Make, Buy, or Share: Understanding Produced Water Reuse in the Marcellus Shale

Matthew O'Keefe*

August 21, 2025

Abstract

Wastewater reuse in the shale gas industry mitigates environmental harms while reducing firms' private costs. Most reuse occurs within the firm boundary, but rival operators often exchange (or "share") wastewater prior to reuse. I quantify private cost savings from reuse and sharing among Pennsylvania producers during the shale boom using an empirical model of transferrable utility matching. Model estimates imply that private cost savings are substantial but modest relevate to producer revenue, ruling out large rebound effects. I also find that sharing is subject to large transaction costs, implying that equilibrium consumption and disposal may be inefficiently high.

Keywords: oil and gas extraction; produced water; firm boundaries; transaction costs; transferrable utility matching; empirical matching models

*Vanderbilt University. Email: matthew.okeefe@vanderbilt.edu. I am grateful for the generous support and guidance of Rob Porter, Mar Reguant, and Vivek Bhattacharya. I also thank Michael Powell, Gaston Illanes, Bill Rogerson, Tom Hubbard, Igal Hendel, Meghan Busse, Ryan Kellogg, Alex MacKay, Ameet Morjaria, Gaston Lopez, Francisco Pareschi, and many others for insightful comments.

1 Introduction

The advent of the shale boom in the United States marked a major turning point for the structure of global energy markets, the clean energy transition, and the fortunes of numerous states and localities. The shale boom is often attributed to rapid declines in the costs of key technologies such as horizontal drilling and hydraulic fracturing ("fracking"). Use of these technologies has greatly increased the oil and gas industry's demand for water.

Because water-related costs constitute a significant share of the costs of completing and operating so-called "unconventional" wells, shale firms face private incentives to manage water efficiently. Nevertheless, firms' water usage could be inefficiently high from a social perspective due to water-related externalities such as localized water stress (Backstrom, 2019; Hitaj et al., 2020), ground- and surface-water contamination from spills (Torres et al., 2016), and earthquakes caused by the use of injection disposal wells (Weingarten et al., 2015), among others. Water-related externalities continue to draw attention from policymakers and the academic researchers, with the Texas Railroad Commission having recently implemented major revisions to waste-handling regulations in that state.

Two important strategies firms employ to manage water efficiently are *reuse* and *sharing*. *Reuse* refers to the practice of repurposing wastewater from existing wells as a substitute for freshwater when fracking new wells. Reuse enables firms to acquire less freshwater and dispose of less wastewater than otherwise. *Sharing* refers to trade in reusable wastewater among otherwise competing firms, which can generate surplus due to mismatches between the supply of wastewater from existing wells and ongoing fracking activity.[1] Both reuse and sharing are prevalent in Pennsylvania. Adoption in other regions has so far been less extensive, but this is likely to change if regulation elsewhere continues to tighten.

The alignment between private incentives to reduce water-related costs and the external benefit of reducing water-related externalities creates a tension similar to the central tension in studies of energy efficiency (Gillingham et al., 2016). On the one hand, reuse and sharing directly mitigate water-related externalities by reducing well-level freshwater consumption and final disposal volumes. On the other hand, if reuse and sharing reduce firms' costs (as revealed preference suggests is the case in Pennsylvania), the number of wells drilled may increase in equilibrium. If the latter effect is sufficiently large, total freshwater consumption could increase despite well-level reductions (a form of backfiring).

The first goal of this paper is to quantify the private cost savings from reuse and sharing among shale gas producers in Pennsylvania in order to assess the net effect of reuse and

[1]For example, a firm that is not currently fracking any new wells may find it cheaper to "share" wastewater with a rival firm rather than incurring the (often significant) cost of final disposal.

sharing on water usage. A key challenge in doing so is that public data on water-related costs is scarce and may present an incomplete picture of producers' true cost structure given the temporal and geographic specificities inherent to reuse. Instead of relying on industry or engineering sources, I combine detailed records of wastewater shipments with an empirical matching model that allows me to recover firms' costs using revealed preference.

The model estimates imply that total private cost savings amounted to a little more than 1% of the value of marketed gas during the analysis period (2017-2020), with benefits concentrated among larger firms and firms located further from injection disposal wells. Long-run supply elasticities taken from the literature suggest that cost savings of this magnitude are consistent an increase in marketed production of around 1%. An effect of this size rules out backfiring: most observed drilling would have occurred (at higher cost) even if reuse and sharing had not been possible. Thus, reuse and sharing likely have a large and negative net effect on Pennsylvania shale producers' water consumption and water-related externalities.

In light of this result, it is interesting to consider whether observed reuse in Pennsylvania is maximally efficient with respect to water usage. The evidence points to frictions in the "sharing market" as one potentially important source of inefficiency. Observed sharing transactions are an important source of private cost savings and improved water efficiency. However, both the descriptive evidence and model estimates suggest that sharing is subject to substantial frictions, which I describe as transaction costs. The presence of transaction costs implies that many efficiency-enhancing sharing transactions never occur, and hence that firms' private incentives alone are insufficient to fully optimize water usage.

The final section of the paper explores the nature these transaction costs. I obtain three main results. First, transaction costs are highly heterogenous across bilateral relationships. Among observed sharing relationships, transaction costs are 60% lower at the 25%-tile than at the 75%-tile. Second, modest reductions in transaction costs can have a large impact. Capping bilateral transaction costs within existing sharing relationships at the 25%-tile would reduce final disposal volume by 43%, capturing about half of the maximum reduction achievable in a world without any transaction costs. Third, transaction costs are lower for shipments to large firms than small firms, and greatest for shipments from the largest firms to smallest firms. The first two results suggest that efforts to reduce transaction costs could be a useful complement to conventional policy instruments such as Pigouvian taxation. I interpret the third as evidence that liability aversion may be a more important driver of the estimated transaction costs than search costs, information frictions, or foreclosure motives.

Pennsylvania is an especially interesting setting in which to study wastewater reuse and sharing for two reasons. First, Pennsylvania is the second-largest natural gas producing state in the United States after Texas. Second, Pennsylvania has essentially prohibited the construction of new injection disposal wells since the start of the shale boom. In recent years, other major oil and gas producing states, including New Mexico and Oklahoma in addition to Texas, have tightened regulation of injection disposal wells. Evidence from Pennsylvania offers a unique perspective on the tradeoffs associated with these developments.

The primary data used in the analysis are well pad-level disposal records published by the Pennsylvania Department of Environmental Protection (DEP). I link disposal facilities reported in the data to records of well ownership and other data sources in order to distinguish between internal reuse, sharing, and final disposal. The data indicate that reuse and sharing activity decline sharply in shipment distance, as might be expected given firms' reliance on costly water-hauling trucks for transportation. However, a case study of a major merger provides evidence that sharing may entail significant transaction Coasean costs (Coase, 1937).

To account for these features I adopt a transferrable utility (TU) matching framework (Choo and Siow, 2006; Galichon and Salanié, 2022). Each month, shipments of wastewater directed to internal reuse and sharing are determined simultaneously through a matching game. Both types of reuse can generate cost savings relative to firms' outside options, but sharing is potentially subject to transaction costs. Crucially, the connection between match stability and the concept of the core provides a justification for decentralizing water management decisions within the firm. This allows me to tractably endogenize the "make-vs-buy" decisions that govern firms' reuse and sharing activity while avoiding the need for a complex multilateral bargaining model. In turn, estimation and counterfactual analysis are greatly simplified relative to alternative models. Under certain parametric assumptions the model yields a gravity equation that can be estimated using standard methods.

Related literature Several authors in industrial organization have examined productivity improvements during the shale boom. Covert (2015) and Steck (2022) estimate learning models with the aim of understanding shale producers' input choices. Agerton (2020) studies location choices. Though each these papers (and Herrnstadt et al., 2024) recognize water as a crucial input to the production process, none has directly addressed reuse, sharing, or water-related externalities. In parallel, the large applied literature studying local impacts of the shale boom has largely set aside the determinants of water usage.[2] Within this literature, Muehlenbachs et al. (2015) and Hill and Ma (2022) discuss the impacts of potential groundwater contamination near well sites on housing prices and infant birthweights, respectively. Koster and van Ommeren (2015) find sizable impacts on housing prices from seismic activity linked to natural gas extraction in the Netherlands. Gibbons et al. (2021) conclude that housing price impacts from seismicity in the UK are largely driven by perception.

[2]Jackson et al. (2014), Mason et al. (2015), and Black et al. (2021) offer comprehensive surveys.

A small number of papers in industrial organization have used TU matching models to study questions pertaining to the organization of production. Akkus et al. (2016) analyze the determinants of bank mergers. Fox (2018) studies production networks for automobile parts. The models in these papers relax Choo and Siow (2006)'s assumption that unobserved heterogeneity is additively separable across agents but require maximum score estimation. TU matching models that retain additive separability have been estimated in the context of marriage (e.g., Chiappori et al., 2017) and labor markets (e.g., Corblet, 2025).

It is striking that the estimator closely resembles that of Atalay et al. (2019), who rely on spatial variation to quantify firm boundaries using a gravity model derived from very different microfoundations.[3] Like Atalay et al. (2019), I find that the costs of crossing the firm boundary (in this case, the transaction costs of sharing) are frequently large in comparison to gains from trade. Because each firm has many geographically dispersed wells which continually produce wastewater, the data in this setting affords the opportunity to precisely estimate relationship-level variation in transaction costs. This distinguishes my estimates from those of Atalay et al. (2019), while my results lack their generality.[4] Despite the large empirical literature on transaction cost economics (TCE), few other papers have sought to directly estimate Coasean transaction costs, in large part due to the scarcity of within-firm transaction data. One exception is Masten et al. (1991), who also find evidence of economically large transaction costs using internal production cost data from a shipbuilder. In the industrial organization literature, MacKay (2022) finds that transaction costs can dissipate a large share of buyer surplus in procurement. Hodgson (2022) rationalizes trade-in programs for durable goods as an equilibrium response to high transaction costs.

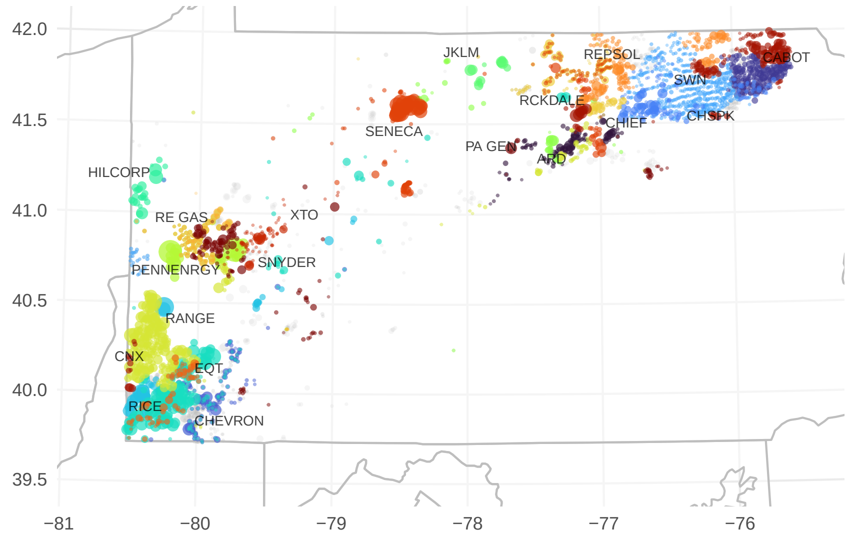
Roadmap The remainder of the paper is structured as follows. Section 2 introduces the data and provides descriptive evidence on reuse, sharing, and transaction costs. Section 3 develops the empirical model. Section 4 describes estimation and introduces the main estimates. Section 5 presents the estimated cost savings from reuse and sharing. Section 6 analyzes the transaction cost estimates and discusses their implications. Section 7 concludes.

