $m_\delta$ 's



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

I assume that  $\epsilon_{i\delta}$  is drawn iid from  $P_K(X_{\kappa\delta}; \theta)$  and  $\eta_{\kappa j}$  is drawn iid from  $P_D(X_{\kappa\delta}; \theta)$ . These distributions may depend observables  $X_{\kappa\delta}$  and a parameter vector  $\theta$ , but not on unobservables, and (importantly) not on whether  $\kappa$  and  $\delta$  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.<sup>31</sup> I assume that  $P_K$  and  $P_D$  have full support and finite expectations.

The core of the matching game consists of the set of all *stable, feasible* matchings of wastewater from  $K$  to  $D$ . A matching  $\mu$  is *stable* if no manager  $m_\kappa$  would prefer to ship a truckload  $i$  allocated for reuse under  $\mu$  to a disposal well, no manager  $m_\delta$  would prefer to replace wastewater received under  $\mu$  with freshwater, and no  $m_\kappa$  and  $m_\delta$  would privately agree to match any  $i$  and  $j$  that are not matched under  $\mu$ . A matching is *feasible* if every truckload  $i$  is matched to some  $\delta \in D_0$ , and every  $j$  is allocated to a truckload of wastewater or freshwater from some  $\kappa \in K_0$ . Unlike the finite assignment model of [Shapley and Shubik \(1972\)](#), there are uncountably many truckloads of wastewater; the matching  $\mu$  can be represented as a probability mass function, where  $\mu_{\kappa\delta}$  represents the probability of observing a shipment of wastewater between  $\kappa$  and  $\delta$ . Using this convention, the set of feasible matches  $\mathcal{M}(\mathbf{Q}, \mathbf{C})$  consists of all matches  $\mu$  such that  $\sum_{\delta \in D_0} \mu_{\kappa\delta} = Q_\kappa$  for all  $\kappa \in K$  and  $\sum_{\kappa \in K_0} \mu_{\kappa\delta} = C_\delta$  for all  $\delta \in D$ , and  $\mu_{\kappa\delta} \geq 0$  for all  $\kappa\delta \in K \times D$ .

[Gretsky et al. \(1992\)](#) establish that the core of the matching game is equivalent to the set of Walrasian equilibria of an exchange economy. In a Walrasian equilibrium characterized

---

cost of reusing wastewater from  $\kappa$  might vary from  $j$  to  $j'$ , resulting in occasional matches.

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

by transfer matrix  $\tau$ , each manager  $m_\kappa$  solves a discrete choice problem for truckload  $i$ :

$$\min_{\delta \in D_0} r_{\kappa\delta}^K + \tau_{\kappa\delta} - \epsilon_{i\delta} \quad (2)$$

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

$$\min_{\kappa \in K_0} r_{\kappa\delta}^D - \tau_{\kappa\delta} - \eta_{\kappa j} \quad (3)$$

where  $r_{\kappa\delta}^D$  are the costs of reuse incurred by the receiver and  $\tau_{\kappa\delta}$  is the utility transfer received from  $m_\kappa$ . In equilibrium, the mass of transfers  $\mu_{\kappa\delta}$  between  $\kappa$  and  $\delta$  is equal to  $m_\kappa$ 's demand for disposal at  $\delta$  and to  $m_\delta$ 's supply of capacity to wastewater from  $\kappa$  when choices are made according to (2) and (3). Note that the systematic costs  $r_{\kappa\delta}^I$  are exactly equal to the sum  $r_{\kappa\delta}^K + r_{\kappa\delta}^D$  for all potential transactions. Furthermore, 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 \(2017\)](#).

[Galichon and Salanie \(2022\)](#) show there is a unique equilibrium matching  $\mu^*$  that minimizes a social cost function:

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

In this expression, the first term captures all systematic cost savings from reuse relative to injection disposal from  $\kappa$  and freshwater sourcing at  $\delta$ . The match entropy term  $\mathcal{E}$  captures the surplus contribution of savings on latent costs. Formally,  $\mathcal{E}$  is a function that depends on  $\mu$ , the distributions  $P_K$  and  $P_D$ , and the marginal probability masses  $\mathbf{Q}$  and  $\mathbf{C}$ :

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

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

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

and

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

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

Note that in contrast to the finite assignment model of [Shapley and Shubik \(1972\)](#), the transfer matrix  $\tau$  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  $r_{\kappa\delta}^K - r_{\kappa 0}^K - \tau_{\kappa\delta} = \frac{\partial G_\kappa^*}{\partial \mu_{\kappa\delta}}$  and  $r_{\kappa\delta}^D - r_{0\delta}^D + \tau_{\kappa\delta} = \frac{\partial H_\delta^*}{\partial \mu_{\kappa\delta}}$ , where  $G_\kappa^*$  and  $H_\delta^*$  are the individual choice entropies for  $m_\kappa$  and  $m_\delta$ , 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 (as recognized by [Shapley and Shubik \(1972\)](#)). An alternative possibility is to consider a model in which sharing decisions are mediated by Nash-in-Nash bargaining with transaction costs.<sup>32</sup> However, it is challenging to model network formation in the Nash-in-Nash context, significantly complicating counterfactual analysis.<sup>33</sup>

Another limitation of my approach is that I assume sharing frictions are exogenous and incurred in proportion to the number of truckloads sent. Linearity of the sharing frictions excludes the possibility that sharing frictions might be amortized over many similar truckloads (although, presumably, estimated frictions would be smaller in this case). In this way, the model differs from trade models that stipulate fixed costs of importing and exporting

---

<sup>32</sup>[Carlton \(2020\)](#) remarks on the absence of transaction costs in prior applications of Nash-in-Nash bargaining.

<sup>33</sup>[Ho and Lee \(2019\)](#) and [Ghili \(2022\)](#) develop models of network formation in Nash-in-Nash environments. One challenge in adapting these models to the wastewater management setting is the lack of a clear distinction between upstream and downstream firms, since these papers exploit institutional differences between upstream and downstream firms to simplify the strategy space (e.g., by assuming that one side of the market can pre-commit to a particular network).

(Antras and Chor, 2022). From a theoretical perspective, it would be natural to endogenize sharing frictions by modeling relationship dynamics (e.g., Chassang, 2010; Gibbons and Henderson, 2012) or favor trading (Samuelson and Stacchetti, 2017). Recent empirical work in the development literature has made progress on quantifying relationship dynamics (e.g., Macchiavello and Morjaria, 2015, 2021), but I do not pursue this approach here, primarily 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  $\Delta r$  denote the  $K \times D$  matrix of systematic cost savings (relative to the outside options) with typical element  $\Delta r_{\kappa\delta} = r_{\kappa\delta} - r_{\kappa 0}^K - r_{0\delta}^D$ . Galichon and Salanie (2022) establish that the systematic cost savings  $\Delta r$  are identified from the observed match  $\mu^*$  (i.e., from the shipments of wastewater) conditional on  $(P_K, P_D, \mathbf{Q}, \mathbf{C})$ .<sup>34</sup> However, the systematic costs of internal reuse  $r^\mathcal{I}$  and sharing frictions  $\phi$  are not separately identified without further restrictions. Thus, I assume that internal costs  $r^\mathcal{I}$  take the following form:

$$r_{\kappa\delta}^\mathcal{I} = g_{\kappa\delta}(X_{\kappa\delta}; \theta) + u_\kappa^\mathcal{I} + u_\delta^\mathcal{I} \quad (5)$$

where  $g_{\kappa\delta}(X_{\kappa\delta}; \theta)$  is a known function of observables  $X_{\kappa\delta}$  and a parameter vector  $\theta$ , and  $u_\kappa^\mathcal{I}$  and  $u_\delta^\mathcal{I}$  are additively separable unobservables. Under this assumption,  $r^\mathcal{I}$  is identified from observed shipments within each firm, while  $\phi$  is identified by shipments between firms. Separability implies that *within the firm*, any unobserved systematic costs of reusing wastewater at  $\delta$  are independent of the source of the wastewater  $\kappa \in K$ . Symmetrically, any unobserved systematic costs of sourcing wastewater from  $\kappa$  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  $\kappa$  and  $\delta$ . This excludes the possibility that some types of wastewater are unobservably better or worse suited for reuse at particular well pads and CTFs.

