Essays on Long-Term Relationships and Networks

By

Thi Mai Anh Nguyen

Bachelor of Economics
The University of Queensland (2016)

Submitted to the Department of Economics
in partial fulfillment of the requirements for the Degree of

DOCTOR OF PHILOSOPHY IN ECONOMICS

at the

MASSACHUSETTS INSTITUTE OF TECHNOLOGY

JUNE 2023

Authored by: Thi Mai Anh Nguyen
Department of Economics
May 12, 2023

Certified by: Glenn Ellison
Gregory K. Palm (1970) Professor of Economics; Department Head
Thesis supervisor

Certified by: Tobias Salz
Castle Krob Career Development Assistant Professor of Economics
Thesis supervisor

Accepted by: Abhijit Banerjee
Ford Foundation International Professor of Economics
Chairman, Departmental Committee on Graduate Studies

Essays on Long-Term Relationships and Networks

By

Thi Mai Anh Nguyen

Submitted to the Department of Economics on May 12, 2023
in partial fulfillment of the requirements for the Degree of

DOCTOR OF PHILOSOPHY IN ECONOMICS

ABSTRACT

This thesis comprises three chapters on long-term relationships and networks. The first and second chapters study long-term relationships between shippers and carriers in the US truckload freight industry. The third chapter studies properties of learning and information aggregation on social networks.

The first chapter, joint with Adam Harris, provides evidence on the scope and incentive mechanisms of long-term relationships in the US truckload freight industry. In this setting, shippers and carriers engage in repeated interactions under fixed-rate contracts that leave scope for inefficient opportunism. We show that shippers use the threat of relationship termination to deter carriers from short-term opportunism. Carriers respond to the resultant dynamic incentives, behaving more cooperatively when their potential future rents are higher. While shippers and carriers often interact on multiple lanes, we find evidence that shippers' incentive schemes do not take advantage of this multi-lane scope for certain classes of carriers.

The second chapter, joint with Adam Harris, builds on the first, exploring a market-level tradeoff that informal long-term relationships present. On the one hand, relationships capitalize on match-specific efficiency gains and mitigating incentive problems. On the other hand, the prevalence of long-term relationships can also lead to thinner, less efficient spot markets. We develop an empirical framework to quantify the market-level tradeoff between long-term relationships and the spot market. We apply this framework to an economically important setting—the US truckload freight industry—exploiting detailed transaction-level data for estimation. At the relationship level, we find that long-term relationships have large intrinsic benefits over spot transactions. At the market level, we find a strong link between the thickness and the efficiency of the spot market. Overall, the current institution performs

fairly well against our first-best benchmarks, achieving 44% of the relationship-level first-best surplus and even more of the market-level first-best surplus. The findings motivate two counterfactuals: (i) a centralized spot market for optimal spot market efficiency and (ii) index pricing for optimal gains from individual long-term relationships. The former results in substantial welfare loss, and the latter leads to welfare gains during periods of high demand.

The third chapter proposes a novel learning model on social networks that captures settings where individuals interact frequently on multiple, relatively short-lived topics. In this model, each period features a new draw of nature and multiple rounds in which information arrives, gets aggregated, and diffuses through network links. The repetitive nature of interactions across periods allows for a separation between learning about the environment and aggregating information about the current state. A class of empiricist learning rules achieve convergence of learning on all networks. On clique trees, these learning rules further achieve strong efficiency in information aggregation. The paper also presents a converse to the positive efficiency result and identifies distinct reasons why efficiency is hard to obtain in general circumstances, even though convergence of learning holds generally.

Thesis supervisor:	Glenn Ellison
Title:	Gregory K. Palm (1970) Professor of Economics; Department Head
Thesis supervisor:	Tobias Salz
Title:	Castle Krob Career Development Assistant Professor of Economics
Thesis supervisor:	Stephen Morris
Title:	Peter A. Diamond Professor of Economics

Acknowledgements

I could not have completed this dissertation without the guidance of my advisors: Glenn Ellison, Tobias Salz, and Stephen Morris. Glenn has always been incredibly generous with his time. Week after week, I learned from him how to separate the good ideas from the bad, dig deeper into my analysis with both theory and empirics, and communicate my work effectively. With his thoughtfulness in research and his genuine care for everyone in the department, he sets an inspiring example of the kind of researcher, teacher, and colleague I will strive to be. I owe much to Tobias for helping me forge my path as an economist. In working with him, I learned how to do empirical work and, with his encouragement, gradually found my confidence. Stephen pushed me to be rigorous while constantly reminding me never to lose sight of the big picture.

In addition to my advisors, I wish to thank Bob Gibbons and Mike Whinston for taking an interest in my line of work with Adam Harris on long-term relationships. We benefited a great deal from the breadth and depth of their knowledge on this topic. I would also like to thank Nikhil Agarwal, Sara Fisher Ellison, Nancy Rose, and the rest of the MIT Industrial Organization faculty, for their warmth and generosity in giving me feedback at every stage of my research.

I am grateful for my coauthors, Adam Harris, Rebekah Dix and Kelsey Moran. From them, I have learned more than I could have imagined. Needless to say, this dissertation could not be completed without Adam, with whom I wrote the first two chapters. But what I am even more grateful for is how our coauthorship has cemented our friendship. Rebekah and Kelsey are my partners in crime, and their brilliance is a constant source of inspiration. I am excited for their journeys and wish them the best of luck.

Graduate school has its ups and downs, and I could not have done it without the best companions. I would like to thank my officemates Alden Cheng, Aaron Goodman, Garima Sharma, John Sturm, and André Sztutman, who have demonstrated great patience with my randomly-arriving and often-missing-the-mark jokes. I also find myself very lucky to know Oğuzhan Çelebi, Jon Cohen, Giacomo Lanzani, Diana Sverdlin Lisker, Kramer Quist, Claire Lazar Reich, and Maggie Yellen. Alek Razdan's care and support carried me through the most chaotic times.

I dedicate this dissertation to my family. I wish to thank my dad for planting the seed of curiosity within me that has grown into this research career and my mom for always believing in me and being the strong and amazing woman she is. Thanks to my brother for always being my devil's advocate. Sometimes I am the devil that he advocates for. None of what I have achieved would have been possible without my family's love and unyielding support. My most heartfelt thanks goes to them.

Lastly, a cheesy line from "Wicked" to sum it up,

"Because I knew you, I have been changed for good."

I truly believe that each of you helped shape the person I am today. And for that, I am forever grateful.

Table of contents

List of figures

List of tables

Chapter 1

Long-Term Relationships in the US Truckload Freight Industry*

Adam Harris† Thi Mai Anh Nguyen‡

The importance and ubiquity of informal interfirm relationships is widely recognized. As the economics, management, and sociology literatures have documented, where contracts do not exist or are incomplete, interfirm relationships are governed by nebulous notions of goodwill, trust, and reciprocity.[1] A wide range of theoretical work has elucidated various reasons why such informal arrangements might exist, what form they might take, and how they might be sustained.[2] A budding empirical literature studies these relationships. This paper contributes to that empirical literature by studying long-term relationships in the US truckload freight industry. The central role of informal long-term relationships, along with the existence of detailed microdata, makes this setting particularly well-suited for studying these relationships.[3] Exploiting this microdata, we ask (1) what is the mechanism governing these relationships, and (2) what is the scope—both temporal and spatial—of that mechanism.

*This research was made possible by the many people in the truckload freight industry who generously shared their time, insights, and data with us. We are especially grateful to Steve Raetz of CH Robinson—whose expertise we refer to throughout the paper—for both the TMC data set, as well as his invaluable feedback on our work. We are also very grateful to Angi Acocella, Chris Caplice, Glenn Ellison, Bob Gibbons, Tom Hubbard, Stephen Morris, Nancy Rose, Tobias Salz, and Mike Whinston for their advice and comments. We also thank all participants in the MIT Industrial Organization Lunch, the MIT Organizational Economics Lunch, the MIT Freight Lab, and the IIOC Rising Stars session for their comments and feedback. We acknowledge the support of the George and Obie Shultz Fund. This material is also based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1745302. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

†Corresponding author. MIT Department of Economics. Email address: asharris@mit.edu.

‡MIT Department of Economics. Email address: anhng@mit.edu.

[1] Early examples include Macaulay (1963).

[2] For an overview, see Malcomson (2010).

[3] About 80% of total industry volume is arranged through long-term shipper-carrier relationships, which, as we will describe in Section 1, are largely informal.

We answer these questions using transaction-level data from a transportation management system (TMS) used by shippers to manage their relationships with carriers. The data records every interaction within these relationships, as well as the requests for proposals (RFPs) through which relationships are formed. Moreover, it provides explicit rankings of the carriers entered into the TMS by the shipper. These rankings summarize both the status of the relationship and the shipper's intended play. The observability of shipper's play and carriers' responses enables us to shed light on both the mechanism governing relationships and the scope of that mechanism.

We first use this data to establish two key facts about the interactions between shippers and carriers. First, the spot market creates a temptation for carrier deviation; however, carriers do not behave as opportunistically as we might expect if they were playing static best response. Second, shippers control relationship termination, a power that can potentially be used to punish carrier opportunism. This hypothesized punishment mechanism could explain carriers' apparent resistance to opportunistic behavior.

Next, we propose a parsimonious model consistent with these facts in which a punishment mechanism governs the shipper-carrier relationship. From the model, we derive testable predictions about qualitative features of the optimal incentive scheme and carriers' dynamic responses. While the model is intentionally minimal, it serves an important role: guiding our approach to econometric challenges and possible alternative mechanisms in our empirical analysis.

Building on the model, we return to our main questions, empirically testing and quantifying (1) the incentive mechanisms and (2) their temporal and spatial scope.