2 Background

In this section, I provide a description of wastewater reuse and sharing in Pennsylvania.

[3]The connection between TU matching and gravity models is recognized in Galichon and Salanié (2024).
[4]Atalay et al. (2019) analyze a large swathe of goods-producing and goods-distributing firms in the United States using Census data.

Figure 1: Well Pad Locations for Selected Firms



Data sources The primary data are derived from monthly disposal records that oil and gas producers file with the Pennsylvania Department of Environmental Protection (DEP). These records indicate the disposal method and destination of all quantities of waste materials leaving each well pad, including every barrel of wastewater.[5] Transfers for reuse are clearly indicated. Beginning in 2017, information identifying the destination facility is included in the case of transfers for reuse. I therefore focus on the period from 2017 to 2020.

Additional data sources include well-level completions data from the FracFocus database and freshwater consumption data from the Susquehanna River Basin Commission (SRBC) for wells in Northeastern Pennsylvania. Further detail is provided in Appendix A.

2.1 Reuse vs. injection disposal

Oil and gas extraction in Pennsylvania is conducted by numerous firms ranging from small, independent firms operating only a few wells to the largest global energy firms (Small et al., 2014). Figure 1 shows the locations of well pads operated by each of the twenty largest operators (by disposal volume) in the period that I study. The clustering visible in the figure reflects economies of density in exploration, leasing, drilling, and marketing, as well as in freshwater and wastewater management, which I discuss in this section.

Fracking is water intensive. A typical completion requires more than one hundred thou-

[5] A well pad is a small parcel of land from which multiple wells (typically less than ten) are drilled in different directions or at different depths. Wells on the same pad share common infrastructure such as access roads and (importantly) wastewater storage tanks, making well-level waste attribution challenging.

sand barrels of water (4.2 million gallons).[6] This demand has only increased in recent years as well length and fracking intensity have increased. During fracking, water is blended with sand or other small particles (known as proppants) along with specialized chemicals before being injected into a well under pressure. After completion, a large share of this fluid returns to the surface, together with significant volumes of groundwater. Wastewater production continues for the life of a well in steadily diminishing volumes.[7] Much like with hydrocarbons, the amount of wastewater that a given well will produce is difficult to predict. Wells in Pennsylvania are considered "dry" in the sense that the volume of wastewater typically amounts to around 50% of the volume of fluid injected. This differs significantly from other shales in the United States such as the Permian in which wastewater volumes can exceed or even double injection volumes (Kondash et al., 2018).

The two primary methods of managing wastewater available to Pennsylvania oil and gas firms are injection disposal and reuse.[8] Wastewater is highly saline and may contain organic compounds, metals, and naturally occurring radioactive materials. Careful handling and specialized disposal are required by law (Groundwater Protection Council, 2019). Injection disposal involves the use of specialized wells to deposit wastewater beneath nonporous rock formations deep below the surface of the earth. Reuse involves transporting wastewater to a new well site, performing inexpensive treatments such as filtering and pH reduction, and then using this treated water in place of freshwater when fracking new wells. In either case, the vast majority of wastewater is transported by tanker truck.[9]

Scarcity of injection disposal capacity Geological considerations render it challenging to drill injection wells in Pennsylvania and West Virginia (McCurdy, 2011). For this reason, most injection wells are located in eastern Ohio. The distance between Pennsylvania gas wells and Ohio injection wells can be significant. This is illustrated in Figure 2, which shows the locations of injection wells relative to gas wells observed in the data. Long shipment distances render injection disposal less economical than reuse for most firms. In the data, injection disposal accounts for only 12.3% of all wastewater volume. Wells located in the dense cluster of drilling activity visible in Northeastern Pennsylvania are roughly 300 miles

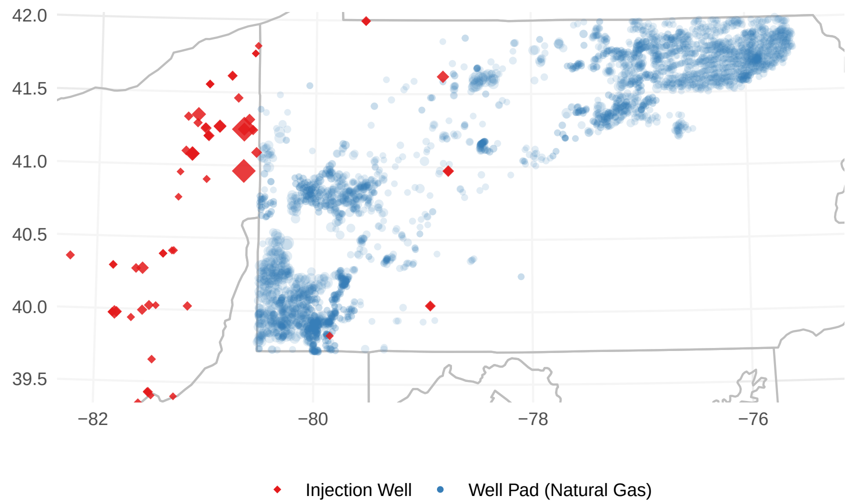
[6] A "barrel" is a standard measure of liquid volume in the oil and gas sector. A barrel holds 42 gallons.

[7] Water produced during or immediately after completion is often referred to as "flowback," while water produced later is more commonly known as "produced water." I use the term wastewater to encompass both.

[8] In western states, evaporation and agricultural application are also used (Groundwater Protection Council, 2019). Another less commonly used option is desalination. In Pennsylvania, reuse outside the oil and gas industry is extremely limited. Disposal at publicly owned treatment works (POTW) is (now) prohibited.

[9] Approximately 5% of produced water is transported by pipeline (Groundwater Protection Council, 2019). According to the DEP, some operators have also used rail transportation.

Figure 2: Injection Well and Well Pad Locations



from Ohio injection wells.[10] In this region, the injection disposal rate is less than 1%.

Reuse as a matching problem Much like the hydrocarbons it accompanies, wastewater is produced in declining volumes over the life of a well. Consequently, the number of well pads generating wastewater (a stock) is large in comparison to the number of new completions (a flow). In the data, 1,712.6 distinct well pads reported wastewater disposals each month, while only 55.0 new wells were completed.[11] The median disposal amount was 415 barrels per month — less than 0.5% of the fluid volume needed to complete a new well. In contrast, well pads appearing as destination received 46,814 barrels of wastewater from 31.4 distinct sources. This information is presented in Table 4, which summarizes the number of facilities appearing in the data each month and associated shipment volumes (in "truckloads").[12]

Centralized treatment facilities (CTFs) Treatment prior to reuse can occur on premises at a well site or at a centralized treatment facility (CTF). Some CTFs are operated by oil and gas producers, while others are operated by third party treatment firms. Producer-affiliated

[10] I define "Northeastern" Pennsylvania as the region coinciding with the Susquehanna River Basin and "Southwestern" Pennsylvania as the region encompassing the Ohio River and Great Lakes Basins.

[11] To obtain the mean number of completions, I take the average number of fracking jobs recorded in FracFocus during the sample period. By comparison, EIA's Drilled but Uncompleted Wells (DUC) data implies 101.5 completions per month for the whole of Appalachia, including Ohio and West Virginia.

[12] The raw data are reported in barrels. I define a truckload as 110 barrels, the modal shipment volume in the data. In practice, water-hauling tanker truck capacity varies from about 80 to 130 barrels.

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.[13]

Limitations The data have a few important limitations. Only the total volume of water transferred between two locations during a month is recorded, rather than the dates, modes, volumes, or circumstances of particular shipments. In particular, the data do not include prices, contract terms, or details pertaining to intermediation.[14] In the case that wastewater is reused, the data do not indicate what treatments were performed or where treatment occurred. Finally, in the case that wastewater is initially transferred to a centralized treatment facility (CTF), the ultimate location of reuse is not reported.[15]

2.2 Wastewater sharing

Wastewater is typically reused by the firm that produced it. However, in some circumstances, firms trade ("share") wastewater with one another prior to reuse.

Table 3 reports the share of wastewater volume by disposal method and destination type. 88.6% of wastewater is transferred to well pads or CTFs for the purpose of reuse. I classify well pads and CTFs as "own," "rival," or "third party" based on the identity of the primary operating firm associated with each facility. For 9.4% of shipments to well pads and CTFs, I can conclusively link the destination facility to a rival firm. Hence, wastewater frequently crosses firm boundaries prior to reuse. I discuss potential motivations for sharing below.

Classification Because the ultimate location of reuse is not reported in the case of shipments to CTFs, it is difficult to gauge the true extent of sharing activity. From Table 3 alone the "true" sharing rate (referring to the share of wastewater ultimately shipped to rival well pads) can be conservatively bounded between 7.1% and 47.6%. If one assumes that all wastewater transferred to operator-affiliated CTF is reused by the recipient, these bounds can be tightened to 9.4% and 23.0%. If one further assumes that all wastewater transferred to third party CTFs is reused by the sender, they collapse to 9.4%. I view this final number as a reasonable (albeit imperfect) estimate of the true sharing rate. For simplicity, the

[13] The choice between CTF and on-pad treatment depends on the tradeoff between scale economies in treatment and transportation costs. The use of CTFs can increase transportation costs because wastewater must be transported twice – once to the CTF, and then again to a location where it can be reused. There are also differences in regulatory treatment (such as bonding requirements).

[14] For example, I do not know whether apparent "sharing" transactions necessary result from direct interaction between two rival firms. This affects the interpretation of the estimated transaction costs.

[15] In contrast, re-transfer of wastewater from one well pad to another is prohibited by the DEP.

remainder of the paper adopts both assumptions: a "sharing transaction" is henceforth a transaction in which the destination well pad or CTF is linked to a rival oil and gas firm.

2.2.1 Which firms share wastewater?

During the sample period, 49 out of 75 firms shared wastewater (i.e., transferred wastewater to a rival firm) on at least one occasion, including 9 of the 10 largest firms. On average, a firm that shared at least once shared in more than half of all sample months. Thus, sharing is widespread and firms that share tend to do so frequently.