---

<sup>34</sup>This result is known for the special case of the logit from Choo and Siow (2006).

Under these assumption,  $\phi_{\kappa\delta}$  is identified whenever  $r_{\kappa\delta}^{\mathcal{I}}$  is identified (since  $\Delta r$  itself is identified, and  $\phi_{\kappa\delta} = \Delta r_{\kappa\delta} - r_{\kappa\delta}^{\mathcal{I}}$ ). In practice, however, I do not observe a sufficient amount of data to estimate  $\phi$  non-parametrically for every  $\kappa\delta$ . I therefore assume that sharing frictions  $\phi$  take the following functional form:

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

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

### 5.1.1 Welfare-relevance

I interpret sharing frictions as a type of transaction costs. While some authors have interpreted transaction costs as strictly real costs, this is not required: transaction costs might instead be viewed as “choice costs” that differ from “true costs” in the same way that “choice utility” differs from “true utility” when behavioral frictions at the consumer level are incorporated into structural models. Under this interpretation, the welfare-relevance of the sharing frictions is not identified: only the gross transaction cost is identified. Examples of welfare-relevant costs could include wages expended in search and negotiation, or quantifiable risks to future profits (e.g, production risks). Examples of welfare-irrelevant costs include shadow costs arising from managerial inattention, loss aversion, or excessive secrecy. Nothing in the data separately identifies welfare-relevant and welfare-irrelevant transaction costs.

## 5.2 Parameterization

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

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

where  $d_{\kappa\delta}$  is the linear component of transportation costs (e.g., distance- or driving time-related costs) between  $\kappa$  and  $\delta$  and  $x_{\kappa\delta}$  is a vector of match-specific observables. The coef-

ficient on  $d_{\kappa\delta}$  is one, so that all other parameters are interpretable in terms of the marginal cost of distance, which I assume is known. I focus on the case of linear (rather than log) over-the-road distance since trucking contracts in similar industries often have a per mile component, and since linear distance appears to deliver a good model fit.<sup>35</sup>  $x_{\kappa\delta}$  includes controls for differences between sending well pads and receiving facilities that could (possibly) shift the cost of reuse. Specifically,  $x_{\kappa\delta}$  includes proxies intended to capture potential differences in fracking fluid formulation and wastewater composition.

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

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

where  $z_{\kappa\delta}$  is a vector of match-specific observables.  $z_{\kappa\delta}$  includes covariates to disentangle potential sources of sharing frictions, including measures of wastewater characteristics, facility characteristics (e.g., well pads vs. CTFs), and firm-pair characteristics (such as differences in environmental compliance rates).

In practice, I cannot feasibly estimate  $\pi_b$  for every pair of firms. Instead, I keep only the 27 firm pairs with the largest bi-directional sharing volume (encompassing 50% of all sharing volume) and aggregate all remaining firms into 6 groups of observably similar firm pairs (for example, large firms in northwestern Pennsylvania sharing with small firms anywhere). I make this decision for two reasons. First, it is difficult to estimate a large number of parameters using the estimation procedure I describe in the next section. Second, an estimate for each firm pair fixed effect  $\pi_b$  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 pair of firms. The cost of this assumption is that any resulting violations of the assumption that  $\pi_b$  is constant within groups introduce bias into the estimates.

Finally, I assume that the latent cost distributions  $P_K(X_{\kappa\delta}; \theta)$  and  $P_D(X_{\kappa\delta}; \theta)$  are type 1 extreme value error distributions with mean zero and scale parameters  $\sigma_K$  and  $\sigma_D$ , respectively. I make this choice for simplicity and computational convenience: computation of the equilibrium is significantly more efficient in this case than with richer forms of heteroskedasticity. Moreover, this assumption implies the independence of irrelevant alternatives which (at the cost of realism) enables me to estimate the main parameters of interest without ad hoc assumptions regarding the shares of the outside options, as I discuss in the next section.

---

<sup>35</sup>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  $u_{\kappa}^I$  and  $u_{\delta}^I$ .

Alternatively, one could consider scale parameters that vary depending on the facility type (for example, for larger and smaller well pads in  $K$  and between larger and smaller well pads and CTFs in  $D$ ), or a similarly constructed nesting structure.

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

### 5.3 Estimation

For each month  $t$ , I observe the total shipment volume  $\hat{\mu}_{\kappa\delta}$  for each  $\kappa \in K$  and  $\delta \in D$ , as well  $\hat{\mu}_{\kappa 0}$  for each  $\kappa \in K$  and (in auxiliary data)  $\hat{\mu}_{0\delta}$  for each  $\delta \in D$ . Using 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.<sup>36</sup> 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  $\epsilon$  and  $\eta$  follow Gumbel distributions, a consistent (but inefficient) conditional maximum likelihood estimator for  $\theta_0$  is:

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

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

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

---

<sup>36</sup>When  $\delta$  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  $\delta$  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.



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

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

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

Importantly, (7) does not depend on  $\hat{\mu}_{\kappa 0}$  or  $\hat{\mu}_{0\delta}$ . Due to Sinkhorn’s Theorem,  $\tilde{u}$  and  $\tilde{v}$  satisfying (8) and (9) exist and are unique up to scale (see, e.g., [Idel, 2016](#)). In Appendix B, I describe how I solve this system for a given  $\theta$ . Together,  $\tilde{u}$  and  $\tilde{v}$  rationalize the observed marginal market shares conditional on  $\theta$  analogously to how additively separable terms representing unobserved heterogeneity rationalize observed market shares in the [Berry et al. \(1995\)](#) (or “BLP”) setting. Due to this similarity, the implementation of (7) is similar to the implementation of BLP-style demand models (and many of the computational suggestions of [Conlon and Gortmaker \(2020\)](#) are directly applicable).<sup>37</sup> 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. To avoid complication, I construct all counterfactuals under the assumption that there is no outside option on either side of the market.<sup>38</sup> 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 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).

---

<sup>37</sup>[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.

<sup>38</sup>This can be formalized with a timing assumption: first, managers choose between reuse and their outside options; then, in a second stage, the reuse market clears. The counterfactuals that I construct can be viewed as counterfactuals of the second stage of this game.

## 6 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 additional miles, or about \$6.14 per barrel (assuming trucking costs are \$5 per mile, a reasonable estimate for this industry). In comparison, the mean shipment distance for the same transactions was 43.6 miles (implying a cost of \$1.98 per barrel). This implies that, on average, bringing a transaction inside the firm has the same effect on private costs as a threefold decrease in shipment distance. However, the estimated sharing frictions are highly heterogeneous. Among the 27 firm-pairs for which I estimate a firm-pair specific sharing friction fixed effect  $\pi_b$ , the expected friction across all shipments (averaging over observable sharing frictions  $z'_{\kappa\delta}\alpha$ ) was 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 therefore reflects the selected composition of the sharing market under the status quo: if the distance-minimizing allocation were implemented for the market as a whole, realized sharing frictions would be 20% greater on average.