First, to assess the question of *temporal scope*, we examine carrier behavior in the contract's final weeks. We use mass RFP events as a plausibly exogenous source of relationship termination. We find evidence of endgame effects, with carriers reducing their tendency to accept loads by 10-18 percentage points after learning of the contract period's imminent end. We argue that these findings indicate that carriers are highly responsive to dynamic incentives before the contract's final weeks and that the incentive mechanism's temporal scope is limited to within rather than across contract periods.

Second, to assess the question of *spatial scope*, we quantify the degree to which relationship status on one lane is conditioned on the carrier's performance on the same lane versus on other lanes.[4] As we would expect carriers' cost structure to differ based on their size and asset ownership, we naturally estimate shippers' incentive schemes separately for different carrier types: large asset-based carriers, small asset-based carriers, and brokers. This both provides a more nuanced answer to the question of scope and allows us to speak to the mech-

[4]The term "lane" refers to an origin-destination pair.

anisms at play. We find that large asset-based carriers face a harsh single-lane punishment scheme; a rejection on one lane can reduce the expected duration of the relationship on that lane by up to five loads. Brokers, in contrast, face a multi-lane punishment scheme.

Third, we quantify carriers' responses to the resultant dynamic incentives by estimating how carrier acceptance responds to own-lane volume, an exogenous proxy for the continuation value of the relationship.[5] The ordering of carrier types by their estimated responses to own-lane volume aligns with the strength of the estimated punishment schemes. Large asset-based carriers are most responsive to volume: doubling volume increases their acceptance probability by 6.2pp. As this would not be true for a carrier playing static best-response, we take this as strong additional evidence that, as suggested by our finding of endgame effects, carriers respond to dynamic incentives.

Finally, having presented empirical evidence consistent with a punishment mechanism, we discuss the potential role of learning. The fact that demotions tend to be permanent in our data refutes learning as the sole mechanism underlying the observed dynamics of shipper-carrier interactions. However, a combination of learning and network adjustments could explain patterns within relationships involving small asset-based carriers, who show less flexibility in adjusting their network of truck movements than do large asset-based carriers and brokers.

Our paper relates to the empirical literature on long-term informal relationships. While this literature is relatively recent, the last two decades have seen the development of a rich body of empirical evidence on the nature and value of these relationships.[6] Different dynamic mechanisms have been explored, including dynamic enforcement (Brugues, 2023), reputation and learning (Macchiavello and Morjaria, 2015), and adaptation (Barron et al., 2020; Gil et al., 2021). The value and effects of these mechanisms can be quantified by exploiting exogenous variation in spot rates (Macchiavello and Morjaria, 2015) or changing prospects of future interactions (Gil and Marion, 2013). Our paper builds on both the conceptual and methodological insights of this literature to study shipper-carrier relationships in the US trucking industry. While most of our empirical evidence points toward dynamic enforcement, we also find suggestive evidence that, depending on carriers' size and asset ownership, other mechanisms may be at play.

Our paper also relates to the literature on the trucking industry. Early papers in this literature include Rose (1985, 1987), Hubbard (2001), Baker and Hubbard (2003, 2004), and Masten (2009). Since these papers, technological improvements have generated rich

[5]This use of volume as a proxy for the expected future relationship value is in keeping with Gil and Marion (2013).

[6]For a review of recent work, see Macchiavello and Morjaria (2023).

transaction-level data on shipper-carrier relationships. Such data has been used in the transportation and logistics literature to analyze carriers' load acceptance, a key measure of performance in the trucking industry.[7] For example, Scott et al. (2017) find that carriers' tendency to accept offers within relationships is positively correlated with the volume and consistency of timing of these offers. Using the same data set as us, Acocella et al. (2020) find that, when shippers maintain high rates during a market downturn, carriers do not reciprocate with higher acceptance during a later market upturn. To the best of our knowledge, our paper is the first to dissect the dynamic mechanisms underlying shipper-carrier relationships.

Relative to these literatures, we make two main contributions. *First*, we study long-term informal relationships in an important yet understudied industry, the US truckload freight industry. Revenue in this industry was $700 billion in 2015, equivalent to about 4% of US GDP. Moreover, two aspects of this industry—(i) *fixed-rate contracts* and (ii) *on-path termination*—differentiate it from other settings in which relational contracts have been studied.[8] On the one hand, the lack of flexible monetary transfers prevents us from using workhorse models of long-term informal relationships (MacLeod and Malcomson, 1989; Baker et al., 2002; Levin, 2003), which rely on relational bonus to support optimal stationary contracts.[9] On the other hand, the fact that relationship termination occurs on-path in our setting allows us to directly estimate the incentive contract, which is nonstationary. Our *second* contribution is that, by exploiting a unique data opportunity, we directly test widely held assumptions in the literature on long-term informal relationships. In particular, we show empirically that relationships do not necessarily exist at the firm-to-firm level as predicted by the multi-market contact literature (Bernheim and Whinston, 1990) and assumed in other empirical studies (Gil et al., 2021). To our knowledge, our paper is the first to test the assumption of relationship scope. Since the relationships studied previously do not feature on-path termination, it would not be possible to test whether incentive power is pooled across all relationship sub-parts (e.g. products or markets). The fact that we observe key aspects of the relationship—its status, the agent's performance, the agent's outside option, and the firm's termination strategy, at a sub-relationship level (in our case, the lane level) permits us to perform such a test.

[7] For an excellent review of this work, see Acocella and Caplice (2023).

[8] Rationalizing these unique features of the truckload freight setting is beyond the scope of our paper. We will instead take these features as given, allowing us to focus on other aspects of the incentive contracts.

[9] Similarly, in long-term relationships where one side has full commitment power, a characterization of the optimal dynamic contracts typically uses a first-order-condition approach that relies on flexible monetary transfers. For example, Brugues (2023) applies Pavan et al. (2014) to characterize dynamic non-linear contracts with limited enforcement between sellers and buyers in the Ecuadorian manufacturing supply chain.

The paper proceeds as follows. Section 1 describes the US for-hire truckload freight industry, its market institutions, and the structures and norms within which long-term shipper-carrier relationships operate. Section 2 describes our data. Section 3 establishes two key facts that motivate the model we present in Section 4—a repeated principal-agent game in which the shipper uses an incentive contract to deter the carrier's opportunism. Guided by the testable predictions of this model, Section 5 empirically quantifies the strength and scope of dynamic incentives within shipper-carrier relationships for three different types of carriers. Section 6 discusses alternative and complementary mechanisms to that of our model. Section 7 offers conclusions about our findings, their implications, and future related research.

1 Setting

We begin by describing our setting: the US for-hire truckload freight industry. This is an economically important industry in which informal interfirm relationships play a central role. We describe the distinguishing features the industry, as well as the market institutions relevant to our analysis.

1.1 The US for-hire truckload freight industry

The freight trucking industry plays a uniquely important role in the US goods economy. In 2015, trucks carried 72% of domestic shipments by value.[10] US trucking firms had revenues of more than $700 billion in 2015, equivalent to nearly 4% of US GDP in 2015.[11]

Within the freight trucking industry, services are differentiated by the contractual relationships between shippers and carriers, by the size of shipments, and by the equipment required. In this paper, we focus on for-hire truckload carriers supplying dry-van services. We will explain each of these terms in turn: First, a *for-hire* carrier is one who sells his services to various different shippers. This is in contrast to a private-fleet carrier, who is vertically integrated with a single shipper. Second, a *truckload* carrier accepts only large shipments that fill all or nearly all of a trailer. Truckload service is "point-to-point": A truckload shipment has a single origin and a single destination. While a truckload carrier must plan his network of truck movements efficiently to minimize empty miles, his problem is far simpler than the optimization problem faced by a less-than-truckload carrier, who aggregates smaller shipments to fill the trailer. Finally, a freight truck consists of a tractor unit, which contains a heavy-duty towing engine and a driver cab, and a cargo trailer, which

[10]See Bureau of Transportation Statistics Freight Facts and Figures 2017.
[11]*American Trucking Trends*, American Trucking Association, 2015.

holds goods being hauled by the tractor. Some common trailer classes include refrigerated, flatbed and tanker. By far the most common trailer type is the *dry van*, used for hauling boxes or pallets of dry goods not requiring refrigeration. We will focus exclusively on dry van truckload services supplied by for-hire carriers. This is the largest subsegment of the trucking industry and one in which carriers' business model and logistical challenges are easy to understand.

1.2 Truckload carriers: Assest-based carriers and brokers

Carriers providing truckload services can be divided into two types: asset-based carriers and brokers. An asset-based carrier owns and operates trucks, which he uses to transport goods.[12] A broker, on the other hand, has no trucks; instead, when a broker accepts a load from a shipper, he subcontracts in a spot arrangement with an asset-based carrier to transport the goods.

Among asset-based carriers, there is enormous heterogeneity in fleet size. On one extreme, the largest US carriers each operate more than 10,000 trucks. On the other extreme, as of December 2020, there were 317,791 registered carriers operating only a single truck.[13] While these tiniest owner-operator carrier will not feature in our analysis, heterogeneity in carrier size—and thus capacity, will play an important role. Our empirical analysis will distinguish between small asset-based carriers (100 trucks or fewer) and large asset-based carriers (more than 100 trucks).

1.3 Market institutions

In the US for-hire truckload freight market, shippers and carriers arrange loads through two primary market institutions: a spot market and (largely informal) long-term relationships.

Typically, about 20% of loads are arranged through the spot market.[14] The dominant spot market platform is organized by DAT Solutions; this online "load board" is a simple post-and-search marketplace that facilitates matches of shippers with loads and carriers with trucks.