One potential motivation for sharing is to realize more efficient matchings of wastewater from old wells to new wells (and CTFs). Because it is costly to transport wastewater, distance appears to be an especially important factor in shaping matching patterns. For more than half of all shipments, the destination facility was one of the five nearest facilities appearing as a destination for shipments from any origin in the same month. The last section of Table 4 shows the distribution of shipment distance by destination type. The mean shipment distance was 30.0 miles. In comparison, the average distance between a randomly selected well pad-destination pair is roughly five times greater. Shipments to rival well pads and CTFs tend to be slightly longer than shipments to internal well pads or CTFs, but shipments to injection disposal wells originating from the same locations would typically be much longer.

Table 5 presents a series of Poisson regressions of monthly disposal shares by destination type on firm characteristics. Firms located further from injection wells send a greater share of wastewater to well pads and CTFs, reflecting the cost of distance. Throughout the paper, I distinguish between three broad categories of firms: "national" firms with substantial operations in multiple oil and gas plays, "large regional" firms that operate exclusively in Appalachia (or nearly so), and a fringe of "small regional" firms (see Appendix A). After controlling for distance to injection wells, small regional firms (large regional firms) send a significantly greater (smaller) share of wastewater to rival well pads and CTFs than national firms. The seven large regional firms account for 69% of wastewater disposal volume and 66% of shipments to well pads and CTFs, but only 28% of shipments to rival well pads and CTFs. Conversely, small regional firms account for only 12% of wastewater and 10% of shipments to well pads and CTFs, but 44% of shipments to rival well pads and CTFs.

2.2.2 Evidence of transaction costs

A merger of two large regional firms that occurred during the sample period provides evidence that sharing can be subject to significant transaction costs.

In November 2017 EQT Corporation ("EQT") and Rice Energy Inc. ("Rice") merged, creating the largest natural gas producer in the United States. Prior to merging, EQT and Rice both operated large numbers of wells on overlapping acreage in far southwestern Pennsylvania.[16] The locations of EQT and Rice well pads are shown in Figure 3. In the six months leading up to the merger announcement, 98% of Rice's wastewater disposals originated at well pads within 20 miles of an EQT facility that concurrently received wastewater. 64% of EQT's wastewater disposals originated at well pads within 20 miles miles of a Rice facility that concurrently received wastewater. Nevertheless, EQT and Rice did not share prior to the merger, despite sharing modest amounts with other firms.[17] This can be seen in Table 6, which summarizes shipments of wastewater to and from EQT/Rice well pads.

Table 6 also shows that shipments between formerly-unintegrated EQT/Rice well pads increased dramatically after the merger. After the merger, 22.5% of wastewater generated at former EQT well pads was transferred to former Rice well pads, while 62.4% of wastewater generated at former Rice well pads was transferred to former EQT well pads. Hence, the removal of the firm boundary was followed by a significant increase in "sharing." It is important to recognize that this outcome need not have occurred. For example, if the main factor preventing sharing in the period prior to the merger were technological in nature, such as a chemical incompatibility between the firms' fracking fluids, the merger need not have induced any "sharing" at all. Thus, I interpret this change as evidence that the merger resulted in the elimination of Coasean transaction costs, which I now define.

Coasean transaction costs Throughout the paper I describe transaction costs in the sharing market as "Coasean." This term lacks a widely accepted definition and could be taken to encompass any fixed or variable costs that erode gains from trade.[18] For the purpose of this paper paper, the essential feature of Coasean transaction costs is that they can be eliminated through integration.

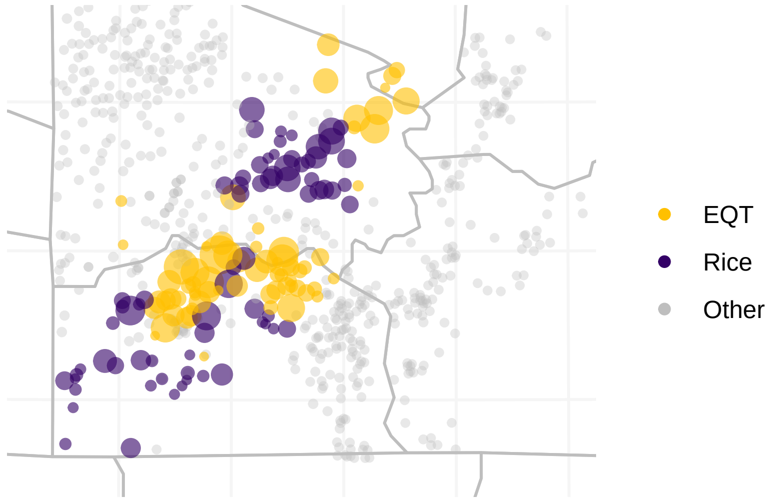
In one classic taxonomy of transaction costs, Dahlman (1979) distinguishes between search and information costs, haggling and decision costs, and policing and enforcement costs. Mechanisms falling within any of these categories could be present in this setting. For instance, it may be difficult for a firm with excess wastewater to learn which other firms are able to accept wastewater, especially if drilling plans are kept secret. If a potential counterparty is found, managers may need to expend scarce time and effort in bargaining over the terms of the transaction (price, quantity, time of delivery, etc.). Formal contracts may need

[16] EQT also had a presence in West Virginia and northeastern Pennsylvania; Rice was present in Ohio.

[17] Rice did not sent wastewater to any rival firms in the pre-merger period. All of the shipments to Rice observed in the pre-merger period originated at well pads associated with Alpha Shale Resources. Rice and Alpha Shale Resources had formerly been partners in a joint venture.

[18] Atalay et al. (2019) offer one useful survey of the vast literature on firm boundaries; Coasean transaction costs are among the most important concepts in this literature. See also Lafontaine and Slade (2007).

Figure 3: EQT/Rice Pre-Merger Well Pad Locations



to be written. Resources may be expended ex post in order to ensure that shipments are timed as expected by both parties and that wastewater quality is acceptable.

I do not attempt to distinguish between costs that fall neatly within this framework and other costs (or shadow costs) arising from interactions between non-integrated firms. Even if there were no search frictions, firms might refuse trade in order to avoid disclosing the composition of fracking fluids or other valuable information.[19] Particular firms might be deemed unsuitable counterparties due to the potential for environmental liability. Frictions of this nature are less commonly described as transaction costs, but function similarly in reducing or eroding gains from trade attainable through integration.

3 Model

This section introduces the model. The goal of the model is to quantify the benefits firms derive from reuse and sharing. To do so, the model must capture firms' tradeoff between reuse and the combination of freshwater acquisition and final disposal, as well as tradeoff between internal reuse and sharing.

I adopt the framework of transferrable utility (TU) matching (see, e.g., Choo and Siow, 2006). Let K_t denote the finite set of well pads generating wastewater in month t, and D_t the finite set of facilities accepting wastewater for reuse. Firm f manages a subset of well

[19]Firms in Pennsylvania are required to publicly disclose fracking fluid composition; however trade secret designations for specific additives are permitted in some cases.

pads $K_{tf} \subset K_t$ and a subset of facilities that accept wastewater $D_{tf} \subset D_t$. In month t, well pad $\kappa \in K_t$ generates Q_κ truckloads of wastewater that must be disposed of immediately. Up to C_δ truckloads of wastewater can be accepted at facility $\delta \in D_t$, but no more. All wastewater must be shipped to some facility in D_t or to final disposal, the outside option. For the remainder of this section I suppress the dependence of all objects on t.

The cost of transporting a truckload of wastewater from $\kappa \in K$ to $\delta \in D$ for the purpose of reuse is given by $r_{\kappa\delta} - \epsilon_{i\delta} - \eta_{\kappa j}$, where $r_{\kappa\delta}$ denotes the systematic cost of reuse and $\epsilon_{i\delta}$ and $\eta_{\kappa j}$ capture shipment-specific cost shocks incurred at the origin and destination facilities, respectively. Similarly, the cost of transporting a truckload of wastewater from κ to final disposal is $r_{\kappa 0} - \epsilon_{i0}$, and the opportunity cost of foregoing acceptance of wastewater at δ is $r_{0\delta} - \epsilon_{0j}$. I interpret the latter as the cost of obtaining freshwater. I assume that the vectors of cost shocks $\epsilon_{i\cdot} = (\epsilon_{i0}, \ldots, \epsilon_{iD})$ and $\eta_{\cdot j} = (\eta_{0j}, \ldots, \eta_{Kj})$ consist of independent draws from centered extreme value type I distributions with scale parameters σ_ϵ and σ_η.[20]

I consider a matching game in which firms choose exactly one shipment destination δ for each truckload i and exactly one supply location κ for each unit of capacity j. In the matching game, each firm seeks to minimize its total cost of wastewater disposal and water acquisition. Firms may choose to transport wastewater to their own facilities or to facilities owned by other firms. If $\kappa \in K_f$ and $\delta \in D_f$, firm f incurs the full cost of reuse. On the other hand, if $\kappa \in K_f$ and $\delta \in D_{f'}$ for some rival firm f', then the systematic cost $r_{\kappa\delta}$ is divided between f and f' while firm f incurs the shock $-\epsilon_{i\delta}$ and firm f' incurs the shock $-\eta_{\kappa j}$. The division of systematic costs occurs in equilibrium. Firm f's equilibrium share of the total cost of shipping truckload i from $\kappa \in K_f$ to $\delta \in D_0 \equiv \{0\} \cup D$ can be denoted by $r^f_{\kappa\delta} - \epsilon_{i\delta}$. Similarly, firm f's equilibrium share of the total cost of allocating capacity j at $\delta \in D_f$ to a shipment from $\kappa \in K_0$ can be denoted by $r^f_{0\delta} - \eta_{0j}$, where $K_0 \equiv \{0\} \cup K$. Utility is assumed to be transferrable in the sense that $r^f_{\kappa\delta} + r^{f'}_{\kappa\delta} = r_{\kappa\delta}$ for any f and f'. The assumption of transferrable utility is reasonable because firms can readily exchange cash.[21] 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.

The equilibrium concept is the core. Suppose the total volume of wastewater generated and accepted in a month grows uncountably large in proportion to $\mathbf{Q} = (Q_1, \ldots, Q_K)$ and $\mathbf{C} = (C_1, \ldots, C_D)$. In this "large market" limit, the core consists of a unique stable matching (Gretsky et al., 1992).[22] A matching can be described by a matrix $\mu \in \mathbb{R}^{K \times D}_{\geq 0}$, where $\mu_{\kappa\delta}$

[20]Galichon and Salanié (2022) establish identification of the systematic match surplus for alternative parametric shock distributions, such as the nested logit. However, Gualdani and Sinha (2023) show that generic non-parametric restrictions on the shock distributions have little identifying power.

[21]The analysis does not exclude non-monetary transfers such as "favors" (Samuelson and Stacchetti, 2017).