#### 6.1.1 Sources of frictions

Understanding the source of sharing frictions is crucial for drawing policy implications from the model. The merger evidence, variation in sharing across firm pairs, and the estimated friction-shifting  $\alpha$  parameters each give insight into the sources of sharing frictions. I discuss all three perspectives in this section.

**Wastewater quality** Certain types of wastewater may be less well suited to reuse than others, either because they require more treatment or because they create risks to well productivity. Roughly, there are two types of fracking fluids: “gels” and “slickwaters.” The distinction between the two is not always sharp, but certain constituents commonly found in slickwater formulations can inhibit the formation of gels in gel-based formulations (Montgomery, 2013; Walsh, 2013). Using the FracFocus data, I construct an indicator for whether a given well was likely to have been fracked with a gel or a slickwater-based fracking fluid,

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 $\alpha$			
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 $\kappa \rightarrow$ well pad	-	-	-
rival $\times$ small $\kappa \rightarrow$ well pad	9.4	0.152	0.43
rival $\times$ small $\kappa \rightarrow$ CTF	25.1	0.260	1.14
rival $\times$ large $\kappa \rightarrow$ CTF	24.6	0.044	1.12
Within-firm cost shifters $\beta$			
gel $\rightarrow$ slickwater	6.8	0.091	0.31
slickwater $\rightarrow$ gel	-8.6	0.046	-0.39
small $\kappa \rightarrow$ 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).

which I infer from the presence of certain key chemicals.<sup>39</sup> I construct a similar measure for whether wastewater received at facility  $\delta$  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  $\delta$  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. Thus, sharing frictions appear to be greater when risks to well productivity are greater, perhaps because firms lack perfect information about the costs and benefits of accepting particular truckloads of wastewater, or 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 after an agreement to share has been reached.

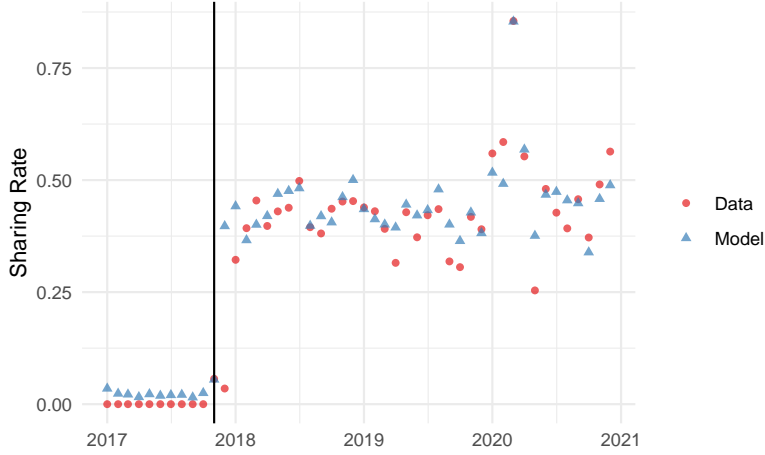
**Inter-operator environmental liability** When a firm shares wastewater with another firm, it assumes some non-contractible risk that the receiving firm will mishandle the wastewater in some way that exposes the sending firm to civil liability, regulatory fines, or reputational damage. The most serious 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  $\delta$  has a worse compliance record than the firm affiliated with well pad  $\kappa$  (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.<sup>40</sup> 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

<sup>39</sup>Guar in the case of gels, and acrylamide in the case of slickwaters (Montgomery, 2013).

<sup>40</sup>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  $\pi_b$  before the merger, so  $\pi_b$  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

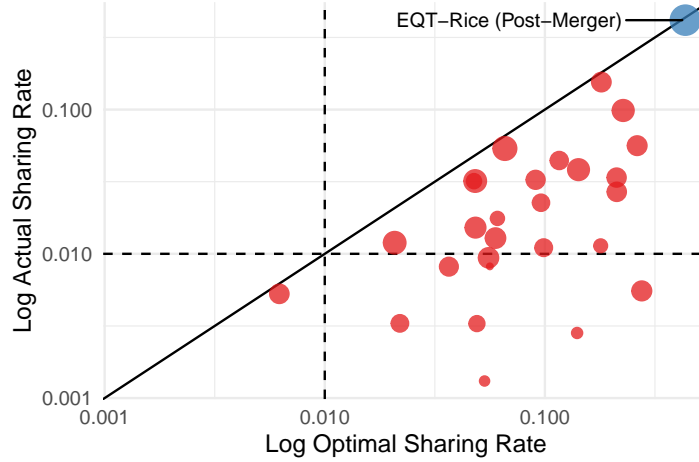


suggests that integration can eliminate sharing frictions. Thus, whatever sharing frictions might consist of, it appears that they can be resolved through integration (consistent with the central but often-disputed claim of transaction cost economics, see e.g., [Demsetz \(1988\)](#)).

**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  $\pi_b$ , 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.<sup>41</sup> The figure shows clearly that for most firm pairs, actual sharing volumes are far below the optimum. At the median, the actual sharing volume was 78% lower than the optimal sharing volume, and only 3 pairs of firms shared more than 75% of the optimal volume. In comparison, sharing volume between EQT and Rice-affiliated facilities was 96% of the optimal sharing volume after the EQT-Rice merger. Thus, by comparison to the merger evidence, it is seemingly 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 from agreements to coordinate 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](#)).

<sup>41</sup>For comparison, Figure 12 plots the actual sharing rate against the fitted sharing rate.

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



**Principal-agent frictions** Loosely, wastewater management for a typical firm can be divided into four different tasks. On the disposal side, managing disposal for mature wells (many locations, small volumes) is a qualitatively different task from managing disposal for recently drilled wells (few locations, large volumes); on the receiving side, managing wastewater sourcing on a well pad before a completion (inelastic demand, tight delivery windows) is qualitatively different from managing receipts at a CTF. These tasks may be delegated to different agents with different incentives. The estimated sharing frictions should be inversely related to the strength of agents' incentives to execute the lowest cost transaction. This perspective suggests a few testable hypotheses. For managers fracking new wells, water sourcing costs are a highly visible capital expenditure; in comparison, managers at operator-affiliated CTFs may have little incentive to source wastewater from rival firms even when doing so would reduce the firms' total cost. 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.<sup>42</sup> 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 costs; 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.

<sup>42</sup>“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).

## 6.2 Substitution between insourcing and outsourcing

To further contextualize the sharing frictions, I analyze the extent to which insourcing is driven by sharing frictions as opposed to transaction-specific differences in the intrinsic cost of reuse. Since each firm’s operations are geographically concentrated (depicted in Figure 1), the intrinsic cost of reuse within the firm is often lower than the intrinsic cost of feasible sharing transactions. Thus, qualitatively large sharing frictions might contribute relatively little to observed insourcing rates in comparison to within firm cost efficiencies. However, I find that the effects of sharing frictions on insourcing rates are modestly large. Aggregating across firms, only about 55% of observed insourcing volume is attributable to intrinsic costs differences as opposed to sharing frictions.