The remaining 80% of US truckload transactions are arranged through long-term relationships between shippers and carriers. While these relationships are formalized by contracts, the contracts are highly incomplete. A contract defines liability for lost or damaged goods

[12] To avoid confusion, we will, throughout the paper, refer to the shipper using she/her pronouns and the carrier using he/him pronouns.

[13] FMCSA, Motor Carrier Management Information System (MCMIS).

[14] medium.com/@sambokher/segments-of-u-s-trucking-industry-d872b5fca913

and establishes the rate the shipper will pay the carrier for each load on the lane. However, it imposes few other restrictions on the parties and does not obligate the shipper and carrier to behave cooperatively toward one another.[15] In particular, the contract does not obligate the carrier to accept any loads offered by the shipper under the terms of the contract. If the carrier rejects loads, the contract does not give the shipper any legal recourse.

The dominance of long-term relationships in this industry suggests that they offer benefits not enjoyed in spot arrangements. Such benefits could take several forms: First, shippers and carriers who interact repeatedly may benefit from a familiarity with each other's facilities and processes, which can improve functions like loading and payments. Second, arranging loads through a long-term relationship might save on costs associated with searching and haggling in a thin spot market.[16] Such costs are likely non-negligible, as demand for transportation services is dispersed across space and time. Third, and closely related, because spot-market demand on a particular lane at a particular time might be scarce, carriers may prefer the more consistent demand from contracted shippers, which facilitates a stable, cost-effective network of truck movements for the carrier.

1.4 Managing relationships: The routing guide

A shipper frequently has contracts with several different carriers on a particular lane. These various carriers, are not, however, equal in status. The shipper explicitly ranks the carriers in a catalog called the *routing guide*. This ranking specifies the order in which carriers are sequentially offered each load that the shipper has on this lane.[17][18]

[15]The fact that these agreements are informal and not legally binding is widely understood in the industry. For instance, Melton Truck Lines Senior Vice President Dan Taylor wrote "The 'rate agreements' and 'load commitments' for the most part have no contractual obligation or penalties on either party." (See Taylor (2011).)

[16]The role of transaction costs in driving the tendency towards contractual arrangements is a well established idea. For related studies in the context of the trucking industry, see Hubbard (2001) and Masten (2009).

[17]For a more detailed discussion of the routing guide and related features of truckload operations, see Section 4 of Caplice (2007).

[18]As we discuss in the next section, the process of sequentially offering loads is automated by software called a transportation management system (TMS). The TMS that allows each carrier only a short amount of time to respond to an offer. A typical response window might be fifteen minutes. This rapidity suggests that the shipper does not have a strategic incentive to rank a carrier higher just because that carrier is in high demand by other shippers; so little time passes between offers that a lower-ranked carrier is unlikely to be "snatched up" by another shipper while higher-ranked carriers are responding to their offers. This means that a shipper's static best response is to rank the carriers according to her preference over the carriers. Thus, rejections by top-ranked carriers are generally undesirable for the shipper. In Table 1, for instance, the fact that the shipper chose to rank A above B indicates that she prefers paying \$1230 for service from A to paying \$1327 for service from B. Furthermore, the fact that she ranked B above C, despite the fact that C has markedly lower contract rate, suggests that B provides the shipper with superior service in some dimension other than rate (e.g. quality, reliability). More generally, differences in non-rate characteristics

Table 1: Example routing guide: Shipper Z, lane City X - City Y (on June 1, 2018)

Order	Carrier	Rate	Type
1	A	$1230	Primary carrier
2	B	$1327	Backup carriers
3	C	$1095	
4	D	$1450	

To illustrate this sequential offering process (sometimes called a *waterfall*), Table 1 gives an example of an (anonymized) routing guide for a shipper Z on the lane from City X to City Y. When Z has a load at City X that she wants to ship to City Y on a particular date, she first offers the load to the *primary carrier*, in this case, A. If A accepts, then A carries the load and receives $1230. If A rejects, then the load is offered to B. If B rejects, the load is offered to C, and so on. If the routing guide is exhausted without a carrier accepting, the shipper will typically turn to the spot market to try to find a carrier to accept the load.

The rationale for the shipper maintaining a routing guide with multiple carriers who have the right to reject loads, rather than a single carrier for whom acceptance is obligatory, is that 100% acceptance by a single carrier is unlikely to be efficient. The demand of a shipper over time is random and, therefore, cannot be perfectly predicted by a carrier.[19] This means that when the shipper offers a load on a particular date, the carrier's trucks may be poorly positioned for carrying this load; doing so may be very costly or infeasible.[20]

There are two features of the routing guide that play an important role in the dynamics of the shipper-carrier relationship:

First, the shipper has discretion to alter the ranking of carriers in the routing guide at any time. Indeed, though Table 1 gives the ordering of carriers on June 1, 2018, the routing guide for the same lane two weeks later is substantially different, with these four carriers ordered C, A, B, D. Why might such a change occur? The shipper might use the power to reorder the routing guide strategically to incentivize carrier cooperation. If a carrier (e.g.,

may also be important to the shipper.

[19]While the timing of loads is random, the demand of a shipper is typically more consistent than the demand of a single consumer in some other transportation industries, e.g. the taxi or ride-hail industry.

[20]One might think that the GPS-equipped devices now installed on most trucks would seem to offer an opportunity for contracts that obligate carriers to accept unless some verifiable conditions are met, e.g. unless all of the carrier's trucks are more than 100 miles from the load pickup location. However, any such simple criterion on current truck locations is unlikely to be satisfactory. First, a load is typically offered and accepted several days before the load's pickup time. At the time the acceptance decision is made, only current truck locations are verifiable, but what is relevant to the carrier's ability to pick up a load is of course the location of his trucks at the pickup time. Second, any such simple condition on truck locations would fail to account for both network-related issues and the carrier's obligations to other shippers. A carrier may have a truck near the load pickup location, but that truck may be needed to fulfill an obligation to another shipper.

Carrier A) were behaving opportunistically, rejecting contract loads in favor of taking higher-paying loads in the spot market, the shipper could punish the carrier by downgrading him to a lower position in the routing guide. Being downgraded diminishes the carrier's future rents from the relationship, as he will now receive fewer offers on this lane. This possibility of punishment via reorganization of the routing guide is the mechanism at the heart of our model and empirical analysis.

Second, at the end of a contract period, the shipper holds a request for proposals (RFP) to determine the set of carriers, their rates, and their initial positions in the new routing guide. In an RFP, a shipper need not award the primary position to the lowest-bidding carrier; non-rate characteristics can be taken into account. While this is intuitively similar to a scoring auction, the way these RFPs are carried out in practice is far more complicated than the formal auctions that have been studied theoretically and empirically in a wide range of economic settings. After a shipper receives carriers' initial bids, multiple rounds of negotiation between the shipper and the various carriers jointly determine carriers' final routing guide positions and rates.

2 Data

We use transaction-level data from the transportation management system software used by shippers to manage their relationships with carriers. The data records every interaction within these relationships. To proxy for carriers' outside option, we use a measure of the going rate for freight services in the spot market from DAT, the gold-standard provider of such spot market data.

2.1 Shipper-carrier microdata

Our analysis is made possible by the fact that shippers use a transportation management system (TMS) to manage their relationships with carriers and to automate the waterfall of offers. The shipper enters carriers' rates and ranks into the TMS, and then, for each load, prompts the TMS to sequentially send electronic offers to the carriers. For each load sent through the TMS, the software records the details of the load, all offers that are made, and whether each is accepted or rejected. These records for one particular TMS software provider, called TMC, are the source of our microdata.[21]

The microdata covers the period from September 2015 through August 2019. In all, the data set includes 1,074,172 loads and 2,130,125 offers. 71% of loads are accepted by the first

[21] TMC is a division of CH Robinson, a third-party logistics firm.

carrier to which they are offered. All loads in the data set have a haul distance of at least 250 miles.[22] The mean distance is 692 miles with a standard deviation of 440 miles. The average per-mile contract rate is $1.85 with a standard deviation of $0.51.

The time between consecutive RFPs on a lane is, on average, 322 days, with an average of 83 loads being offered during this time. The carrier who wins an RFP on a lane is demoted from primary status before the next auction is held 21% of the time.

Shipper-carrier relationships and networks The microdata includes 40 shippers with at least 500 loads. The median shipper has 8,094 loads with, on average, 192 active lanes and 53 active carriers each year.

On many lanes, shippers and carriers interact infrequently. For example, the median lane of the median shipper has only one load per month. However, among the top 10% of lanes for the median shipper, each has, on average, a load for every four days. Such variation in frequency of interactions will be important for our test of carriers' response to dynamic incentives.

Multilane interactions between a shipper and a carrier are also common. The top five and top ten carriers of the median shipper deliver, respectively, 58% and 73% of her loads. Relatedly, it is common for a carrier to serve as a shipper's primary carrier on multiple lanes. For example, the top five carriers of the median shipper hold primary status on an average of 21 lanes each. There is thus significant potential for strategic exploitation of multilane interactions: a shipper might condition a carrier's primary position on one lane on his behavior on another lane. This question of the scope of the incentive mechanism—whether shippers exploit multilane interactions to create cooperative incentives—is one of the key questions this paper addresses.

2.2 Spot rate data

We will use data on the average rate for truckload services in the spot market to capture the relevant outside option—the alternative opportunities available to shippers and carriers outside of their long-term relationships. This data comes from DAT Solutions, the leading provider of data on truckload spot markets. For our sample period, the data set gives us seven-day trailing average spot rates for a set of narrowly-defined lanes that cover the continental United States.[23]

[22]For shorter-distance hauls, the prevailing market institutions are somewhat different. These loads are therefore excluded from our analysis.

[23]Each lane is defined by a pair of key market areas (or KMAs). The continental US is partitioned into 135 KMAs, so there are 135^2 KMA-to-KMA lanes.