[22]The core of the matching game consists of the set of all *stable*, *feasible* matchings. A matching is *feasible*

represents the mass of shipments between κ and δ. In the unique stable matching μ^*, firm f ships truckload i from κ to the least-cost destination among all destinations in D_0:

$$\min_{\delta \in D_0} \quad r^f_{\kappa\delta} - \epsilon_{i\delta}. \tag{1}$$

Simultaneously, each unit of capacity j at δ is allocated to the least-cost origin among all origins in K_0:

$$\min_{\kappa \in K_0} \quad r^f_{\kappa\delta} - \eta_{\kappa j}. \tag{2}$$

Hence, no firm would prefer a destination for i or origin for j to the one assigned by the stable match. It follows immediately that for any $\kappa\delta \in K \times D$

$$\mu^*_{\kappa\delta} = Q_\kappa P\left(\delta = \arg\min_{\delta' \in D_0} r^f_{\kappa\delta} - \epsilon_{i\delta'}\right) = C_\delta P\left(\kappa = \arg\min_{\kappa' \in K_0} r^f_{\kappa\delta} - \eta_{\kappa j}\right).$$

In words, $\mu^*_{\kappa\delta}$ coincides with the operator of well pad κ's demand for shipments to facility δ and the operator of facility δ's demand for shipments from well pad κ when choices are made independently according to (1) and (2).[23]

Galichon and Salanié (2022) demonstrate that the stable matching μ^* is the solution to a regularized optimal transport (OT) program. In particular, it is obtained by minimizing a social cost function:

$$\min_{\mu \in \mathcal{M}(\mathbf{Q},\mathbf{C})} \sum_{\kappa \in K} \sum_{\delta \in D} \mu_{\kappa\delta} \left\{r_{\kappa\delta} - r_{\kappa 0} - r_{0\delta}\right\} - \mathcal{E}\left(\mu, \mathbf{Q}, \mathbf{C}\right) \tag{3}$$

where $\mathcal{M}\left(\mathbf{Q}, \mathbf{C}\right)$ is the set of feasible matchings and $\mathcal{E}\left(\mu, \mathbf{Q}, \mathbf{C}\right)$ depends on the distributions of $\epsilon_{i\cdot}$ and $\eta_{\cdot j}$.[24] Let $\mu_{\kappa 0} = Q_\kappa - \sum_{\delta \in D} \mu_{\kappa\delta}$ and $\mu_{0\delta} = C_\delta - \sum_{\kappa \in K} \mu_{\kappa\delta}$ denote the mass of unmatched wastewater originating at κ and the mass of unmatched capacity at δ, respectively. Fixing the scale parameters of the distributions to be $\sigma_e = \sigma_\eta = \sigma$, the match entropy

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 $\kappa \in K_0$. A matching μ is *stable* if no firm would prefer to ship a matched truckload i to final disposal or to allocate a matched unit of capacity j to freshwater, and no two firms (possibly the same) would privately agree to match any i and j not matched under μ.

[23] For intuition, one can imagine individual "managers" located at each well pad and CTF making decisions about wastewater disposal or water acquisitions on a truckload-by-truckload basis.

[24] A matching is feasible if its rows sum are weakly less than than Q and its column sums are weakly less than C.

function $\mathcal{E}(\mu, \mathbf{Q}, \mathbf{C})$ is given by

$$\mathcal{E}(\mu, \mathbf{Q}, \mathbf{C}) = -\sigma \sum_{\kappa\delta \in K \times D_0} \mu_{\kappa\delta} \log\left(\frac{\mu_{\kappa\delta}}{Q_\kappa}\right) - \sigma \sum_{\kappa\delta \in K_0 \times D} \mu_{\kappa\delta} \log\left(\frac{\mu_{\kappa\delta}}{C_\delta}\right). \tag{4}$$

The match entropy $\mathcal{E}(\mu, \mathbf{Q}, \mathbf{C})$ quantifies the contribution of the cost shocks $\epsilon_{i\cdot}$ and $\eta_{\cdot j}$ to the match surplus. In the limit as $\sigma \to 0$, the systematic component $r_{\kappa\delta} - r_{\kappa 0} - r_{0\delta}$ dominates and (3) reduces to a linear program (LP).

As (3) makes clear, equilibrium shipping patterns are driven by the structure of $r_{\kappa\delta} - r_{\kappa 0} - r_{0\delta}$: the systematic costs of reuse less the cost of the outside option (injection disposal paired with freshwater acquisition). I allow the systematic costs of reuse to depend shipment distance and the presence of firm boundaries. Parameterization is discussed below.

Discussion The model assumes that firms make decisions on a truckload-by-truckload basis, both when sending and receiving wastewater. This assumption is without loss in the context of a matching model: the firm can be viewed as a coalition of individual managers, one located at each facility, and in the core no coalition of managers can achieve lower costs than managers acting independently. This feature of the model represents an important simplification: I can avoid formally modeling what would otherwise be a highly complex multilateral bargaining process. The cost is that firms have no market power and do not behave strategically. For example, a firm cannot earn more surplus by threatening to abstain from sharing (Shapley and Shubik, 1971). The "Nash-in-Nash" model associated with Horn and Wolinsky (1988) offers one alternative framework that could be used to capture some important forms of strategic behavior. Apart from the lack of price data, one practical challenge in adopting the Nash-in-Nash framework in this setting is the difficulty of constructing counterfactuals given the large number of potential trading partners.[25] Moreover, models of Nash-in-Nash bargaining typically do not account for transaction costs (Carlton, 2020).

Another important restriction imposed by the model is that systematic costs are incurred in proportion to shipment volume. In fact, many plausible types of transaction costs are best understood as fixed costs that might be incurred once for multiple shipments (e.g., search costs), or even once per relationship (e.g., the cost of drafting a first contract). Such costs are rationalized by the model in the form of higher variable transaction costs. As a result, estimated costs unrelated to distance (including the transaction costs) should be regarded as reduced form objects rather than true variable costs. Explicitly accounting for fixed

[25] Ho and Lee (2019) and Ghili (2022) develop models of network formation in Nash-in-Nash environments. These papers simplify the strategy space by exploiting institutional differences between upstream and downstream firms that have no clear analogues in this setting.

transaction costs would require a considerably more complex model given the size of the market and potential for dynamics.

4 Estimation

I estimate the model using a moment-matching procedure recommended by (Galichon and Salanié, 2022, 2024). This section motivates the estimator.

Parameterization The key object that must be estimated is the systematic component of the cost savings from reuse (analogous to the match surplus in a marriage model). I specify this object to be a linear function of observable origin and destination facility characteristics:

$$\{\sigma_\epsilon + \sigma_\eta\}^{-1} (r_{\kappa 0} - r_{0\delta} - r_{\kappa\delta}) = x'_{\kappa\delta}\theta. \tag{5}$$

where the left hand side is first normalized by the sum $\sigma_\epsilon + \sigma_\eta$. For the main specification, the covariate vector $x_{\kappa\delta} = \left(x^{(1)}_{\kappa\delta}, \ldots, x^{(J)}_{\kappa\delta}\right)$ includes distance (in miles), a constant, and two sets of variables that capture heterogeneity in the opportunity costs of reuse and heterogeneity in transaction costs, respectively. The first includes a proxy for the distance from κ to injection wells and fixed effects corresponding to the type of firm at δ (small or large). These variables give the model flexibility to account for differences in disposal and treatment costs. The second set of variables includes a (symmetric) fixed effect for every pair of firms that shared at least 50,000 barrels of wastewater during the sample period as well as a single "residual" fixed effect for all other pairs of firms within which some sharing occurred.[26] These variables are intended to capture transaction costs as flexibly as possible.

Moment-matching procedure Standard arguments imply that the equilibrium matching conditional on θ is described by the following equations:

$$\mu^\theta_{\kappa\delta} = \exp\left\{-x'_{\kappa\delta}\theta + F_\kappa + H_\delta\right\}$$
$$\mu^\theta_{\kappa 0} = \exp\left\{\gamma^{-1} F_\kappa\right\}$$
$$\mu^\theta_{0\delta} = \exp\left\{(1-\gamma)^{-1} H_\delta\right\}$$

where the vectors $\mathbf{F} = (F_1, \ldots, F_K)$ and $\mathbf{H} = (H_1, \ldots, H_D)$ uniquely solve the market clearing conditions $\mu^\theta_{\kappa 0} + \sum_\delta \mu^\theta_{\kappa\delta} = Q_\kappa$ and $\mu^\theta_{0\delta} + \sum_\kappa \mu^\theta_{\kappa\delta} = C_\delta$ for all $\kappa \in K$ and $\delta \in D$. Intuitively,

[26] These pairs of firms account for approximately 90% of sharing. If a pair of firms never shared, I assume that transaction costs were infinite. Transaction costs are not point identified in this case.

the market clearing conditions correspond to the feasibility restriction on the equilibrium stable match. The fixed effects are interpreted as the expected cost per truckload at each facility. Under the maintained assumptions on the distributions of ϵ and η, they are sufficient to summarize the equilibrium division of surplus between sending and receiving firms.

Turning to estimation, for each element j of covariate vector I construct moment functions of the form:

$$g_j(\theta) = \sum_{(\kappa,\delta)\in A} \left(\hat{\mu}_{\kappa\delta} - \mu^{\theta}_{\kappa\delta}\right) x^{(j)}_{\kappa\delta} \tag{6}$$

where $A = K \times D \cup K \times \{0\} \cup D \times \{0\}$ is the set of all origin-destination combinations. The estimator $\hat{\theta}$ is a root of $g(\theta) = 0$. In order for (6) to be a valid moment condition, it must be the case that $E\left[\hat{\mu}_{\kappa\delta} | x_{\kappa\delta}\right] = \mu^{\theta}_{\kappa\delta}$. This assumption may be violated if there are unobserved factors that shift the systematic cost of reuse for both the sending and receiving firm (as opposed to their separable, idiosyncratic shocks ϵ and η). However, any potential bias is likely mitigated by the fact that (6) aggregates over many different origin-destination pairs.

The model is just identified. The constant in $x_{\kappa\delta}$ ensures that the total volume of reuse in the model exactly matches the data. Likewise, the firm pair fixed effects ensure that bilateral sharing volumes in the model exactly match the data for all relationships in which sharing volume exceeds 50,000 barrels.[27] The remaining moments target the mean shipment distance among shipments for reuse, wastewater disposal conditional on location, and freshwater consumption conditional on firm type.