This finding reflects the interplay between two forces in equilibrium. Sharing frictions directly raise the cost of potential sharing transactions, and indirectly mediate price levels in the sharing market. When sharing frictions are removed, increased competition in the sharing market can dampen firms’ incentives to substitute towards the sharing markets. Figure 13 presents the distribution of insourcing rates for the twenty largest firms in the data, the fitted model, and in two counterfactuals. Among these firms, the average firm reuses 70% of wastewater internally under the status quo. The purple series shows that in a world without sharing frictions, 29% of wastewater would be reused internally on average. But if only the frictions directly affecting potential transactions to which firm  $f$  is a counterparty were eliminated (i.e.,  $f$ ’s *degree one* frictions), leaving frictions between all other pairs of firms unchanged,  $f$ ’s insourcing rate would fall to 14% on average (shown in green), implying that price effects mute the direct effects of the removal of sharing frictions.

## 6.3 Additional results

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

### 6.3.1 Non-transportation costs within the firm

Table 5 also reports point estimates for within-firm cost shifters  $x_{\kappa\delta}$ . In general,  $\kappa$  and  $\delta$  specific characteristics are not separately identified from the fixed effects in (5), which have no effect on the conditional match. Thus, I only consider a small number of covariates that interact characteristics of  $\kappa$  with characteristics of  $\delta$ . 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 findings 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.

### 6.3.2 Latent costs

The point estimate for  $\sigma_K + \sigma_D$  indicates how much of the total match surplus is created by matching on latent costs rather than systematic costs. My preferred estimation approach does not separately identify  $\sigma_K$  and  $\sigma_D$  (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  $\mu$  and  $\mu'$ , 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  $\sigma_K + \sigma_D \rightarrow 0$ . 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  $\sigma_K + \sigma_D \rightarrow \infty$ , 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 observables).

### 6.3.3 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 14 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, demonstrating good model fit overall. The final row shows the time series of the sharing rate and aggregate mean shipment distance. This provides visual evidence that the aggregate model fit does not mask a poor model fit in specific time periods (e.g., during the spring of 2020).

### 6.3.4 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  $\pi$ , 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  $\kappa$  and  $\delta$  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  $\pi_b$ . 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  $\epsilon$  and  $\eta$ ). 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. The green line indicates the mean shipment distance if the ownership of each well pad  $\kappa$  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. Thus, the main finding would reverse if firms were geographically dispersed or if non-transportation costs were relatively less important. Under the status quo, eliminating sharing frictions increases matching on transportation and non-transportation cost simultaneously. Because firms are geographically clustered, extramarginal matches increasingly involve non-transportation synergies at greater and greater distances.

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

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.<sup>43</sup> 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.<sup>44</sup>

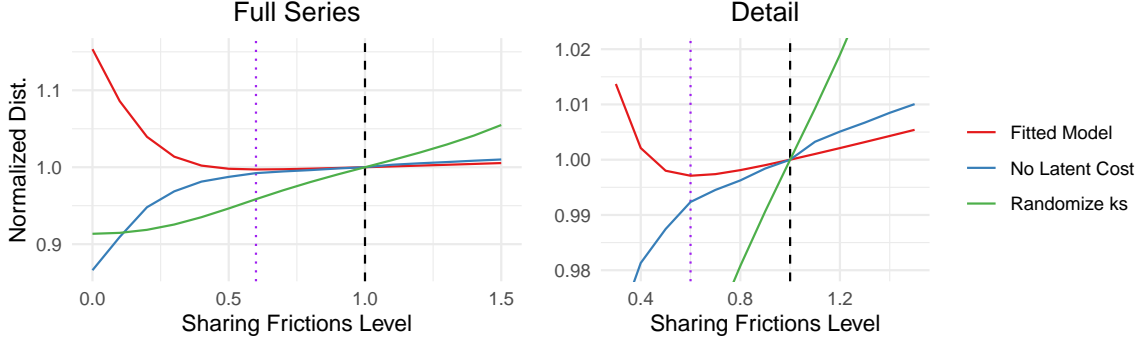
These results have a few immediate policy implications. Most immediately, market design improvements that aim to make sharing easier may not lead to large reductions in trucking-

---

<sup>43</sup>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.

<sup>44</sup>I do not predict how Rice and EQT's drilling activity would have evolved absent the merger. Nevertheless, this finding demonstrates that integration can entail many changes in economic activity aside from the elimination of sharing frictions, which fall outside the scope of the model. Thus, the elimination of sharing frictions should not be interpreted as a "merger-to-monopoly."

Figure 6: Shipment Distance as a Function of Sharing Frictions



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 social planner's 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  $\phi$ . This means that for each  $\kappa\delta$ , the social planner should internalize  $s\phi_{\kappa\delta}$  but not  $(1 - s)\phi_{\kappa\delta}$ .<sup>45</sup> In this section, I assume that the social planner observes the estimated cost parameters, the marginal external cost of trucking  $\gamma$ , and the welfare relevance parameter  $s$ . Holding fixed the overall level of wastewater generation and

<sup>45</sup>Of course,  $s$  could differ across  $\kappa\delta$ . I abstract from this possibility.

reuse, the Pigouvian match  $\mu_s^*$  in state  $s$  is the solution to:

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

where  $\Gamma(\mu)$  represents the external cost of the next truckload of wastewater shipped under shipment plan  $\mu$ :

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

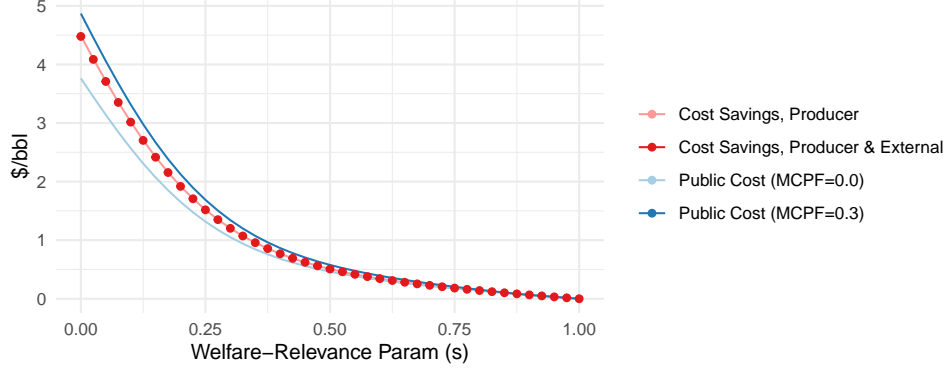
and  $C_s(\mu)$  represents the welfare-relevant component of the private cost of shipping the next truckload of wastewater under  $\mu$  in state  $s$ . In Appendix C, I show that the optimal match  $\mu_s^*$  is implemented by Pigouvian tax on truck-miles where the optimal tax (or subsidy) on shipments between  $\kappa$  and  $\delta$  in state  $s$  is:

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

If sharing frictions are fully welfare-relevant ( $s = 1$ ), the social planner implements a uniform tax equal to the marginal external cost of trucking  $\gamma$ . 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 greager 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  $\lambda$ . 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