Across all lanes and dates, the overall mean spot rate per mile is \$1.68 with a standard deviation of \$0.60. The first quartile, the median, and the third quartile are \$1.26, \$1.53, and \$1.93, respectively. A notable feature of the data is persistent differences in rates across lanes; a regression of spot rates on a set of lane fixed effects has an R^2 of 0.78, indicating that across-lane differences are large relative to within-lane variation. In later empirical analysis, we pool observations across lanes for the purpose of estimating the strategies of shippers and carriers. To make for appropriate comparisons across lanes, we will use residualized, rather than raw, spot rates, partialling out lane fixed effects. For the time series of average monthly spot rates over our sample period, see Figure 1 in the next section.

3 Two Key Facts

In this section, we use our shipper-carrier microdata, together with the data on spot rates, to establish two key facts which suggest that the shipper-carrier relationship can be thought of as a repeated principal-agent game in which an incentive mechanism deters carrier opportunism.

3.1 Fact 1: Temporary spot-contract rate differences create temptation for carriers

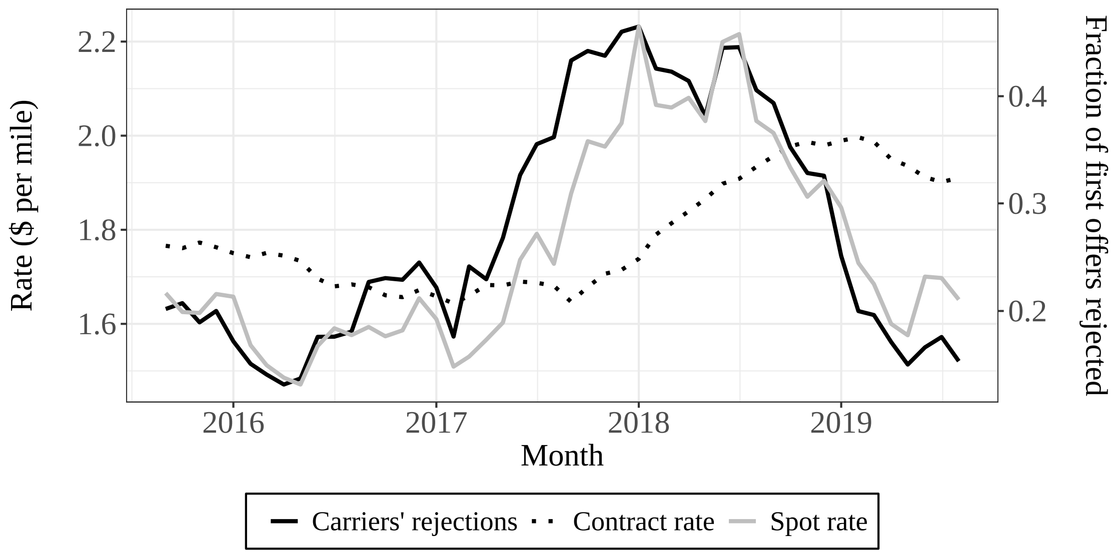
We begin by arguing that when spot rates are higher than contract rates, carriers are tempted by short-term opportunism. As such opportunism is detrimental to the shipper, a moral hazard problem exists within the shipper-carrier relationship.

Figure 1 illustrates the potential for carrier opportunism by depicting two key aggregate trends in our relationship microdata and aggregate data on spot rates. First, there are periods in which spot rates are significantly higher than contract rates. Second, these periods coincide with a high proportion of rejections by carriers.

Figure 1 shows considerably greater intertemporal variation in spot rates than in contract rates. While spot rates were generally lower than contract rates in the first two years of our sample, an aggregate demand shock in late 2017 and early 2018 resulted in a sharp increase in spot rates.[24]

[24]Contemporary articles from various trade publications (including Transportation Topics and FreightWaves.com) describe the high spot rates of the 2017-2018 period as being driven by increased spending on e-commerce, booming US industrial production, and the December 2017 corporate tax cut. Some sources also cite various supply factors, including the December 2017 introduction of a rule requiring for-hire trucks to be equipped with electronic logging devices (ELDs), though these supply factors seem to be considered of secondary importance. See, for instance, www.freightwaves.com/news/market-insight/forecasting-2019.

Figure 1: Aggregate trends: National averages of rejection, contract, and spot rates



Notes: The monthly rejection rate is constructed from the TMS microdata as the fraction of loads rejected by the first carrier in the routing guide. The average monthly contract rate is constructed from the TMS microdata. The average monthly spot rate is constructed from the DAT data, on the same set of lanes covered by the TMS microdata. The rejection rate, contract rate, and spot rate are all volume-weighted averages.

These spot market premia create the potential for short-term opportunism. Recall that a carrier in a long-term relationship always has the option to reject loads offered to him by the shipper. Thus, when the spot rate exceeds the contract rate, the carrier may reject contract loads and instead opt to provide service in the spot market. Figure 1 shows evidence consistent with this hypothesis: The period of high relative spot rates coincides with a large increase in the proportion of offers rejected by primary carriers. This observation strongly suggests that the spot market represents a key outside option for carriers.

Such opportunism by the carrier presents a moral hazard problem. The fact that long-term relationships exist in the first place—rather than all transactions being arranged through the spot market–suggests that there is relationship surplus that would be foregone were the carrier to opportunistically choose to service the spot market.[25] Furthermore, the shipper has imperfect monitoring: the shipper cannot distinguish between an inefficient opportunistic rejection and an efficient rejection resulting from the carrier's current cost of service being high.

Yet Figure 1 also gives us reason to believe that some mechanism exists to alleviate the moral hazard problem. When spot rates peak in January 2018, they are on average 20% higher than contract rates. Despite this strong incentive for carriers to reject loads, the majority of loads are still accepted by primary carriers in this month. That many carriers

[25] See Section 1.3 for a discussion of possible sources of this surplus.

are willing to forgo significant short-term profits suggests that their opportunistic tendencies are restrained by some other force.

One such force could be an incentive scheme in which the promise of future rents helps alleviate the carrier's short-term opportunism. Necessary conditions for such an incentive scheme to be effective are that (i) the shipper has the power to deny the carrier future rents if he behaves opportunistically and (ii) the carrier's future rents from the relationship are sufficiently large. The next subsection establishes the former; the latter is established in Sections 5.1 and 5.3.

Before addressing the role of shippers in the next subsection, we note an asymmetry between shippers and carriers: While carriers might be tempted to short-term opportunism by high spot rates, shippers seem not to be tempted to opportunism by low spot rates. In Appendix C, we show that shippers continue offering loads to their primary carriers even when lower-rate alternatives are available in the spot market.[26]

3.2 Fact 2: Shippers control relationship termination

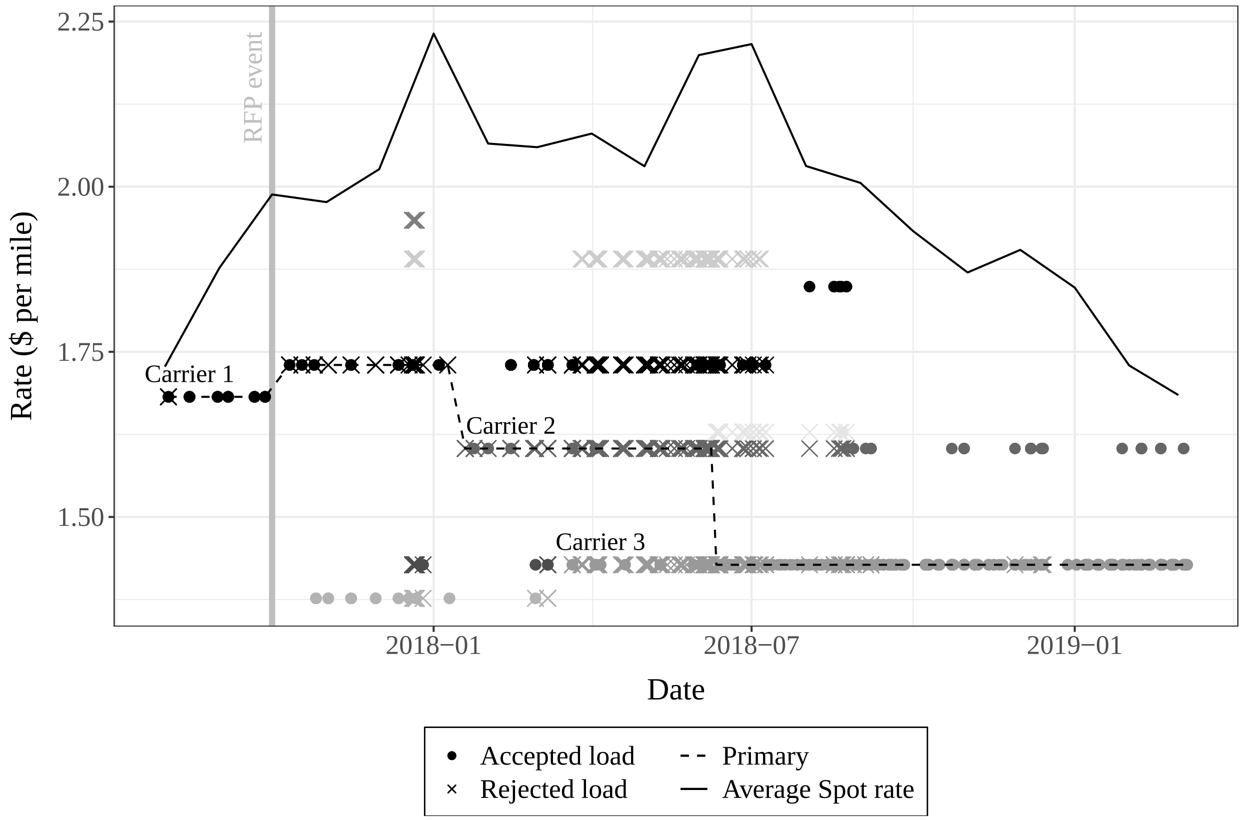
We next use our shipper-carrier microdata to show that shippers control relationship termination and provide suggestive evidence on the form of shippers' termination strategies.