The use of the method of moments rather than the generalized method of moments (GMM) is motivated by the observation that moment conditions of the form (6) comprise the first order conditions in a Poisson pseudomaximum likelihood (PPML) estimation:

$$\begin{aligned} \max_{\theta,F,H} \quad & \sum_{\kappa\delta\in KD} \left\{\hat{\mu}_{\kappa\delta}\log\mu^{\theta}_{\kappa\delta} - \mu^{\theta}_{\kappa\delta}\right\} \\ & + \sum_{\kappa\in K} \gamma \left\{\hat{\mu}_{\kappa 0}\log\mu^{\theta}_{\kappa 0} - \mu^{\theta}_{\kappa 0}\right\} + \sum_{\delta\in D} (1-\gamma) \left\{\hat{\mu}_{0d}\log\mu^{\theta}_{0d} - \mu^{\theta}_{0d}\right\} \end{aligned} \tag{7}$$

where $\gamma = \sigma_{\eta}^{-1} / \left(\sigma_{\epsilon}^{-1} + \sigma_{\eta}^{-1}\right)$ captures the relative dispersion of $\epsilon_{i\cdot}$ and $\eta_{\cdot j}$. This approach is appealing because the wastewater shipment data resembles typical trade datasets in which errors are heteroskedastic and the dependent variable frequently takes a value of zero. Despite being inefficient, PPML is known to exhibit good performance in such datasets (Santos Silva and Tenreyro, 2022). Moreover, (7) is computationally feasible: like standard gravity models,

[27] In the case that a pair of firms did not engage in trade in any period, I assume that $r_{\kappa\delta} + r_{\kappa 0} + r_{0\delta} = -\infty$.

it can be solved quickly using iteratively reweighted least squares (IRLS).[28] The parameter γ is treated as known; I fix it to $\gamma = 0.5$ and normalize $\sigma_\epsilon = \sigma_\eta = 1$.[29]

The objective in (7) is not a standard PPML objective due to the presence of the weights as well as the fact that the fixed effects F_κ and H_δ enter non-linearly into the model shares of the outside options $\mu^\theta_{\kappa 0}$ and $\mu^\theta_{0\delta}$. In this respect the model differs from a standard gravity model. In the context of gravity models, Fernandez-Val and Weidner (2016) demonstrate that estimation of θ is not affected by the incidental parameters problem even when the number of fixed effects is large. However, the incidental parameters problem does contribute to biased standard errors and incorrectly centered confidence intervals for inference based on the score of the pseudolikelihood (Kauermann and Carroll, 2001). Zylkin (2024) proposes to address this issue using a two-step bootstrap. I adopt the same procedure.

Freshwater consumption Implementation of (7) requires knowledge of $\mu^t_{0\delta}$, the amount of unmatched reuse capacity at facility δ in month t. This quantity is not directly observable. Maintaining the assumption that $\mu^t_{0\delta}$ represents freshwater consumption, I construct a proxy measure $\hat{\mu}^t_{0\delta}$ by comparing each firms' total water usage as reported in FracFocus to wastewater receipts observed in the disposal data. In Appendix A.1, I show that the freshwater consumption rates implied by the proxy measure are similar to those obtained in a sample of well-level freshwater consumption records provided by the SRBC.

Estimation sample I exclude certain types of shipments from the estimation sample. First, I omit shipments to third party CTFs. These facilities are difficult to accommodate into the modeling framework because they face binding capacity constraints but do not require freshwater, making it difficult to estimate the share of the outside option. Second, I omit shipments originating from well pads that were concurrently recorded as destinations for wastewater from other well pads. By excluding shipments from (but not to) these well pad-months, I avoid predicting that wastewater produced by one fracking event should have been reused as an input for that same fracking event.[30]

[28]IRLS is the typical method for estimating Poisson regression models with high-dimensional fixed effects. The Frisch-Waugh-Lovell theorem is exploited to avoid inverting the full design matrix (Correira et al., 2020). In this case, the fixed effects are updated at each step by using Sinkhorn's algorithm to solve the system of market clearing conditions described above.

[29]In principle γ is identified from the relative rates at which firms substitute to the outside options as the sizes of K_t and D_t vary. However, estimation of γ significantly complicates inference.

[30]Because flowback volumes are significant, this results in a loss of 11.4% of wastewater from the sample. An alternative approach would be to assign an infinite cost to self-shipments.

4.1 Model fit and interpretation

In this subsection I briefly explore model fit. I first explain how the parameter estimates are interpreted in the context of a series of preliminary specifications. This exercise also serves to validate the model. Then I introduce the main estimates.

Preliminary estimates Table 1 presents parameter estimates for a series of preliminary models that I use to demonstrate the interpretation of the parameter estimates. Each model includes only a small number of moments. All include an intercept and a measure of distance. The reported pseudo-R^2 statistic is the square of the correlation between the predicted and realized shipment volumes among shipments for reuse. The models in Columns (1) and (2) use the log of over-the-road mileage as the distance measure. The second column indicates that after controlling for the presence of a firm boundary, the elasticity of shipments for reuse with respect to distance is -0.8, similar to typical estimates in the trade literature.[31] If firm boundaries were not accounted for, the estimated elasticity would be much larger, and the model fit (as measured by the pseudo-R^2 value) would be comparatively poor.

The models in Columns (3) and (4) differ in replacing log mileage with over-the-road mileage itself ("Miles"). Using linear distances yield slightly higher pseudo-R^2 values than the corresponding log specifications. While the coefficient on log mileage has a convenient interpretation, linear distance may be a more suitable proxy for firms' costs in the context of wastewater management: distances are short relative to the distances between foreign countries, and trucking contracts may feature a per-mile component. This finding motivates me to use linear distance rather than log distance in the main specification.

The coefficient estimate for the effect of a firm boundary in Column (4) implies that crossing a firm boundary (in other words, sharing) raises the total cost of reuse by an amount equivalent to the cost of shipping a truckload of wastewater $\frac{4.154}{0.030} = 138.5$ additional miles, more than four times the mean shipment distance observed in the data. For comparison, the optimized objective function in (3) provides a measure of the total cost savings created by reuse and sharing (in other words, the net gains from trade). For every truckload delivered to a well pad or CTF for reuse, total surplus is equivalent to the cost of shipping a truckload of wastewater $\frac{6.186}{0.030} = 206.2$ additional miles. Thus, crossing the firm boundary is estimated to incur a cost which is similar in magnitude to the typical cost savings from reuse and sharing. I return to the interpretation of cost savings in Section 5.

Columns (5) and (6) report similar estimates obtained with alternative models in which

[31] Head and Meyer (2014) report -1.1 to be the mean distance elasticity of trade across a large number of studies estimating structural gravity models. Looking at manufacturing establishments in the US, Atalay et al. (2019) obtain an estimate of -0.96. See also Chaney (2018).

Table 1: Preliminary Model Estimates

	Dependent Variable: # of Truckloads					
	(1)	(2)	(3)	(4)	(5)	(6)
Constant	-4.790^{*}	-1.682^{*}	-0.339^{*}	0.580^{*}	-1.475^{*}	0.794^{*}
	(0.023)	(0.016)	(0.022)	(0.012)	(0.018)	(0.015)
Distance	-1.639^{*}	-0.811^{*}	-0.058^{*}	-0.030^{*}	-0.998^{*}	-1.007^{*}
	(0.008)	(0.007)	(0.001)	(0.0004)	(0.010)	(0.016)
Firm Boundary		-4.432^{*}		-4.154^{*}	-4.565^{*}	-4.386^{*}
		(0.040)		(0.038)	(0.038)	(0.047)
EQT-Rice Post-Merger		-0.026		-0.160^{*}	-0.072^{*}	-0.176^{*}
		(0.022)		(0.024)	(0.021)	(0.027)
Savings/Truck (All)	4.987	5.499	4.971	5.539	5.598	5.656
Savings/Truck (Reuse)	5.569	6.141	5.551	6.186	6.252	6.316
Distance Measure	Log Miles	Log Miles	Miles	Miles	Log Hours	Hours
κ-month FEs	Yes	Yes	Yes	Yes	Yes	Yes
δ-month FEs	Yes	Yes	Yes	Yes	Yes	Yes
Pseudo R^2	0.384	0.767	0.460	0.791	0.760	0.775
Observations	3,843,919	3,843,919	3,843,919	3,843,919	3,843,919	3,843,919

Notes: Significance levels: $^{*}p < 0.05$.

distance is measured in terms of trucking time rather than over-the-road mileage. Model fit is slightly worse than for the corresponding specifications in miles. In Column (6), crossing the firm boundary raises the total cost of reuse by an amount equivalent to the cost of shipping a truckload of wastewater $\frac{4.386}{1.007} = 4.36$ additional hours.

Coasean transaction costs can be eliminated through integration. To assess whether the estimated firm boundary affects can be interpreted as Coasean transaction costs, the specifications that include firm boundaries also include a dummy for post-merger transactions across the pre-merger boundary between EQT and Rice. If mergers eliminate transaction costs, then the estimated coefficient should be equal to zero. For the log distance specification, the estimate is negative but insignificant at the 5% level. For the linear distance specifications, the estimates are negative and statistically significant but small. A small and negative estimate implies a small and positive cost of crossing the pre-merger firm boundary. In Column (4) this cost is estimated to be 3.4% as large as the cost of sharing. Although this estimate is not a statistical zero, I view this result as broadly consistent with a Coasean interpretation.[32] This informs my interpretation of the boundaries between other firms.

Trucking costs Converting point estimates such as those in Table 1 into dollars requires an estimate of the marginal cost of trucking. There is little public information on the cost of wastewater hauling, which is typically intermediated at arms length between an operating firm and a trucking company. My preferred estimate is \$69 dollars per hour, or \$2.5 per mile. This estimate is obtained by adding together estimates of driver wages and fuel costs and then adjusting for empty ("deadhead") return trips (see Appendix A). The estimate is intended to be conservative: I do not account for vehicle depreciation, maintenance, overtime wages, or markups. I revisit these assumptions in Section 5.

Main specification The main specification builds on the model in Column (4) by allowing for greater flexibility in how costs may differ across potential transactions. In this section I briefly illustrate model fit by examining features of the model that are not directly targeted during estimation. Discussion of the main results is deferred to the next sections.

Figures 5a and 5b plot the empirical distributions of shipment distance for internal reuse and sharing, respectively, against the distributions implied by the fitted model. In each case fit is close throughout the distribution. The Kolmogorov-Smirnov statistic is 0.009 for internal shipments and 0.031 for shipments between rivals. This degree of goodness-of-fit may be surprising given that only a single moment in the main specification features distance. In part, this reflects the large number of degrees of freedom in the model (via the facility-month

[32]Residual within-firm friction is not surprising if post-merger integration is not completed immediately.

fixed effects). Another explanation is that firms in practice may solve what amounts to an optimal transport problem much like the one in (3).

Figures 5c and 5d plot the time series of the aggregate reuse and sharing rates, respectively, in the data and the fitted model. The coefficient vector x in the main specification does not include any time-varying variables. Hence, variation in model-implied reuse and sharing rates across months is fully explained by changes in the total amount of wastewater produced and freshwater consumed (i.e., by changes in $\mathbf{Q}$ and $\mathbf{C}$). Despite this, the model reasonably captures both the level and time trend of both series. The mean absolute error in the estimated reuse rate was 1.9 percentage points, compared to a mean of 89.7% in the estimation sample. The mean absolute error in the sharing rate (i.e., the proportion of reuse that involved sharing) was 1.8 percentage points compared to a mean of 10.6%.[33]

5 Private costs

The estimated model implies that firms achieve significant cost savings from reuse and sharing. This section summarizes these results and discusses the implications for aggregate freshwater consumption, wastewater disposal volumes, and external costs.