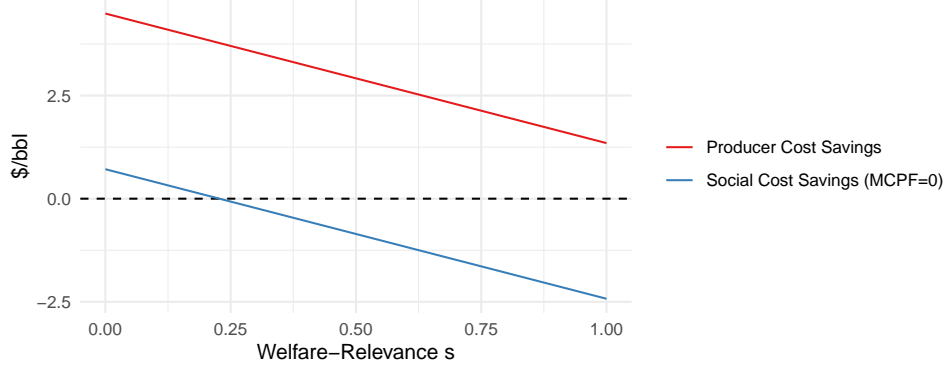


## 8.2 Pigouvian taxation under ambiguity

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  $tax_{\kappa\delta}^{(1)}$  relative to uniform tax  $tax_{\kappa\delta}^{(0)}$ , 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  $\phi$  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, such as the establishment of a clearinghouse or digital exchange. Nevertheless, the benefits of any such interventions are likely modest at best, and could be counterproductive, due to the geographic configuration of firms and the misalignment between private and social costs. 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 stronger assumptions) it would be straightforward to quantify the impact of sharing frictions on injection well disposal rates and freshwater withdrawals. Freshwater withdrawals in particular are an important issue in shale basins in the western United States, and the framework in this paper can be used to model regulation in this context. Second, it is natural

to consider whether quantity regulations have any advantages over corrective taxation in the presence of sharing frictions, although the more general 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.
- Covert (2015) “Experiential and Social Learning in Firms: The Case of Hydraulic Fracturing in the Bakken Shale,” *Working Paper*.
- Covert and Kellogg (2023) “Environmental Consequences of Hydrocarbon Infrastructure Policy,” *Working Paper*.
- Dahlman (1979) “The Problem of Externality,” *The Journal of Law and Economics*, 22 (1), 141–162.
- Decker, Lieb, McCann, and Stephens (2013) “Unique equilibria and substitution effects in a stochastic model of the marriage market,” *Journal of Economic Theory*, 148, 773–792.
- Demsetz (1988) “The Theory of the Firm Revisited,” *Journal of Law, Economics, & Organization*, 4 (1), 141–161.
- EPA (2016) “Hydraulic Fracturing for Oil and Gas: Impacts from the Hydraulic Fracturing Water Cycle on Drinking Water Resources in the United States (Final Report).”
- Fowlie (2009) “Incomplete Environmental Regulation, Imperfect Competition, and Emissions Leakage,” *American Economic Journal: Economic Policy*, 1 (2), 71–112.
- Galichon and Salanie (2022) “Cupid’s Invisible Hand: Social Surplus and Identification in Matching Models,” *Review of Economic Studies*.
- Ghili (2022) “Network Formation and Bargaining in Vertical Markets: The Case of Narrow Networks in Health Insurance,” *Marketing Science*, 41 (3), 501–527.
- Gibbons and Henderson (2012) “Relational Contracts and Organizational Capabilities,” *Organization Science*, 23 (5), 1350–1364.
- Goldin and Reck (2018) “Rationalizations and Mistakes: Optimal Policy with Normative Ambiguity,” *AEA Papers and Proceedings*, 108, 98–102.
- Graham (2011) “Econometric Methods for the Analysis of Assignment Problems in the Presence of Complementarity and Social Spillovers,” *Handbook of Social Economics*, 1B.

- (2013) “Comparative static and computational methods for an empirical one-to-one transferable utility matching model,” *Mimeo*.
- Graham, Irving, Tang et al. (2015) “Increased traffic accident rates associated with shale gas drilling in Pennsylvania,” *Accident Analysis and Prevention*.
- Gretsky, Ostroy, and Zame (1992) “The Nonatomic Assignment Model,” *Economic Theory*, 2 (1), 103–127.
- Grossman and Hart (1986) “The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration,” *Journal of Political Economy*, 94 (1986), 691–719.
- Groundwater Protection Council (2019) “Produced Water Report: Regulations, Current Practices, and Research Needs.”
- Hahn and Stavins (2011) “The Effect of Allowance Allocations on Cap-and-Trade System Performance,” *Journal of Law and Economics*, 54 (4), S267–S294.
- Handel and Kolstad (2015) “Health Insurance for "Humans": Information Frictions, Plan Choice, and Consumer Welfare,” *American Economic Review*, 105 (8), 2449–2500.
- Handel and Schwartzstein (2018) “Frictions or Mental Gaps: What’s Behind the Information We (Don’t) Use and When Do We Care?” *Journal of Economic Perspectives*, 32 (1), 155–78.
- Harris and Nguyen (2023) “Long-Term Relationships and the Spot Market: Evidence from US Trucking,” *Working Paper*.
- Hausman and Kellogg (2015) “Welfare and Distributional Implications of Shale Gas,” *Brookings Papers on Economic Activity*, 71–125.
- Head and Meyer (2014) “Gravity Equations: Workhorse, Toolkit, and Cookbook,” *Handbook of International Economics*, 131–195.
- Heo, Adams, and Gao (2016) “Reduced-form modeling of public health impacts of inorganic PM2.5 and precursor emissions,” *Atmospheric Environment*.
- Ho and Lee (2019) “Equilibrium Provider Networks: Bargaining and Exclusion in Health Care Markets,” *American Economic Review*, 109 (2), 473–522.
- Hodgson (2022) “Trade-ins and Transaction Costs in the Market for Used Business Jets.”
- Hubbard (2001) “Contractual Form and Market Thickness in Trucking,” *The RAND Journal of Economics*, 32 (2), 369–386.

- Idel (2016) “A review of matrix scaling and Sinkhorn’s normal form for matrices and positive maps,” *arXiv: Rings and Algebras*, arXiv:1609.06349.
- Kellogg (2011) “Learning by Drilling: Interfirm Learning and Relationship Persistence in the Texas Oilpatch,” *The Quarterly Journal of Economics*, 126 (4), 1961–2004.
- Kondash, Lauer, and Vengosh (2018) “The intensification of the water footprint of hydraulic fracturing,” *Science Advances*.
- Leslie (2018) “Tax induced emissions? Estimating short-run emission impacts from carbon taxation under different market structures,” *Journal of Public Economics*.
- Luxen and Vetter (2011) “Real-time routing with OpenStreetMap data,” *Proceedings of the 19th ACM SIGSPATIAL international conference on advances in geographic information systems*.
- Macchiavello and Morjaria (2015) “The Value of Relationships: Evidence from a Supply Shock to Kenyan Rose Exports,” *American Economic Review*, 105 (9), 2911–45.
- (2021) “Competition and Relational Contracts in the Rwanda Coffee Chain,” *The Quarterly Journal of Economics*, 136 (2), 1089–1143.
- MacKay (2022) “Contract Duration and the Costs of Market Transactions,” *American Economic Journal: Microeconomics*, 14 (3).
- Maloney, Baruch-Mordo, Patterson et al. (2017) “Unconventional oil and gas spills: Materials, volumes, and risks to surface waters in four states of the U.S.,” *Science of the Total Environment*.
- Mansur (2007) “Do Oligopolists Pollute Less? Evidence from a Restructured Electricity Market,” *The Journal of Industrial Economics*, 55 (4), 661–689.
- Masten, Meehan, and Snyder (1991) “The Costs of Organization,” *Journal of Law, Economics, & Organization*, 7 (1), 937– 976.
- McCurdy (2011) “Underground Injection Wells For Produced Water Disposal.”
- Menefee and Ellis (2020) “Wastewater management strategies for sustained shale gas production,” *Environmental Research Letters*.
- Montgomery (2013) “Fracturing Fluid Components,” *Effective and Sustainable Hydraulic Fracturing*.
- Muehlenbachs, Staubli, and Chu (2021) “The Accident Externality from Trucking: Evidence from Shale Gas Development,” *Regional Science and Urban Economics*.
- Newell and Raimi (2014) “Implications of Shale Gas Development for Climate Change,” *Environmental Science and Technology*, 48 (15), 8360–8368.