Figure 2 presents an example of a lane history that motivates the way we think about the shipper's decisions. Recall that the chief decision faced by the shipper is that of when and how to change the routing guide. Such changes can be made at any time. Some are the result of RFPs, while others take place within the contract period, i.e. in the time between RFPs. Our analysis will focus on the latter and, in particular, on those changes that replace one primary carrier with another. We refer to such a change as a *demotion* of the current primary carrier.

In the example in Figure 2, Carrier 1 initially holds primary status and accepts most of the offers he receives. Around early October 2017, the shipper holds an RFP for this lane; Carrier 1 retains his primary status and gets a rate increase of about 5 cents per mile. However, over the next three months, a period of high spot rates, Carrier 1 rejects many of the loads offered to him. In January 2018, Carrier 1 is demoted from primary status and replaced by Carrier 2. Over the next five months, Carrier 2 rejects most of the loads offered to him. Ultimately, he too is demoted in favor of Carrier 3, who maintains primary status

[26] In addition, industry experts tell us that it is "very rare" for a shipper to go directly to the spot market before the routing guide when spot rates are low. Perhaps because shipping represents only a small component of the operations of shippers, who are usually non-transportation firms (e.g. manufacturers or retailers), taking advantage of short-term opportunities to reduce shipping costs is not a priority. Shippers allow day-to-day shipping decisions to be automated by the TMS and make strategic decisions only on a medium-term basis.

Figure 2: An example: Offers for Shipper X, City Y - City Z



Notes: Each point represents an offer: circles represent offers that are accepted while crosses represent offers that are rejected. Each carrier is indicated by a different color. The dotted black line indicates the rate of the primary carrier at each point in time. Carrier 1, Carrier 2 and Carrier 3 each serve as primary carrier for this shipper and lane for some subset of the period from October 2017 to May 2019. The offer data comes from the TMS microdata. The average monthly spot rate on the same lane is constructed from the DAT data set.

for the rest of the sample period.

This figure, which illustrates patterns that are common to many lanes, motivates two key conclusions about shipper-carrier relationships:

First, while shippers have almost unlimited discretion in what kind of routing guide changes they make, in practice, they do not switch primary carriers frequently; rather, a shipper maintains a primary carrier for a time before ultimately—and usually permanently—demoting that primary carrier.[27] From this observation, it seems appropriate to think of the

[27]See Table 5. We find that in more than 95% of instances where a carrier is demoted from primary status, he never regains primary status on the lane in our sample period (2015-2019). While there is a truncation issue here (a demoted carrier may regain primary status after the end of our sample period), it nevertheless seems clear that demotions are typically permanent. While it is possible that shippers are actually employing reward-punishment cycles like those described by Green and Porter (1984) or Li and Matouschek (2013), the cycles would have to be very long. We observe four years of data—quite a long period of time when one considers that the typical time between consecutive loads on a lane is on the order of a few days—, and yet we almost never observe a demoted carrier regaining primary status. Hereafter, we assume that demotion is permanent and use the terms demotion and termination interchangeably.

shipper-carrier relationship in terms of the following kind of principal-agent model: the shipper controls relationship termination and, at each point in time, decides between continuation and (permanent) termination.

Second, a clear pattern on this lane is that a series of rejections by the primary carrier often is followed by a demotion. This pattern is documented more systematically in Section 5.2. This evidence is consistent with an incentive mechanism where the shipper generates dynamic incentives for the carrier by conditioning relationship continuation on acceptance.

The evidence in this subsection indicates that shippers have the power to terminate relationships. Whether the threat of termination is effective in deterring carrier opportunism will be addressed in Sections 5.1 and 5.3, both of which show strong evidence of carriers responding to dynamic incentives.

4 A model of the incentive contract

In this section, we develop a model of long-term shipper-carrier relationships that will serve as a theoretical framework for our empirical analysis. For tractability, we focus on the relationship between a shipper and a primary carrier, abstracting away from the existence of backup carriers. Features of our model are motivated by the two facts established in Section 3:

Fact 1. Temporary spot-contract rate differences create deviating temptation for carriers,

Fact 2. Shippers control relationship termination and can use this power to generate an incentive scheme.

The model also generates testable predictions for our empirical analysis.

Relationship parameters A tuple $(\psi, \eta_1, \eta_2, p, \delta, F, G)$ summarizes the key characteristics of a relationship. Here, ψ is the relationship-specific gain to the shipper from transacting with the carrier; $\eta = \eta_1 + \eta_2$ is the relationship-specific gain to the carrier from transacting with the shipper. Some of these gains are publicly observed (η_1), such as the consistency of timing of shippers' requests, which helps the carrier's planning. Some are privately observed (ψ, η_2), such as the quality of on-road communication or efficiency of loading and docking. In addition, p is the contract rate; $\delta \in (0, 1)$ is the discount factor, measuring the frequency of shipments.[28] F is the distribution of the carrier's cost of servicing a shipment on the contracted lane; and G is the distribution of the spot rate on that lane.

[28]An underlying assumption is that the shipper's demand is perfectly inelastic with respect to spot rates. Appendix C shows evidence that shippers do not reduce load offers within the routing guide when spot rates are low.

Let $\tilde{p}_t$ and c_t denote, respectively, the spot rate and the carrier's cost draw in period t. The shipper's period-t payoff is $u_t = \psi - p$ if she is served by the contracted carrier and $u_t = -\tilde{p}_t$ if she is served by the spot market. The carrier's period-t payoff is $v_t = \eta + p - c_t$ when serving the contracted shipper and $v_t = \tilde{p}_t - c_t$ when serving the spot market (on the same lane). If the carrier chooses to remain idle (or serve a different lane), he gets zero payoff in that period. That is, c_t captures the opportunity cost of servicing the contracted lane in period t. This means that the cost distribution F captures both the contracted lane's average alignment with the rest of the carrier's network and the day-to-day variation of such alignment.

When $\psi + \eta > 0$, shipments fulfilled within relationships generate surplus over spot transactions. In this case, it is never efficient for the carrier to reject the shipper's offer in order to serve the spot market on the same lane. However, requiring the carrier to always accept the shipper's load is also not efficient, since the carrier's (opportunity) cost in some periods might be very high.[29] The inability of the shipper to distinguish between rejections due to high cost draws and rejections due to high spot rates represents a potential source of inefficiency in this setting, one that the shipper may hope to alleviate using the threat of relationship termination.

Timing and informational assumptions In period $t = 0$, the shipper holds an RFP to select a primary carrier. This RFP also reveals to both the shipper and carrier the characteristics $(\psi, \eta_1, \eta_2, p, \delta, F, G)$ of their relationships.[30] From period 1 onward, the shipper and the primary carrier interact repeatedly, with spot rates and costs being drawn independently and identically over time.[31] The stage game is summarized in Figure 3. In each period t, a spot rate $\tilde{p}_t$ is drawn from G and publicly observed, and a cost draw c_t is drawn from F and privately observed by the carrier. The shipper then decides whether to "keep" the carrier as the primary carrier or "end" their relationship. If the relationship is maintained, the carrier chooses whether to accept (A) or reject (R) the shipper's load in that period. If he rejects, then he can either serve the spot market or remain idle. Otherwise, if the relationship ends, both sides resort to the spot market for future transactions; the shipper gets expected payoff

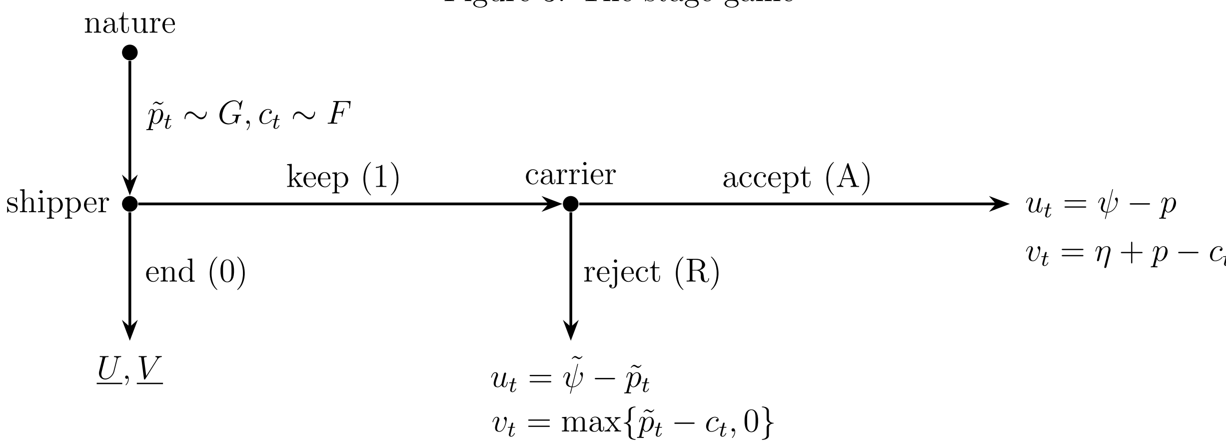
[29] Specifically, when $c_t > (\psi + \eta) + \tilde{p}_t$, the efficient outcome is that the carrier remains idle or serves a different lane.

[30] See Harris and Nguyen (2023) for modeling assumptions under which relationship characteristics are revealed via the auction.