The model estimates are directly informative about firms' cost savings from reuse and sharing. As noted above, the optimized value of the objective in (3) provides one useful measure of total cost savings. Building on this logic, the dual of (3) can be used to quantify the cost savings accruing to particular firms. For any origin-destination pair, the sending and receiving firms' expected cost savings are given by the inclusive values of (1) and (2), respectively, after normalizing the utilities of the outside options to zero. These inclusive values correspond to the dual variables for the feasibility constraints.

Table 2 summarizes the average expected cost savings conditional on reuse for different types of firms. I report estimates both in terms of mileage and in terms of dollars per barrel of wastewater. I find that the average cost savings conditional on reuse are equal to \$4.18 per barrel of wastewater. This estimate encompasses heterogenous benefits for different firms. A large regional firm in Southwestern Pennsylvania (such as EQT or Rice) saves \$3.76 per barrel. Conditional on location, large regional firms save more than national firms, while national firms save more than small regional firms. In Northeastern Pennsylvania, which is both further from Ohio injection wells and more water-scarce, firms of all sizes save more. The finding that larger firms enjoy larger cost savings can be attributed at least in part to

[33] In addition to illustrating model fit, these results suggest that unmodeled dynamics play at most a limited role empirically during the analysis period, lending support to the use of a (repeated) static framework. Dynamic factors may be more important in the context of commodity price swings. Natural gas prices are widely considered to have been in a "bust" throughout the entire analysis period.

economies of scale: larger firms complete more wells, providing more opportunities to reuse wastewater at shorter distances without incurring transaction costs.

These cost estimates are essential for gauging the potential equilibrium impacts of reuse and sharing on aggregate preduction levels, which drive total water consumption. Aggregating across all shipments, model-implied cost savings from reuse and sharing amount to 1.4% of the total value of gas marketed by Pennsylvania producers during this period.[34] Newell et al. (2019) estimate a supply elasticity of 0.89 for US gas production (including conventional production). Hausman and Kellogg (2015) estimate a long-run supply elasticity of 0.81 for shale gas. Supply elasticities in this range imply that a drop in price equivalent to the ceteris paribus cost increase from eliminating reuse and sharing would reduce aggregate supply by 1.1-1.2%. Put differently: by increasing the effective price received by producers, reuse and sharing may have increased aggregate supply by 1.1-1.2% during the sample period. As water use and wastewater production are proportional to production, this implies that cost savings from reuse and sharing are unlikely to have induced enough marginal production to offset direct reductions in freshwater consumption and final disposal.

This qualitative finding is robust to some violations of the maintained assumptions. If trucking costs are underestimated by a factor of two, the effect on supply would double to 2.3-2.5%. If the supply elasticity is too large, the effect would be smaller. In either case, the magnitude of cost savings suggest that reuse and sharing have only modestly stimulated aggregate production. Nevertheless, the estimates in Table 2 suggest that reuse and sharing have played a more important role in the development of Northeastern Pennsylvania than Southwestern Pennsylvania. Moreover, by disproportionately benefiting larger firms, reuse and sharing are likely to have played a role in shaping the firm size distribution.

Comparison to industry reports The point estimates in Table 2 capture the foregone costs of disposal and freshwater acquisition net of the total costs of reuse (including transaction costs). Without relying on revealed preference, each element of this net benefit would be difficult to quantify. Nevertheless, information from industry sources and engineering studies can be a useful guide to whether the point estimates are reasonable. In an email exchange, one firm indicated that cost savings for firms in Northeastern Pennsylvania are on the order of $10-15 per barrel.[35] As can be seen in the table, I obtain smaller point estimates. One reason for the difference is that the estimated cost savings include the transaction costs of

[34] During the sample period, oil and gas firms in Pennsylvania produced 8.65 barrels of wastewater for every million cubic feet (MMcf) of marketed gas (including conventional production). The average spot price for natural gas was $2.64 per million Btu. (Note that this calculation relies on the Henry Hub spot price. In reality, firms faced a negative basis during this period due to binding pipeline constraints.)

[35] This includes $1 for freshwater, $8-12 per barrel for transportation, and $2-4 per barrel for disposal, net of $0.25 per barrel for treatment.

Table 2: Estimated Cost Savings from Reuse and Sharing

		Estimated Cost Savings	
Region	Firm Type	Miles/Truck	$/Barrel
Southwest	Large regional	166.9	3.76
Southwest	National	117.7	2.65
Southwest	Small regional	80.8	1.82
Northeast	Large regional	384.3	8.65
Northeast	National	247.7	5.58
Northeast	Small regional	141.1	3.18
All	All	185.7	4.18

Notes: (1) $/barrel refers to dollars per barrel of wastewater reused; (2) Assumes a marginal trucking cost of $69 per hour.

sharing and potentially other types of shadow costs (e.g., transaction costs associated with injection disposal). A larger part of the difference is likely explained by the use of relatively conservative trucking cost assumptions. If the marginal cost of trucking were on the order of $150 per hour, which would not be unreasonable, point estimates would fall within the $10-15 range for most firms in Northeastern Pennsylvania.

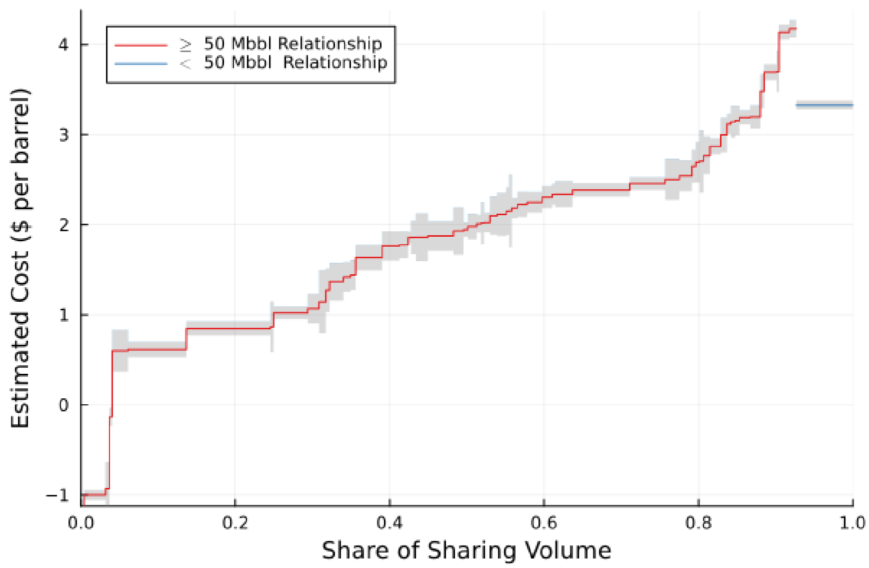
6 Transaction costs

The presence of transaction costs implies that the extent of reuse and sharing may be inefficiently low. This section describes the main transaction cost estimates.

The estimated transaction costs exhibit significant dispersion. Figure 4 plots the distribution of transaction costs. Each step represents a different bilateral relationship. The length of a step indicates the proportion of all sharing activity which occurred within that relationship, while the height indicates the estimated transaction cost. (Recall that relationship fixed effects are estimated only for firm-pairs in which the total volume of sharing exceed 50,000 barrels during the sample period, with a single fixed effect for all other sharing activity.) Estimated transaction costs are $1.02 per barrel at the 25%-tile and $2.46 per barrel at the 75%-tile. The estimates are relatively precise due to the large sample size.

One implication of Figure 4 is that the scope for reducing transaction costs may be large. To provide a sense of scale, I simulate the effect of capping transaction costs among firm-pairs

Figure 4: Distribution of Estimated Bilateral Transaction Costs



Notes: Gray bar denotes 95% confidence interval.

that already share at \$1.02 per barrel (the 25%-tile level). I find that final disposal volume falls by 43% before accounting for changes in drilling rates. For comparison, eliminating transaction costs altogether would reduce final disposal volume by 91%.[36] Thus, even modest reductions in transaction costs limited to pairs of firms that already engage in sharing could have large equilibrium effects on water usage. It is therefore relevant to ask why transaction costs are higher within some relationships than others.

One approach to this question is to explore how the point estimates presented in Figure 4 correlate with relationship characteristics. A challenge with this strategy is that estimates are obtained only for firm-pairs that are observed to share; as a result, the sample is highly selected. Regressions of the estimated firm-pair fixed effects on firm characteristics suggest that transaction costs are positively correlated with measures of relationship surplus (such as firm size or nearness). These results can be explained by selection bias: when the gains from trade are large, larger transaction costs can be sustained in equilibrium (Demsetz, 1988).

To circumvent this issue, I re-estimate the model dropping the firm-pair fixed effects and replacing them with interactions between firm characteristics (such as firm type).[37] Table 7 presents the results. I focus on two general patterns. First, for all types of sending

[36]Changes in freshwater consumption rates are 1% and 2%, respectively.

[37]Making this change also enables me to allow for asymmetries in transaction costs depending on the characteristics of the sending and receiving firms. For many bilateral relationships such asymmetries are not identified because observed sharing is unidirectional.

firms transaction costs tend to be lower when a large regional firm is the recipient. This pattern would be surprising in a world with strong vertical foreclosure incentives or significant potential for moral hazard on the part of the sender (e.g., poor wastewater quality, late deliveries). Though the bargaining process is not formally modeled, one would expect large regional firms to have greater bargaining power than other types of firms on account of completing the largest number of wells and consuming the most water. Given that the costs of final disposal are significant (and much larger than the costs of obtaining freshwater), a large regional firm that sought to drive up rivals' costs would benefit from refusing to accept wastewater for reuse. If moral hazard were an important issue, large regional firms would likely be able to exercise greater selectivity when accepting shipments, resulting in elevated transaction cost estimates.

Second, shipments from large regional firms to small regional firms have the greatest transaction costs. Thus, it appears that the largest firms are especially averse to sending wastewater to the smallest firms.[38] This results is difficult to reconcile with search costs: it is unclear why search costs would be greater for larger senders than for smaller ones. It is also difficult to reconcile with asymmetric information related to the sender's productivity (e.g., trade secrets embodied in wastewater). If such asymmetries are present, the frictions they create are likely more severe among firms of similar size and productivity who compete more closely. One alternative mechanism that better fits explains results is liability aversion: small firms in the oil and gas sector tend to have worse environmental records (Boomhower, 2019), while large firms have more to lose in the event of spills or other accidents.

7 Conclusion

The shale boom lead to significant changes in water use within the oil and gas sector. Because of negative externalities associated with water, including localized groundwater depletion and seismic activity from injection disposal wells, water use within the oil and gas sector has attracted attention both from policymakers and from researchers interested in the local impacts of fracking. In parallel, work in industrial organization on the sources of productivity gains during the shale boom has highlighted the importance of firms' water input decisions and the cost of water. Reuse and sharing are practices that can simultaneously mitigate environmental harms and reduce firms' private costs, reducing external costs but potentially inducing rebound effects. Focusing on the case of Pennsylvania, which has seen more extensive adoption than any other region, I make two contributions. First, I show that

[38] The other possibility is that small firms are especially averse to accepting wastewater from larger firms. It is difficult to imagine plausible mechanisms that would have this implication.

producers' private cost savings from reuse and sharing are substantial but not large enough to have significantly raised equilibrium production levels. Second, I present evidence suggesting that sharing is subject to large transaction costs, especially in the case of shipments from the largest firms to the smallest ones. Despite Pennsylvania's strict regulation of injection disposal, these transaction costs may result in inefficient levels of freshwater consumption and injection disposal.