- Preonas, Louis (2023) “Market Power in Coal Shipping and Implications for U.S. Climate Policy,” *Mimeo*.
- Samuelson and Stacchetti (2017) “Even up: Maintaining relationships,” *Journal of Economic Theory*, 169, 170–217.
- Shapley and Shubik (1972) “The assignment game I: The core,” *International Journal of Game Theory*.
- Small, Stern, Bomberg et al. (2014) “Risks and Risk Governance in Unconventional Shale Gas Development,” *Environmental Science and Technology*.
- Stavins (1995) “Transaction Costs and Tradeable Permits,” *Journal of Environmental Economics and Management*.
- Sweeney and Covert (2022) “Secrecy Rules and Exploratory Investment: Theory and Evidence from the Shale Boom.”
- Vreugdenhil (2022) “Booms, Busts, and Mismatch in Capital Markets: Evidence from the Offshore Oil and Gas Industry.”
- Walsh (2013) “Water Management for Hydraulic Fracturing in Unconventional Resources:Part 1,” *Journal of Petroleum Technology*.
- Williamson (1971) “The Vertical Integration of Production: Market Failure Considerations,” *American Economic Review*, 61 (1971), 112–123.
- Wilson (2022) “Moving to economic opportunity: the migration response to the fracking boom,” *Journal of Human Resources*.
- Wunz Associates (2014) “Summary of Shale Gas Wastewater Treatment and Disposal In Pennsylvania,” *Report*.
- Xu and Xu (2020) “Fraccidents: The impact of fracking on road traffic deaths,” *Journal of Environmental Economics and Management*.

# Appendix

## A Data Preparation

The main dataset consists of Oil and Gas Well Waste Reports collected from the Pennsylvania Department of Environmental Protection web site. For the main analysis, I consider waste reports for all unconventional wells and for all production periods between January 2017 and December 2020. This choice of analysis period reflects the fact that the waste reporting format was modified in 2017 to consistently indicate the location of reuse. I choose to retain data from the Covid pandemic period. Although drilling rates in general fell during this period, the demand for disposal did not, and overall reuse rates remained relatively stable, as evidenced by Figure 10.

Operators are required to report disposal method for various waste products, including solids such as drill cuttings and shredded containment liners. I rely on the classifications from [Wunz Associates \(2014\)](#) as well as internet research on the functions performed at different waste facilities (e.g., landfills vs. injection wells) to identify presumably reusable wastewater. This procedure is inevitably imperfect. Reporting errors are possible, and not all liquid waste in fact represents reusable wastewater. In particular, sludges produced as a byproduct of the treatment process are in some cases disposed via injection well. Although these volumes are presumably small in comparison to the volumes of reusable wastewater, my preferred approach to estimation avoids relying on injection well rates to avoid data contamination, as discussed in the main text.

As described in the main text, the waste reports do not report the dates or quantities associated with specific transfer events, but rather the aggregate quantities of different types of waste transferred from a given well to a given disposal location during a specified month. Wastewater intended for reuse can be transferred either to a CTF prior to reuse or directly to another well pad for reuse. These cases appear differently in the data. In the former case, it is not possible to identify the ultimate location of reuse. However, whether the treatment facility is operated by the reporting firm or by a third party can be inferred from the reported permit information and facility names (although in some cases this requires consulting separate DEP resources). In the latter case, if the destination well pad is located in Pennsylvania, a numeric identifier associated with the destination well pad is also provided. I use this numeric identifier to determine whether a given amount of wastewater was transferred for internal or external reuse. In particular, I classify reuse locations as internal or external depending on whether the reporting firm is currently listed as an operator for any well at the destination well pad (in a separate DEP data source). If the destination



well pad is located outside of Pennsylvania (primarily in West Virginia), no such identifier is provided, and I do not attempt to infer the ownership of the destination well pad. I identify firms by their DEP OGO Number (where OGO is an acronym for “Oil and Gas Operator”). I rely on press releases and changes in the data over time to account for changes in ownership over time (the Rice-EQT merger was the most significant but not the sole merger during the sample period). It is rare for multiple operators to be associated with the same well pad, but when this is the case I treat the well pad as “internal” for both parties

Typically several wells are located at a single well pad, which encompasses common infrastructure such as access roads and storage tanks. Technically operators are required to report waste information on a well-by-well basis, but 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.<sup>46</sup> 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  $\delta$  is a CTF, I take a volume-weighted average of the indicators for well pads operated by the operator affiliated with  $\delta$ .

**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

---

<sup>46</sup>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,  $\hat{\mu}_{11}, \dots, \hat{\mu}_{KD}$  where  $\sum_{\delta \in D} \hat{\mu}_{\kappa\delta} > 0$  for all  $\kappa$  and  $\sum_{\kappa \in K} \hat{\mu}_{\kappa\delta} > 0$  for all  $\delta$ . We do not observe the shares of the outside options,  $\hat{\mu}_{10}, \dots, \hat{\mu}_{K0}$  or  $\hat{\mu}_{01}, \dots, \hat{\mu}_{0D}$ . The conditional likelihood of observing the data is:

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

Galichon and Salanie (2022) Theorem 4 implies that:

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

where:

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

and:

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

For the logit, we can show that:

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

and hence:

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

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

$$\begin{aligned}\mu_{\kappa 0} + \sum_{\delta \in D} \mu_{\kappa\delta} &= Q_{\kappa} \quad \forall \kappa \in K \\ \mu_{0\delta} + \sum_{\kappa \in K} \mu_{\kappa\delta} &= C_{\delta} \quad \forall \delta \in D\end{aligned}$$

where the marginals  $Q_{\kappa}$  and  $C_{\delta}$  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  $Q_{\kappa}$  and  $C_{\delta}$  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:

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

Expanding terms and re-arranging gives:

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

Now consider the system of equations:

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

Clearly,  $\tilde{u}_\kappa = (\sigma_\kappa + \sigma_\delta)^{-1} \psi_\kappa - (\sigma_\kappa + \sigma_\delta)^{-1} \sigma_\kappa \{\log c_n + \log \mu_{\kappa 0}\}$  for all  $\kappa$  and  $\tilde{v}_\delta = (\sigma_\kappa + \sigma_\delta)^{-1} \psi_\delta - (\sigma_\kappa + \sigma_\delta)^{-1} \sigma_\delta \{\log c_n + \log \mu_{0\delta}\}$  for all  $\delta$  is a solution to this system. Moreover it is easy to see that  $\tilde{u}_\kappa = \alpha + (\sigma_\kappa + \sigma_\delta)^{-1} \psi_\kappa - (\sigma_\kappa + \sigma_\delta)^{-1} \sigma_\kappa \{\log c_n + \log \mu_{\kappa 0}\}$  for all  $\kappa$  and  $\tilde{v}_\delta = -\alpha + (\sigma_\kappa + \sigma_\delta)^{-1} \psi_\delta - (\sigma_\kappa + \sigma_\delta)^{-1} \sigma_\delta \{\log c_n + \log \mu_{0\delta}\}$  all  $\delta$  is also a solution for any  $\alpha \in \mathbb{R}$ . Indeed, provided that  $\sum_{\delta} \hat{\mu}_{\kappa\delta} > 0$  for all  $\kappa$  and  $\sum_{\kappa} \hat{\mu}_{\kappa\delta} > 0$  for all  $\delta$ ,  $\tilde{u}$  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.<sup>47</sup>