[31] Realistically, spot rates exhibit autocorrelation (Figure 1). With autocorrelation, we would expect an increase in the spot rate to have several different effects, as it would change the continuation value of both the shipper and the carrier in addition to increasing contemporaneous temptation for the carrier. Harris and Nguyen (2023) present a model in which spot rates are AR(1) and the shipper keeps track of a rejection index summarizing all past rejections. The qualitative properties of our predictions extend to this more general model.

of $\underline{U}$ and the carrier gets expected payoff of $\underline{V}$.

Figure 3: The stage game



To generate clean predictions, we focus on the simplest class of shipper's incentive schemes, those that condition only on the carrier's decision in the last period. Denote such an incentive scheme by $\sigma_0 : \{A, R\} \to [0,1]$, where $\sigma_0(d_{t-1})$ is the probability that the shipper maintains the relationship following decision d_{t-1} of the carrier in the last period. We interpret $\sigma_0(A)$ as the level of rewards following cooperation and $1 - \sigma_0(R)$ the level of punishment following noncooperation. We examine the carrier's optimal response to such scheme. Since the shipper conditions only on the last period's decision, we can focus on the carrier's stationary play, $\sigma_1 : \text{supp}(G) \times \text{supp}(F) \to [0,1]$, where $\sigma_1(\tilde{p}_t, c_t)$ is the probability that the carrier accepts the offered load given spot rate $\tilde{p}_t$ and cost draw c_t.

Model predictions Next, we derive testable predictions on the carriers' optimal stationary play and shipper's incentive scheme.

Proposition 1. *(Carrier acceptance) Suppose that $\eta + p \in Supp(G)$.*[32] *The carrier's optimal response takes a threshold form: accept if and only if $\bar{p} \geq \max\{\tilde{p}_t, c_t\}$, where*

$$\overline{p} = \eta + p + \frac{\delta}{1-\delta}(V(A) - V(R)),$$

and $V(d_t)$ is the carrier's expected payoff following $d_t \in \{A, R\}$. Moreover, the following comparative statics hold:

i) (Dynamic incentives) If $\sigma_0(A) > \sigma_0(R)$, then $\overline{p} > \eta + p$. That is, the carrier accepts more often than their static best response.

[32] That is, there are periods in which spot transactions are more attractive than the contracted offer and periods in which the contracted offer is more attractive than spot transactions.

ii) For a fixed incentive scheme with $\sigma_0(A) > \sigma_0(R)$ and unobserved characteristics (ψ, η_2, F),

$$\frac{\partial \overline{p}}{\partial \delta} \geq 0, \quad \frac{\partial \overline{p}}{\partial \eta} \geq 1, \quad \text{and} \quad \frac{\partial^2 \overline{p}}{\partial \delta \partial \eta} \geq 0.$$

iii) For fixed carriers' parameters (η_1, η_2, F) and δ, and for every $\sigma_0(A) > \sigma_0(R)$,

$$\frac{\partial \overline{p}}{\partial \sigma_0(A)} \geq 0 \quad \text{and} \frac{\partial \overline{p}}{\partial \sigma_0(R)} \leq 0.$$

Proof. See Appendix A.1. □

Intuitively, the shipper can use her control over relationship continuation as an incentive scheme to induce dynamic incentives for the carrier to accept more loads. The strength of such dynamic incentives depends on both the value of current and future loads to the carrier, as well as the levels of rewards and punishments induced by the shipper's incentive scheme.

Proposition 2. *(Shipper's single-lane incentive scheme) Suppose that shipper's match-specific gain is sufficiently large, $\psi \geq p - \mathbf{E}[\tilde{p}_t | \tilde{p}_t \leq p]$. The optimal single-lane incentive scheme for the shipper satisfies*

i) (Maximum rewards) $\sigma_0^*(A) = 1$ for any parameter values of $(\psi, \eta_1, \eta_2, \delta, F, G)$,

ii) (Soft punishment) $\sigma_0^*(R) \in (0, 1)$ for some parameter values of $(\psi, \eta_1, \eta_2, p, \delta, F, G)$.

Proof. See Appendix A.2.1. □

The incentive scheme affects the shipper's payoffs via two channels: the effect on carrier's acceptance probability (the incentive-inducing effect) and the direct effect on the probability of ending the relationship (the regime-switching effect). Since the shipper faces no tradeoff between these two effects when deciding on the reward, the maximum reward is guaranteed relationship continuation. In contrast, harsher punishment increases the carrier's acceptance probability but also increases the likelihood of relationship termination when the carrier rejects. It is possible that this tradeoff is not resolved by extreme punishment, but rather by soft punishment.

While Proposition 2 assumes that the shipper's incentive scheme operates at the lane level, Example 1 considers the possibility of a broader scope, operating across multiple lanes within the shipper-carrier relationship. Importantly, part (ii) of this example shows that using a simple scheme to pool incentives across heterogeneous lanes can backfire, making the shipper worse off than if she just using the optimal single-lane incentive schemes.

Example 1. Suppose that the shipper and the carrier interact on two lanes $\ell = 1, 2$, each characterized by $(\psi, \eta_1^\ell, \eta_2^\ell, p^\ell, F, G)$. Let $F \sim \alpha U(0,1) + (1-\alpha)\delta_K$ for some large K. That is, a cost draw is with probability α distributed as a standard uniform random variable and with probability $(1-\alpha)$ equal to some $K \gg 1$. Let $G \sim U(0,1), \alpha = 0.75, \delta = 0.9$, and $\psi = 0.3, p = 0.6$. Focus on multi-lane incentive schemes that map the average multi-lane rejections $\frac{1}{2}\sum_{\ell=1}^{2} \mathbf{1}\{d_{t-1} = R\}$ of the carrier in the last period to a probability of the shipper keeping the relationship (on both lanes). Denote by $\hat{\sigma}_0$ the shipper's optimal multi-lane incentive scheme and by $\sigma_0^\ell : \{A, R\} \to [0,1]$ the shipper's optimal single-lane incentive scheme for $\ell = 1, 2$.

i) If $\eta_1^1 + \eta_2^1 + p^1 = \eta_1^2 + \eta_2^2 + p^2 = 0.65$, the shipper is better off using $\hat{\sigma}_0$ than (σ_0^1, σ_0^2).

ii) If $\eta_1^1 + \eta_2^1 + p^1 = 0.65$ and $\eta_1^2 + \eta_2^2 + p^2 = 0.8$, the shipper is worse off using $\hat{\sigma}_0$ than (σ_0^1, σ_0^2).

Proof. See Appendix A.2.2. □

Empirical challenges In testing the predictions generated by this model, the econometrician only observes a component of carrier's gain (η_1), the contract rate (p), the frequency of interactions (δ), the realized spot rate ($\tilde{p}_t$) and the carrier's decision (d_t) in each period. Our empirical approaches focus on addressing endogeneity issues due to the unobservability of other relationship-specific characteristics (ψ, η_2) and carrier's cost distribution F. We will focus on addressing three channels: First, these unobservable characteristics affect the carrier's decisions (d_t). Second, they affect the contract rate (p) via the RFP process. Third, they potentially correlate with the frequency of interactions (δ). The last of these objects plays an important role in our empirical analysis, serving as a potential shifter for the continuation value of the relationship.

5 Empirical Evidence

In this section, we present empirical evidence on two key questions. First, what is the scope—both spatial and temporal—of the incentive mechanisms that govern shipper-carrier relationships? And, second, how do carriers respond to the dynamic incentives created by these mechanisms?

Our analysis of these questions proceeds in three subsections. In Section 5.1, we begin by presenting empirical evidence on the temporal scope of the relationship using carrier acceptance decisions in the weeks preceding the end of the contract period. This exercise also provides preliminary evidence that carriers respond to dynamic incentives. Next, in Section

5.2, we present empirical evidence on the scope of the relationship in the spatial dimension by estimating shippers' demotion strategies. Finally, in Section 5.3, we provide further evidence that carriers respond to dynamic incentives using carrier behavior throughout the relationship. In the latter two subsections, our analysis accounts for an additional dimension of carrier heterogeneity, analyzing behavior separately for three different types of carriers—large asset-based carriers, small asset-based carriers, and brokers.[33]

5.1 Empirical Evidence: Carrier behavior at the end of the contract period

This subsection presents evidence that (i) carriers respond to dynamic incentives and (ii) the scope of the incentive scheme carriers face is within, not across, contract periods.

We showed in Proposition 1(i) that a carrier facing dynamic incentives has $\bar{p} > p + \eta$, meaning that he is *ceteris paribus* more likely to accept load offers than a carrier with only static incentives. If an exogenous event were to unexpectedly eliminate the carrier's dynamic incentive, then we would expect a decrease in his probability of acceptance. In our data, *mass RFP events* act as a natural experiment which results in such a shock to dynamic incentives.

While we have established that a relationship sometimes ends because the shipper demotes the carrier, it may also end because the shipper holds a new RFP and selects a different primary carrier. Suppose an RFP is held and the primary carrier learns that he has "lost" the RFP, so he will soon lose his primary position. This alters his dynamic incentives. Typically, about four or five weeks pass between the announcement of the RFP outcome and the enactment of the new routing guide that results from that RFP. This means that the carrier experiences a one-month "lame duck" period in which he knows that, after the end of the month, there is no prospect of future relationship surplus. During this lame duck period, we might expect to observe *endgame effects*, where the carrier's tendency to accept loads is diminished. Observing such endgame effects for losing carriers would strongly support the notion that prior to the last few weeks of the contract period, $\bar{p} > p + \eta$; that is, carriers' future relationship surplus induces a cooperative response.