Table 3: Wastewater Disposal by Destination Type

Destination Type	%
Own well pad	46.5
Own CTF	21.9
Rival well pad	6.3
Rival CTF	2.0
3rd party CTF	12.0
Injection well	8.1
Other	3.3

Notes: (1) Share of wastewater during the sample period classified by disposal method and destination type. (2) Classification of own and rival facilities is described in the text. (3) 3rd party CTFs facilities are centralized wastewater treatment facilities not operated by oil and gas operators. (4) Shipments outside Pennsylvania (including shipments for reuse in West Virginia) and are classified as "Other."

Table 4: Distributions of Selected Statistics

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

Notes: (1) Facility counts: Observations are facility type-months. For well pads (origin) an observation is the number of well pads recording wastewater disposal in a month. For well pads (dest) an observation is the number of well pads appearing as wastewater shipment destinations in a month. Producer CTFs (dest) and 3rd Party CTFs (dest) are defined similarly. (2) Truckloads sent or received: Observations are facility-months. For well pads (origin) an observation is the amount of wastewater shipped form a single well pad in month. For well pads (dest) an observation is the amount of wastewater shipped to a single well pad in a month. Producer CTFs (dest) and 3rd Party CTFs (dest) are defined similarly. A "truckload" is defined as 110 barrels of wastewater (i.e., 4,620 gallons). (3) Miles per truckload: Observations are origin-destination pairs. For own pad or CTF an observation is the distance between a well pad and an own pad or CTF to which it shipped wastewater. Remaining variables are defined similarly for differnt types of wastewater destinations.

Table 5: Poisson Regressions of Disposal Shares on Firm Characteristics

	Share of Wastewater Disposal by Firm-Month					
	Any Well Pad or CTF			Rival Well Pad or CTF		
	(1)	(2)	(3)	(4)	(5)	(6)
Small regional	-0.161		0.179	0.560		0.886^{**}
	(0.276)		(0.273)	(0.372)		(0.397)
Large regional	-0.052		0.311	-1.292^{***}		-0.942^{**}
	(0.393)		(0.333)	(0.470)		(0.468)
Distance to Injection Wells		0.698^{***}	0.772^{***}		0.524^{**}	0.744^{***}
		(0.193)	(0.214)		(0.231)	(0.232)
Constant	-0.843^{***}	-4.464^{***}	-4.996^{***}	-1.820^{***}	-4.135^{***}	-5.816^{***}
	(0.241)	(1.022)	(1.229)	(0.340)	(1.182)	(1.366)
R^2	0.004	0.115	0.117	0.061	0.031	0.101
Observations	2,043	2,043	2,043	2,043	2,043	2,043

Notes: (1) An observation is a firm-month. Standard errors clustered at the firm level. (2) The "National firm" category is excluded from the firm type dummies. (3) Pseudo-R^2 is the square of the correlation between the predicted value and dependent variable. (4) Distance to injection wells is calculated at the firm level as a quantity-weighted average of well pad-month level distances to injection wells. The latter is calculated as a weighted average from the well pad to all observed injection wells. The weights are constructed as ratio of the injection well receipt quantity divided by distance (hence giving more weight to higher capacity injection wells and less to more distant ones). Finally, I take the log of this measure. Significance levels: $^{*}p < 0.1$; $^{**}p < 0.05$; $^{***}p < 0.01$

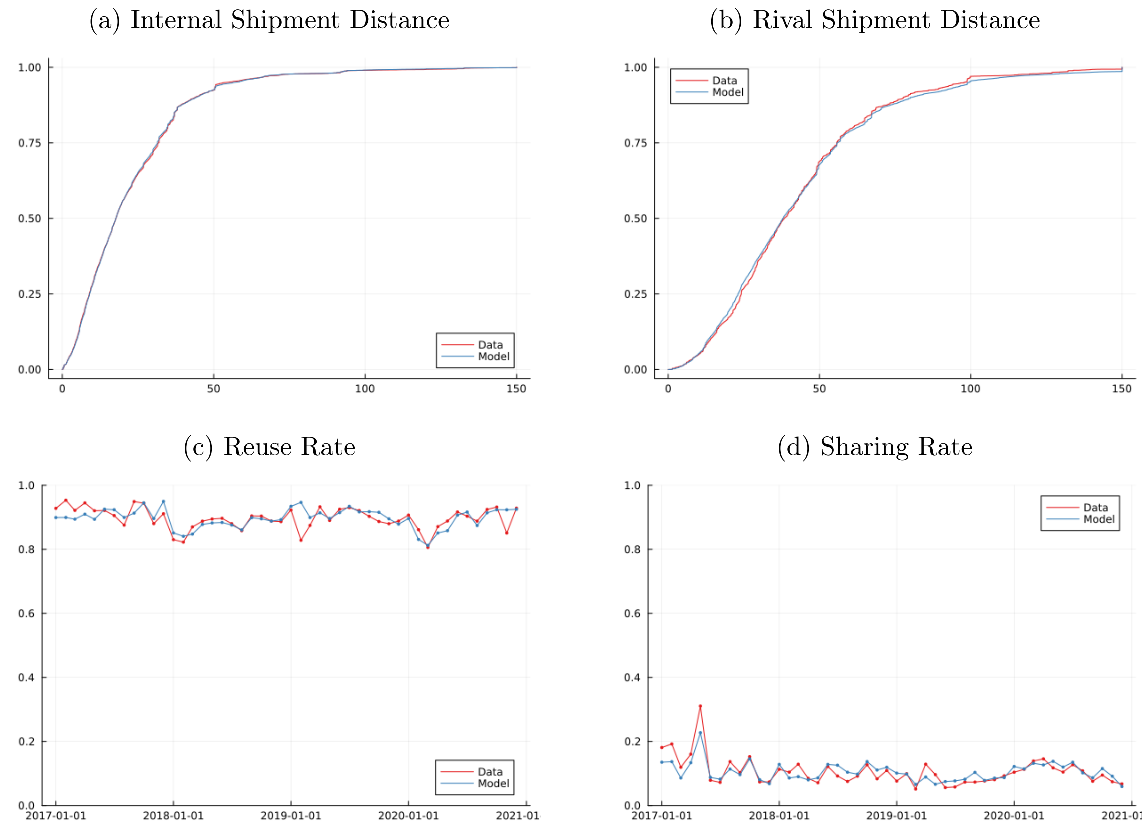
Table 6: Monthly Average Shipment Volumes at EQT/Rice Pre-Merger Well Pads (Mbbl)

	Destination, pre-merger (2017-01 through 2017-06)				
Origin	EQT	Rice	Other Rival	3rd Party CTF	Injection Well
EQT	434.4	-	46.5	40.0	-
Rice	-	173.1	-	71.9	0.1
Other	7.5	11.9			

	Destination, post-merger (2017-11 through 2020-12)				
Origin	Former EQT	Former Rice	Other Rival	3rd Party CTF	Injection Well
Former EQT	427.7	151.7	17.9	55.6	8.1
Former Rice	303.0	154.8	2.7	22.2	6.9
Other	1.1	6.0			

Notes: Monthly average shipment volumes between indicated facility types before and after the EQT-Rice merger presented in thousands of barrels (Mbbl).

Figure 5: Model Fit Results



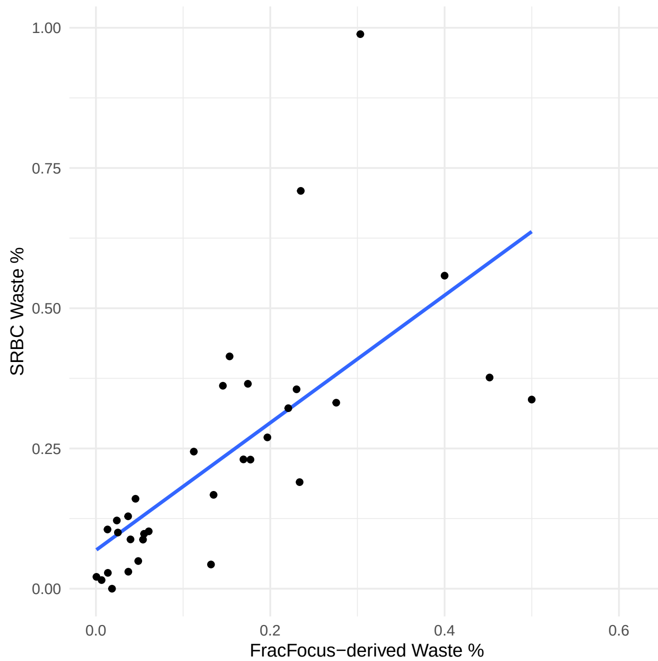
Notes: (1) Top line: observed and fitted empirical distributions of shipment distance weighted by volume. (a) includes only internal shipments (i.e., within a firm); (b) includes only shipments between rivals. (2) Bottom line: observed and fitted reuse (c) and sharing (d) rates by month. Reuse rate defined as the share of all wastewater transferred to own or rival-linked facilities and not transferred to an injection disposal well. Sharing rate defined as the share of all shipments for reuse sent to rival-linked facilities.

Table 7: Alternative Model Estimates with Firm Type Interactions

	Dependent Variable: # of Truckloads	
	(1)	(2)
Intercept	-1.526^{*}	-1.565^{*}
Distance	-0.033^{*}	-0.033^{*}
Firm Boundary	-3.145^{*}	-3.118^{*}
Firm Boundary × Injection Well Distance	-0.139	-0.179
Firm Boundary × Mutual Distance	0.089^{*}	0.060^{*}
Large regional →	0.548^{*}	
→ Large regional	-0.893^{*}	
Small regional →	0.360^{*}	
→ Small regional	0.645^{*}	
Large regional → Large regional		-0.743^{*}
Large regional → National		-0.091
Large regional → Small regional		1.564^{*}
National → Large regional		-1.175^{*}
National → Small regional		0.168
Small regional → Large regional		-0.688^{*}
Small regional → National		-0.136
Small regional → Small regional		0.501^{*}
Distance Measure	Miles	Miles
κ-month FEs	Yes	Yes
δ-month FEs	Yes	Yes
Pseudo R^2	0.802	0.802
Observations	3,843,919	3,843,919

Notes: (1) Standard errors omitted for legibility. (2) In both regressions, "National → National" is the excluded category. (3) In addition to the listed variables, each specification includes a proxy for the distance from κ to injection wells and fixed effects corresponding to the type of firm at δ (not reported). (4) Distance to injection wells is calculated at the firm level as a quantity-weighted average of well pad-month level distances to injection wells. The latter is calculated as a weighted average from the well pad to all observed injection wells. The weights are constructed as ratio of the injection well receipt quantity divided by distance (hence giving more weight to higher capacity injection wells and less to more distant ones). Finally, I take the log of this measure. (5) Mutual distance calculated as the arithmetic average of each firm's quantity-weighted average distance to a counterparty's reuse capacity. The weights are constructed as in (4). Significance levels: $^{*}p < 0.05$.