Finally, observe that:

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

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

---

<sup>47</sup>To apply the theorem, note that  $\exp \left\{ -(\sigma_\kappa + \sigma_\delta)^{-1} (d_{\kappa\delta} + x'_{\kappa\delta}\beta + z'_{\kappa\delta}\alpha + \pi_b) \right\} > 0$  for all  $\kappa$  and  $\delta$ . 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,  $\kappa$  and  $\delta$ -specific covariates have no effect on the equilibrium match under the maintained assumptions.<sup>48</sup> Therefore, it is only necessary to include covariates for economically relevant interactions between  $\kappa$  and  $\delta$ -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 comparison 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:

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

Even more compactly, this is:

$$\begin{aligned} U \circ ZV &= \tilde{Q} \\ V \circ Z'U &= \tilde{C} \end{aligned}$$

where  $U = (U_1, \dots, U_K)$  and  $V = (V_1, \dots, V_D)$ , 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:

$$\begin{aligned} U^{(s+1)} &= \tilde{Q} \circ (ZV^{(s)})^{-1} \\ V^{(s+1)} &= \tilde{C} \circ (ZU^{(s+1)})^{-1} \end{aligned}$$

---

<sup>48</sup>To illustrate, suppose reuse at  $\delta$  incurs an additional cost of  $c_{\delta}$  per truckload. If this cost is the same for truckloads from all origins  $\kappa$ , then the magnitude of  $c_{\delta}$  has no effect on the relative probability that trucks from  $\kappa$  or  $\kappa'$  are matched to  $\delta$ .

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  $\phi$ .) 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  $\mu_s^*$  satisfies:

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

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

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

Inspection of (13) shows that the Pigouvian tax on truck-miles between  $\kappa$  and  $\delta$  is:

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

**Robustness of uniform tax within  $\mathcal{S}$**  Consider  $\mathcal{S} = \{\alpha : \alpha \in [0, 1]\}$ , where the state  $s = \alpha$  implies that a fraction  $\alpha$  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  $C_{\alpha}$  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  $\mu$  by the concavity of  $\mathcal{E}$ , while  $f(\mu, \cdot)$  is linear in  $\alpha$  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  $\alpha$ , it follows that  $f(\mu, \alpha)$  obtains a minimum at  $\alpha = 0$ .

## D Additional Tables and Figures

Figure 9: Centralized Treatment Facility Locations

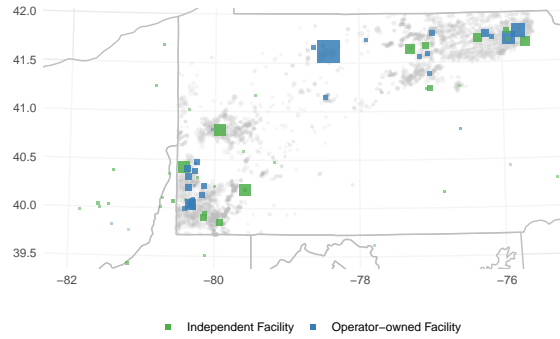
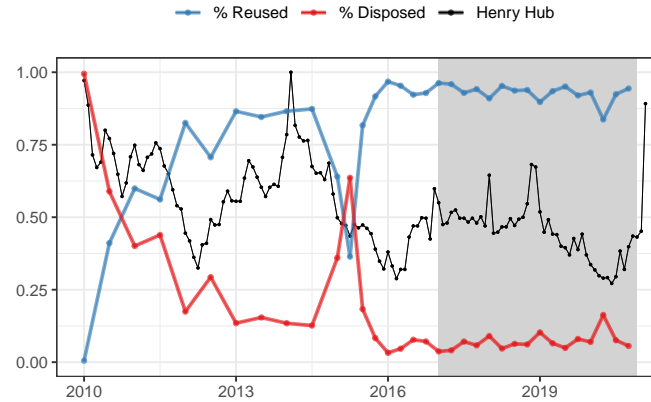