Even for a carrier who "wins" the RFP and will maintain the primary position in the next contract period, dynamic incentives may still be altered in the last few weeks of the current contract period. If a shipper's incentive scheme were conditioned only on carrier behavior *within a contract period*, then a winning carrier's dynamic incentives would be greatly lessened by imminent end of the contract period; he would, in effect, get a "free

[33]See Section 1.2 for a discussion of these three carrier types.

pass," knowing that the slate will be wiped clean at the start of the next contract period. Observing this kind of endgame effect for winning carriers would therefore not only provide further evidence of a response to dynamic incentives, but would also speak to the *scope of the incentive mechanism* in the time dimension.

While we might worry that the timing of an RFP is not exogenous, we address this concern by restricting our attention to *mass RFP events*, where the shipper holds RFPs simultaneously on at least 30 lanes. We think it is unlikely that poor carrier performance on one lane will affect the shipper's decision of when to hold an RFP on such a large set of lanes.[34]

To study these hypothesized endgame effects, we estimate a linear probability model

$$\begin{aligned} \text{Accepted}^{\ell}_{sct} = &\beta_0 + \beta_1 \left(\text{spot rate}^{\ell}_{t} - \text{contract rate}^{\ell}_{sct}\right) \\ &+ \sum_{k=1}^{18} \alpha_k \mathbb{1}\{k \text{ weeks until end of contract}\} + \epsilon^{\ell}_{sct} \end{aligned} \tag{1}$$

regressing an indicator for the primary carrier's acceptance of a tender on a set of dummies for the number of weeks until the end of the contract period (when new rates are enacted), along with the deviation profit (the difference between the spot and contract rates), which captures the carriers' short-run incentives.[35] The pattern of week fixed effects $\{\alpha_k\}$ over time will provide insight into the proposed end-of-contract effects. As we are interested in the potential endgame effects for both losing (lame duck) carriers and winning carriers, we estimate (1) separately for these two groups of carriers. The estimated coefficients $\{\hat{\alpha}_k\}$ on the weeks-to-end-of-contract dummies, along with 95% confidence intervals, are plotted in Figure 4.[36]

In the final month of the contract period, losing primary carriers (solid lines) significantly reduce acceptance, with carriers 17pp less likely to accept load offers in the last week (as compared with a baseline rate of 71%). These economically and statistically significant endgame effects provide strong evidence that—prior to learning the RFP outcome—carriers respond to dynamic incentives. The large magnitude of these endgame effects suggests that $\bar{p} \gg p + \eta$, which would results from either large relationship rents or harsh shipper rejection penalties.

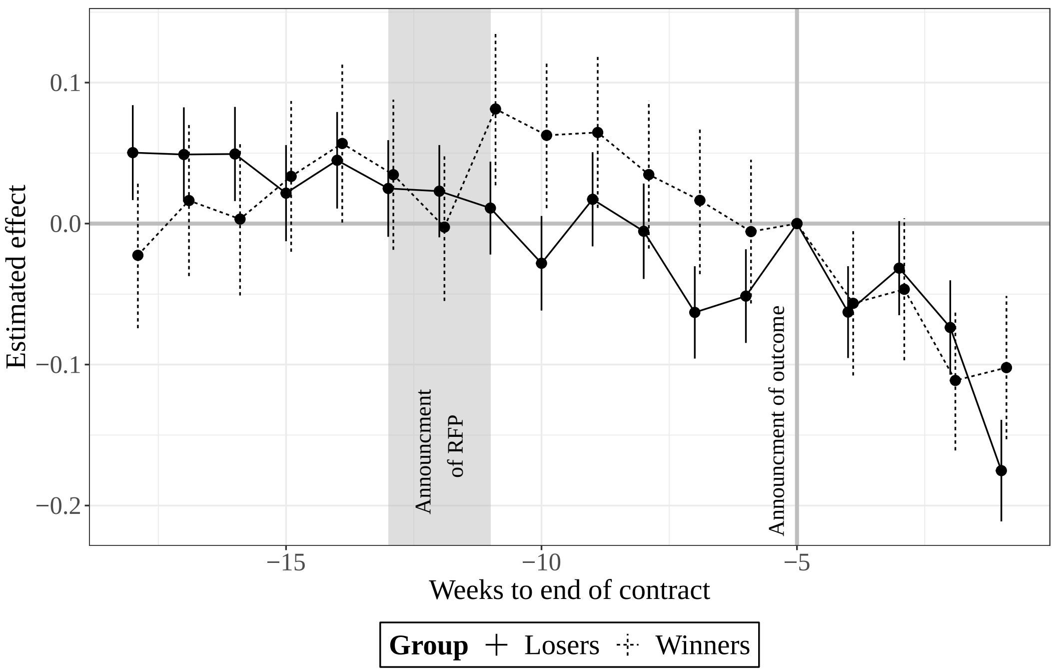
The estimated coefficients for winning carriers (dashed lines) similarly show a decline (albeit a smaller one) in acceptance after the RFP outcomes are announced. This happens

[34] This approach is intuitively similar to the "mass layoff" approach used to address worker selection issues in the labor literature.

[35] Notation: s indexes shippers, c indexes carriers, t indexes tenders, and ℓ indexes lanes.

[36] The omitted level is $k = 5$, i.e. we normalize $\hat{\alpha}_5 = 0$.

Figure 4: End-of-contract effects on tender acceptance for winning and losing carriers



Notes: This figure plots the estimated coefficients $\{\alpha_k\}$ from equation (1) with normalization $\alpha_5 = 0$. This choice reflects the fact that the RFP outcome is typically announced about 5 weeks before the end of the contract period. Also labeled in this plot is the approximate timing of the shipper announcing to carriers that an RFP will be held (6-8 weeks prior to the announcement of the RFP outcome).

despite the fact that the winning carriers now know they will continue to be primary carriers. This is consistent with the winning carrier anticipating that the slate will be wiped clean at the start of the new contract period. Once his primary status in the next contract period is secured, the carrier's incentive to perform well in the current contract period is much weakened. This strongly suggests that the scope of the incentive mechanism is *within* rather than *across* contract periods.

Another feature of Figure 6, however, does hint at the possibility of an across-contract period incentive mechanism. We see that, 9-11 weeks before the end of the contract period, the carriers who go on to win the RFP are slightly more likely to accept load offers. It is common for 6-8 weeks to pass between the shipper informing carriers of an upcoming RFP and announcing the winner of the RFP. Thus, by 11 weeks before the end of the contract period, carriers are likely aware that an RFP is imminent. If a carrier believed the RFP outcome to be conditional on his acceptance decisions, this would create an extra dynamic incentive to accept, possibly resulting in a kind of *window-dressing effect*. Indeed, we find evidence that carriers can affect RFP outcomes: while a primary carrier's current-period acceptances have a negligible effect on his probability of winning in the next RFP, they have a positive effect on the new contract rate conditional on winning. Specifically, we find that a 10pp reduction in rejection rate increases the probability of winning by only 0.7pp but, conditional on winning, increases the contract premium by 4 cents per mile (2% of the typical spot rate of $2 per mile).[37]

From the evidence presented in this subsection, we draw two conclusions: First, the temporal scope of the incentive mechanism is largely within the contract period. Second, the evidence on endgame effects strongly supports the hypothesis that carriers respond to dynamic incentives. This second conclusion, however, comes with a caveat: this evidence on dynamic incentives is limited to a selected subset of relationships (those with mass RFP events) and to a selected time period (the last 18 weeks of the contract period). To further support our conclusion, Section 5.3 will present additional evidence of carriers responding to dynamic incentives based on acceptance decisions across and throughout all relationships in our sample.

5.2 Empirical Evidence: Shippers' Strategies

Having established the temporal scope of the incentive mechanisms, we now assess its spatial scope, as well as its magnitude. To do this, we study shippers' decisions to demote primary carriers to determine whether and how much such demotion decisions on lane ℓ are conditioned on the carrier's performance on all lanes within the shipper-carrier relationship or only on the carrier's performance on lane ℓ. Rather than assuming that this spatial scope is the same across all shipper-carrier relationships, we estimate the shipper's demotion strategy separately for large asset-based carriers (large ABCs), small asset-based carriers (small ABCs), and brokers.[38] While we find that the incentive mechanism for brokers is at the firm-to-firm level, we find that the incentive mechanism for large ABCs is at the narrower lane level. For small ABCs, we do not find evidence of punishment.

At a high level, our approach is a simple one: We estimate the following linear probability model of the shipper's demotion strategy:

$$\begin{aligned}\text{Demotion}^{\ell}_{sct} = \gamma_0 &+ \gamma_{\text{Rej}(\ell)}\text{Rejection rate}^{\ell}_{sct} + \gamma_{\text{Rej}(-\ell)}\text{Rejection rate}^{-\ell}_{sct} \\ &+ \gamma_{\text{Rej}(\ell)\times\text{Rej}(-\ell)}\text{Rejection rate}^{\ell}_{sct} \times \text{Rejection rate}^{-\ell}_{sct} \\ &+ \gamma_X X^{\ell}_{sct} + \gamma_{\text{Rej}(\ell)\times X}\text{Rejection rate}^{\ell}_{sct} \times X^{\ell}_{sct} + \epsilon^{\ell}_{sct},\end{aligned} \tag{2}$$

[37]This is a sizable effect, in light of the fact that profit margins are generally low in the trucking industry, at around 2.5% to 6%. See Table 7 in Appendix C for a detailed description of our estimation.

[38]To define carrier types, we use an NMFTA crosswalk to convert the Standard Carrier Alpha Code (SCAC) identifiers in our microdata to US DOT codes. We then match US DOT codes to carriers' DOT registration for the year 2020. This method matches 90% of carriers in our data set to a fleet size variable and a broker/non-broker indicator. The latter two groups are for carriers that have divisions operating both a brokerage and an asset-based business.

where $\text{Demotion}_{sct}^{\ell}$ is an indicator for primary carrier c being demoted from primary status on lane ℓ between load t and load $t+1$, and X_{sct}^{ℓ} is a vector that includes the time-invariant relationship characteristics, along with the spot rate at the time of load t.[39] The critical regressors are $\text{Rejection rate}_{sct}^{\ell}$ and $\text{Rejection rate}_{sct}^{-\ell}$, which capture, respectively, the frequency of rejections by carrier c on lane ℓ and on all other lanes on which c is the primary carrier of shipper s. The coefficients on these variables and their interaction speak to the spatial scope of the incentive mechanism.