Figure 6: Predicted vs. Observed Wastewater Percentage in SRBC Sample



Notes: (1) Observations are firm-years. (2) Blue line is an OLS regression line. (3) SRBC Waste % is the average reuse rate of a firm's wells among wells appearing in the SRBC sample of well-level water usage in Northeastern Pennsylvania. (4) FracFocus-derived Waste % is described in Appendix A.1.

A Data preparation

Classification of reuse and sharing The main dataset consists of Oil and Gas Well Waste Reports collected from the Pennsylvania Department of Environmental Protection web site.[39] Operators are required to report the method of disposal for various waste products, including solids such as drill cuttings and shredded containment liners. I rely on the classifications from Wunz (2015) as well as internet research on the functions performed at different waste facilities (e.g., landfills vs. injection wells) to identify presumably reusable wastewater. This procedure is inevitably imperfect. Reporting errors are possible, and not all liquid waste necessarily represents reusable wastewater.[40]

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 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). If the destination is a well pad located in Pennsylvania, as occurs most often, a numeric identifier associated with the destination well pad is 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 reporting operator over time to track changes in corporate ownership (the Rice-EQT merger was the most significant but not the sole merger during the sample period). It is rare for multiple operators to be associated with the same well pad, but when this is the case I treat the well pad as "internal" for both parties.

Typically several wells are located at a single well pad, which encompasses common infrastructure such as access roads and storage tanks. Technically operators are required to report waste information on a well-by-well basis, but because wastewater is often stored in

[39] PA DEP (2021). Other referenced data sources include GWPC and IOGCC (2021), SRBC (2021), Ohio DNR (2024).

[40] For instance, sludges produced as a byproduct of treatment for reuse are often sent to injection wells.

a single location on the pad many 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.[41] 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 in a month).

Analysis period 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 January 2017 to consistently indicate the location of reuse.[42] Prior to January 2017, it is possible to determine whether reuse occurred, but not whether reuse occurred internally or via sharing.

Distance calculation I calculate over-the-road shipment distances using the Open Source Routing Machine (Luxen and Vetter, 2011) and data from OpenStreetMaps.[43]

Classification of large regional, national, and small firms I define "Large regional" firms to include Range, EQT, Seneca, Rice, Pennenergy, CNX, and SWN. These firms were each among the largest firms by wastewater disposal volume during the sample period, and none has significant operations outside Appalachai. Many, but not all, are privately held. I define "national" firms to include Cabot, Chief, XTO (an Exxon subsidiary), Chevron, Repsol, Hilcorp, SWEPI (a Shell subsidiary), Chesapeake, BKV, EOG, and Noble Energy. Each of these firms has a significant presence outside Appalachia. Many, but not all, are publicly held. A classify all remaining firms as "Small regional" firms.[44]

Trucking cost assumptions My preferred estimate is based on the following assumptions:

1. Trucker wages are assumed to be $22.62 per hour based on the BLS estimate of the average hourly wage of Pennsylvania truckers in 2018.

2. Average speed is assumed to be 27.8 miles per hour based on the ratio of duration and distance along calculated optimal routes for observed shipments.

[41] Line items in the data are frequently reported in integer multiples of a truck capacities ranging between 80 and 130 barrels.

[42] 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 the overall reuse rate remained stable.

[43] I do not account for roadway-specific vehicle weight or hazardous material restrictions that could alter optimal shipment routes for wastewater-hauling trucks in comparison to passenger vehicles.

[44] This designation includes some affiliates of national firms with particularly small operations.

3. Tanker-truck fuel efficiency is assumed to be 13.5 gallons of gasoline equivalent per 100 miles unladen and 15.4 gallons per 100 miles laden (assuming vehicle weights of 30,000 and 70,000 pounds, respectively) based on Bureau of Transportation Safety figures. Assuming 50% utilization this implies fuel efficiency is 6.9 miles per gallon of gasoline, equivalent to 7.9 miles per gallon of diesel.[45]

4. Diesel price is assumed to be $3.361 per gallon based on the 2018 average diesel retail price reported by EIA for PADD1B (East Coast - Mid-Atlantic).

A.1 Validation of Freshwater Usage Rates

$\hat{\mu}_{0\delta}$ is constructed as the sum of obesrved wastewater shipments divided by an estimated firm-year level wastewater consumption rate. The latter is defined as the difference between the sum of "TotalBaseFluid" for all fracking events appearing in the FracFocus data and observed disposals. A well's freshwater usage is equal to one minus its wastewater usage. Freshwater usage is observabe for a subset of wells in Northeastern Pennsylvania from the SRBC data. For firm-years appearing in the SRBC data, Figure 6 plots (i) the calculated wastewater usage rate based on the FracFocus "TotalBaseFluid" and disposal records against on the x-axis and (ii) the true wastewater usage rate in the SRBC data on the y-axis. An observation is a firm-year. A linear regression yields a coefficient 1.13 (SE 0.2) and an R^2 value of 0.49.

[45]Unlike dry van trucking, backhauls for produced water are rare due to the necessity of cleaning the tank between loads.

References

Agerton, M. (2020) "Learning Where to Drill: Drilling Decisions and Geological Quality in the Haynesville Shale," Working Paper.

Akkus, O., J.A. Cookson, and A. Hortaçsu (2016) "The Determinants of Bank Mergers: A Revealed Preference Analysis," *Management Science*, 62 (8), 2241–2258.

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

Backstrom, J. (2019) "Strategic Reporting and the Effects of Water Use in Hydraulic Fracturing on Local Groundwater Levels in Texas," Center for Growth and Opportunity at Utah State University Working Paper 2019.017.

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

Boomhower, J. (2019) "Drilling like there's no tomorrow: Bankruptcy, insurance, and environmental risk," *American Economic Review*, 109 (2), 391–426.

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

Chaney, T. (2018) "The Gravity Equation in International Trade:An Explanation," *Journal of Political Economy*, 126 (1), 150–177.

Chiappori, P., B. Salanié, and Y. Weiss (2017) "Partner Choice, Investment in Children, and the Marital College Premium," *American Economic Review*, 107 (8), 2109–2167.

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

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

Corblet, P. (2025) "Education, Sorting and Wages: A Structural Matching Approach," *Working Paper*.

Correira, S., P. Guimaraes, and T. Zylkin (2020) "Fast Poisson estimation with high-dimensional fixed effects," *The Stata Journal* (20), 95–115.

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

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

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

Fernandez-Val, I. and M. Weidner (2016) "Individual and time effects in nonlinear panel models with large N , T," *Journal of Econometrics* (192), 291–312.

Fox, J. (2018) "Estimating matching games with transfers," *Quantitative Economics*, 8 (1), 1–38.

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

——— (2024) "Estimating separable matching models," *Journal of Applied Econometrics*, 30, 1021–1044.

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

Gibbons, S., S. Heblich, and C. Timmins (2021) "Market tremors: Shale gas exploration, earthquakes, and their impact on house prices," *Journal of Urban Economics*, 122, 103313.

Gillingham, K., D. Rapson, and G. Wagner (2016) "The Rebound Effect and Energy Efficiency Policy," *Review of Environmental Economics and Policy*, 10 (1), 66–88.

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

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

Gualdani, C. and S. Sinha (2023) "Partial Identification in Matching Models for the Marriage Market," *Journal of Political Economy*, 131 (5), 1109–1171.

GWPC, Ground Water Protection Council and Interstate Oil and Gas Compact Commission IOGCC (2021) "FracFocus," https://www.fracfocusdata.org (accessed January 23, 2021).

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

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

Herrnstadt, E., R. Kellogg, and Lewis E. (2024) "Drilling Deadlines and Oil and Gas Development," *Econometrica*, 92 (1), 29–60.

Hill, E. and L. Ma (2022) "Drinkingwater, fracking, and infant health," *Journal of Health*

Economics, 82, 102595.

Hitaj, C., A.J. Boslett, and J.G. Weber (2020) "Fracking, farming, and water," *Energy Policy*, 146, 111799.

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

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

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

Jackson, R., A. Vengosh, J. Carey, R. Davies, T. Darrah, F. O'Sullivan, and G. Pétron (2014) "The Environmental Costs and Benefits of Fracking," *Annual Review of Environment and Resources*, 39, 327–362.

Kauermann, G. and R.J. Carroll (2001) "A Note on the Efficiency of Sandwich Covariance Matrix Estimation," *Journal of the American Statistical Association*, 96 (456), 1387–1396.

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

Koster, H.R.A. and J. van Ommeren (2015) "A shaky business: Natural gas extraction, earthquakes and house prices," *European Economic Review* (80), 120–139.

Lafontaine, F. and M. Slade (2007) "Vertical Integration and Firm Boundaries: The Evidence," *Journal of Economic Literature*, 45, 629–685.

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

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

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

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

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

Muehlenbachs, L., E. Spiller, and C. Timmins (2015) “The Housing Market Impacts of Shale Gas Development,” *American Economic Review*, 105 (12), 3633–3659.

Newell, R., B. Prest, and A. Vissing (2019) “Trophy Hunting versus Manufacturing Energy: The Price Responsiveness of Shale Gas,” *Journal of the Association of Environmental and Resource Economists*, 6 (2), 391–431.

Ohio DNR, Ohio Department of Natural Resources (2024) “Ohio Earthquake Database,” https://ohiodnr.gov/discover-and-learn/land-water/earthquakes/ohio-earthquake-database?=undefined (accessed November 17, 2024).

PA DEP, Pennsylvania Department of Environmental Protection (2021) “Oil and Gas Waste Reports,” https://greenport.pa.gov/ReportExtracts/OG/OilGasWellWasteReport (accessed May 11, 2021).

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

Santos Silva, J.M.C. and S. Tenreyro (2022) “The Log of Gravity at 15,” *Portuguese Economic Journal* (21), 423–437.

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

Small, M.J. et al. (2014) “Risks and Risk Governance in Unconventional Shale Gas Development,” *Environmental Science and Technology*, 48 (15), pp. 8289–8297.

SRBC, Susquehanna River Basin Commission (2021) “Post-Hydro Frac Data.”

Steck, A. (2022) “Industry Dynamics with Social Learning: Evidence from Hydraulic Fracturing,” Working Paper.

Torres, L., O.P. Yadav, and E. Khan (2016) “A review on risk assessment techniques for hydraulic fracturing water and produced water management implemented in onshore unconventional oil and gas production,” *Science of the Total Environment*, 539, 478–493.

Weingarten, M., S. Ge, J.W. Godt, B.A. Bekins, and J.L. Rubinstein (2015) “Induced Seismicity. High-rate injection is associated with the increase in U.S. mid-continent seismicity,” *Science*, 348 (6241), 1336–40.

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

Zylkin, T. (2024) “Bootstrap for Gravity Models,” *Working Paper*.