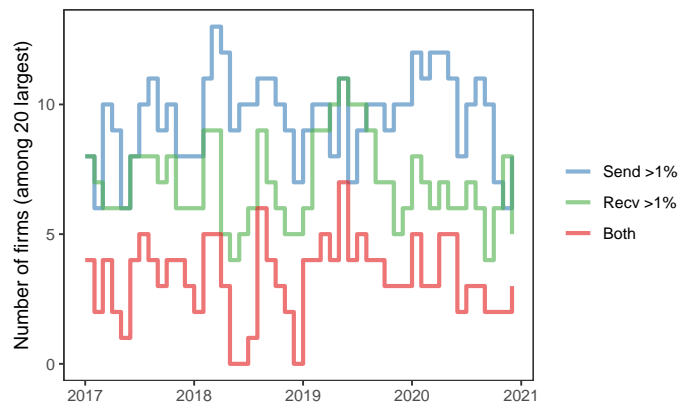


Figure 10: Wastewater Reuse Over Time in PA



*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

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

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

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

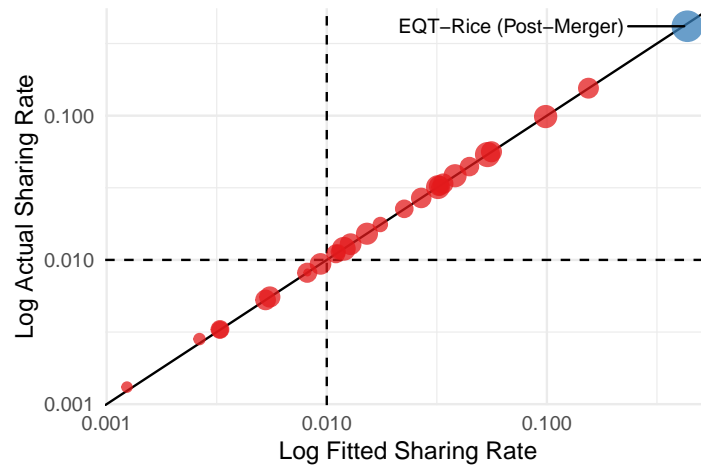


Figure 13: Insourcing Rates for 20 Largest Firms

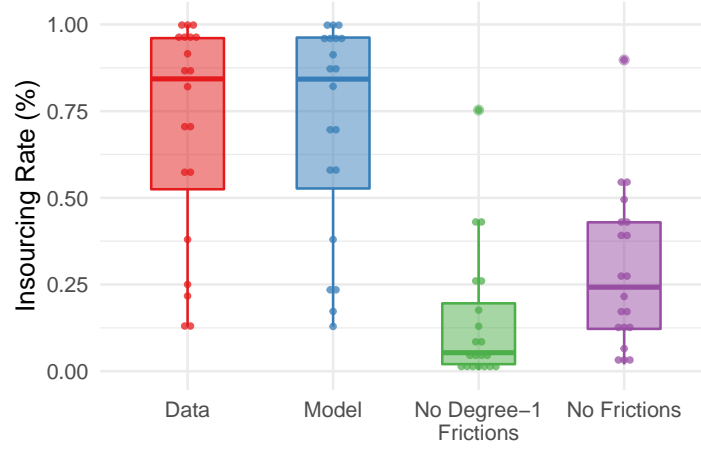
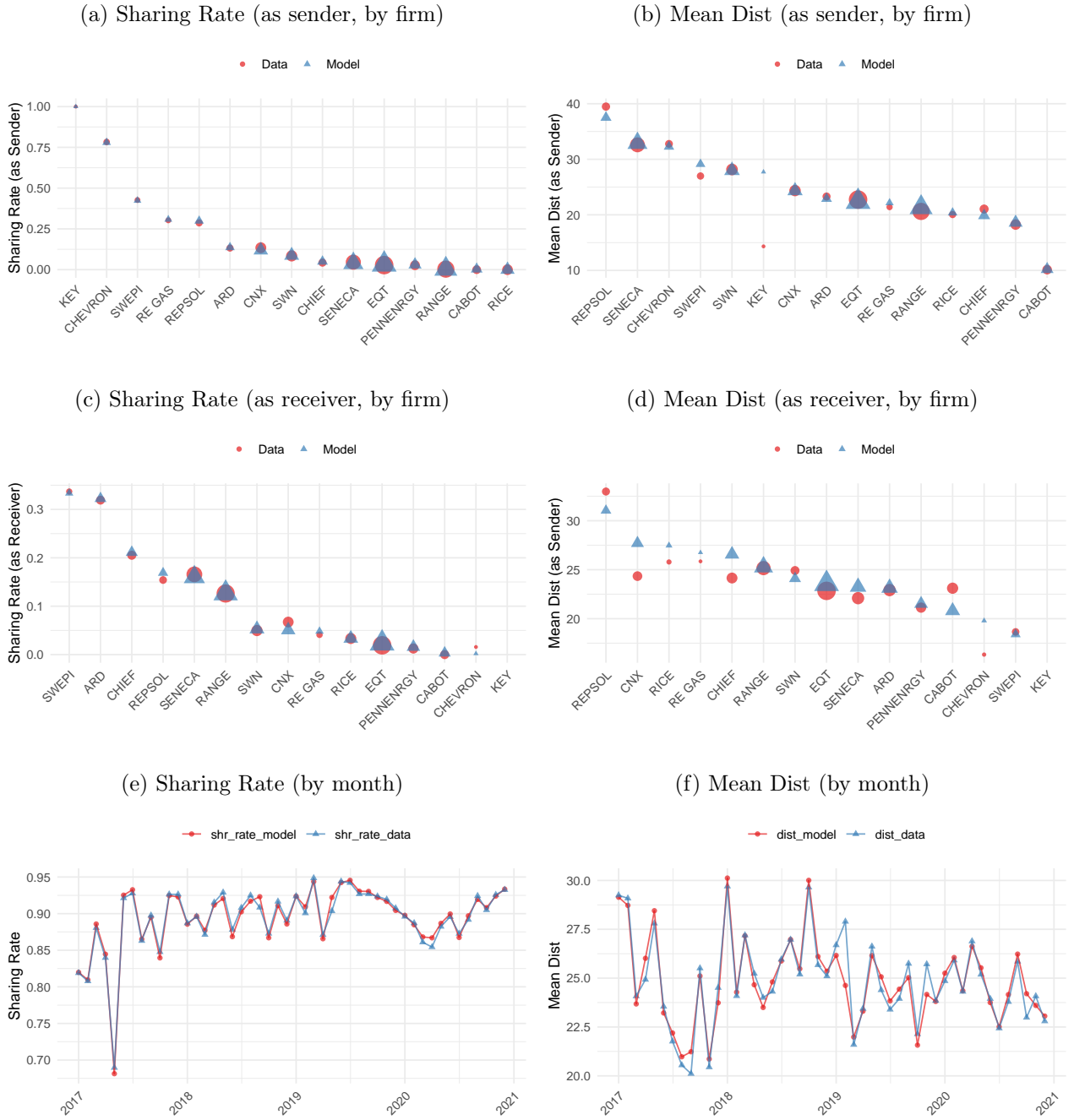


Table 7: Model Fit & Key Counterfactuals

	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		
$\sigma_K + \sigma_D \rightarrow 0$	21.61	9.72
$\sigma_K + \sigma_D \rightarrow \infty$	146.99	84.37

*Notes:* The  $\phi$  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  $\sigma_K + \sigma_D \rightarrow 0$  counterfactual is constructed by solving a linear program where the cost function is the estimated systematic cost  $r_{\kappa\delta}$ . The  $\sigma_K + \sigma_D \rightarrow \infty$  is constructed by assuming that each truckload  $i$  is sent to each facility  $\delta$  with equal probability, and likewise that each delivery slot  $j$  is allocated to a truckload from  $\kappa$  with equal probability.

Figure 14: Model Fit Diagnostics



*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 $\phi$	127.6	135.1	132.3	131.7	4.1	6.4	291.7	250.2
$\alpha$ 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
$\beta$ 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  $\kappa$  and  $\delta$  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  $\kappa$  and  $\delta$  is less than 30 miles. Likewise, when distance is included non-linearly in  $z_{\kappa\delta}$  in (4) and (8), it is represented with indicators for 30-mile increments. The log-likelihood is the negative of the objective function in equation (7) in the body of the text. The model fit statistics report the median deviation between the observed and fitted expected distance and sharing probability across the indicated category (months, or firm-months).

Figure 15: Firm Locations and Transportation Efficiency (Illustration)

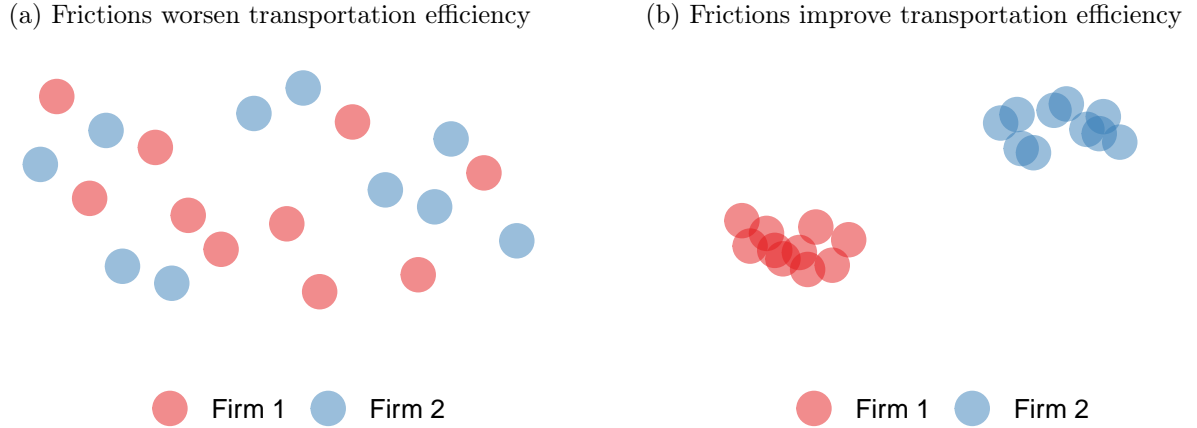


Table 9: Counterfactual Shipment Distances & Sharing Rate Within EQT-Rice

	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.