Defining rejection rates In contrast to the parsimonious model in Section 4, our empirical analysis allows shippers to have memories longer than one load. We adopt a functional form that allows the shipper's strategy to condition on a rejection rate index that summarizes the entire history of rejections in the relationship, though potentially giving greater weight to more recent rejections than less recent ones. For a shipper s, lane ℓ, and carrier c, this index takes the following form:

$$\text{Rejection rate}_{sct}^{\ell} = \frac{\sum_{k=0}^{t-1} \alpha^{\text{days}(t-k,t)} \text{Rejection}_{sct-k}^{\ell}}{\sum_{k=0}^{t-1} \alpha^{\text{days}(t-k,t)}}, \tag{3}$$

where $\text{Rejection}_{sct}^{\ell}$ is an indicator for carrier c rejecting a load t from shipper s on lane ℓ; $\text{days}\,(t-k,t)$ indicates the number of days between load $t-k$ and load t; and $\alpha \in [0,1]$ is a daily decay rate.[40] The other-lanes rejection rate $\text{Rejection rate}_{sct}^{-\ell}$ is defined analogously using acceptance/rejection decisions by carrier c on lanes other lane lane ℓ on which c is the primary carrier for s.

Relationship characteristics In estimating the shipper's strategy, we control for relationship characteristics relevant to the payoffs of the shipper and/or carrier.

First and foremost, X_{sct}^{ℓ} includes the *log of average monthly volume*, which is a proxy for δ, the frequency of interactions between shipper s and carrier c and lane ℓ. We measure monthly volume as the number of loads offered by the shipper on that lane in an active month.[41]

Second, X_{sct}^{ℓ} includes two measures of the *inconsistency of load timing*: The first measures

[39] Recall that by demotion, we mean a change to the lane-ℓ routing guide within a contract period that results in carrier c losing his primary position and being replaced by a new primary carrier. Since this definition is limited to changes within contract periods, any change in primary carrier that coincides with a change in rates (i.e. an RFP) on lane ℓ is not considered a demotion in our analysis.

[40] Note that the special case $\alpha \downarrow 0$ corresponds to the single-load memory restriction imposed in the model.

[41] By not conditioning on the identity of the accepting carrier, this measure avoids potential endogeneity issues. Moreover, since the primary carrier is the first carrier to receive an offer for each load, this measure approximates the expected number of offers the primary carrier receives per month during the relationship.

the degree to which the number of weekly loads varies from week to week. The second measures the degree to which the distribution of loads across days of the week varies from week to week. We treat these measures of load consistency as a component of the carrier's match-specific gain (η_1).[42] Intuitively, if the timing of loads is more consistent, it is easier for the primary carrier to plan his network of truck movements around the expected timing of offers. As discussed in Section 1.2, in the context of the trucking industry, such network planning could be important for reducing wasteful expenditures on fuel and labor.[43]

Finally, X^{ℓ}_{sct} includes the difference between the prevailing spot rate for lane ℓ at time t and carrier c's contract rate on lane ℓ. This captures the carrier's short-run incentive for deviation.

Identification strategy In estimating equation (2), we face a potential identification challenge stemming from the fact that the shipper's match-specific value (ψ) may shape the shipper's optimal strategy but is unobserved and therefore omitted. Several variables on the right-hand side of (2) are endogenous and are likely correlated with ψ. First, the rejection rate is the result of the carrier's endogenous response to the shipper's strategy. If the shipper's strategy is shaped by the omitted ψ, then the rejection rate will also be a function of ψ and therefore correlated with the regression error ϵ^{ℓ}_{sct}. Second, the contract rate, which enters X^{ℓ}_{sct}, is an endogenous outcome of the RFP process and is likely to be positively correlated with the shipper's match-specific value.

To address these endogeneity concerns, we use an instrumental variables approach that exploits exogenous variation in spot rates. First, we instrument for past acceptance/rejection decisions using the spot rates at the time at which each acceptance/rejection decision was made. To that end, we construct an index of past spot rates analogous to the construction of the rejection rate index, which serves as an instrument for the rejection rate.[44] We likewise construct an index of past spot rates on other lanes, which serves as an instrument for the other-lanes rejection rate. Second, we instrument for the contract rate using the spot rate at the time of the RFP in which the contract rate was established.[45] The idea is that at the RFP stage, the current spot rate serves as a competitive pressure on proposed contract rates.[46]

Using past spot rates as instruments is attractive because variation in spot rates is plausi-

[42]See Appendix B for a detailed description of how the two inconsistency measures are constructed.

[43]This network-planning explanation for carriers valuing consistent timing of load offers is widely accepted in the truckload industry.

[44]Specifically, Past spot index$^{\ell}_{sct} = \sum_{k=0}^{t-1} \alpha^{\text{days}(t-k,t)} \tilde{p}^{\ell}_{t-k} / \sum_{k=0}^{t-1} \alpha^{\text{days}(t-k,t)}$.

[45]To be more precise, we use Spot rate$^{\ell}_{sct}$ − Spot rate$^{\ell}_{sc0}$ to instrument for Spot rate$^{\ell}_{sct}$ − Contract rate$^{\ell}_{sc}$.

[46]Figure 1 shows that the average contract rate tends to adjust with spot rates, though with some lag. This supports the idea that spot rates create competitive pressure on contract rates.

bly exogenous for our purposes. While endogenous factors at the market level do determine spot rates, the industry is competitive enough that no one shipper or carrier has significant power to influence them. To satisfy the exclusion restriction, however, past spot rates must not directly affect the shipper's strategy. An identifying assumption is therefore that only the period-t spot rate affects the shipper's period-t demotion decision, something which aligns with industry experts' descriptions of typical demotion decision processes. According to these experts, shippers track carrier performance and make demotion decisions using a scorecard that records various performance aspects, including rejection history, but does not record spot rate history.[47]

Estimation We jointly estimate the parameters (α, γ) by GMM. The parameters γ are identified by the standard 2SLS moments. To identify α, we include a set of additional instruments, the prevailing spot rate on lane ℓ during each of the last five weeks, which, under the exclusion restriction, should be conditionally independent of the outcome.[48] For computational efficiency, we implement this GMM estimation via a nested algorithm. For a given value of α, the inner step uses 2SLS to obtain estimates of the linear parameters γ, while the outer loop searches for the value of α that minimizes the GMM objective function.

For the different carrier types, we estimate daily decay rates ranging from $\hat{\alpha} = 0.9742$ to $\hat{\alpha} = 0.9992$. This means that the shipper puts between 2.3% and 54.3% less weight on a rejection one month ago than on a rejection today. Our estimates of the linear parameters $\hat{\gamma}$ are reported in Table 2. For each carrier type, the parameter estimates for our main specification are the GMM estimates in the second column. For contrast and to illustrate the endogeneity problem described above, we also report in the first column the OLS estimates of the parameters γ.[49] We discuss the GMM estimates and their interpretation below.

Spatial scope and magnitude of the incentive mechanism The results in Table 2 show substantial heterogeneity in shipper strategies across carrier types. For brokers, we see that $\hat{\gamma}_{\text{Rej}(\ell)\times\text{Rej}(-\ell)}$ is positive and statistically significant, indicating that shippers punish multi-lane rejections more harshly. In contrast, for large ABCs, we see that $\hat{\gamma}_{\text{Rej}(\ell)}$ is positive and significant, while both $\hat{\gamma}_{\text{Rej}(-\ell)}$ and $\hat{\gamma}_{\text{Rej}(\ell)\times\text{Rej}(-\ell)}$ are insignificant, indicating that shippers

[47]Industry experts with whom we discussed these issues include Steve Raetz, Director of Research and Market Intelligence at CH Robinson; other members of Steve's team; and Chris Caplice and Angi Acocella of the MIT Center for Transportation and Logistics.

[48]These five lagged spot rates are instruments for the past acceptance/rejection decisions over the last five weeks. Under the functional form assumption in (3), however, individual rejection decisions enter into the shipper's strategy only through the rejection rate index. Thus, at the true α, these instruments for the individual acceptance/rejection decisions should be uncorrelated with the error term.

[49]These are the OLS estimates of γ with rejection rate measures constructed using the GMM estimate of the daily decay rate.

Table 2: Estimation of shipper's strategy

	All carriers		Large asset-based carriers		Small asset-based carriers		Brokers	
	(OLS)	(GMM)	(OLS)	(GMM)	(OLS)	(GMM)	(OLS)	(GMM)
Rejection rate	0.00842	0.00233	0.00823	0.084	0.012	0.000847	0.00642	-0.0151
	(0.000483)	(0.00467)	(0.000815)	(0.0337)	(0.00127)	(0.0123)	(0.00103)	(0.011)
Other lane rejection rate	-0.00547	-0.0166	-0.00241	-0.0408	0.000477	0.0364	0.00281	-0.0754
	(0.00086)	(0.0111)	(0.00136)	(0.032)	(0.00211)	(0.0463)	(0.00149)	(0.023)
Rejection rate × Other lanes rejection rate	0.0188	0.0743	0.0137	0.018	-0.00557	0.0102	0.00652	0.257
	(0.00142)	(0.0216)	(0.00218)	(0.0428)	(0.00357)	(0.0914)	(0.00264)	(0.0549)
Volume	-0.00689	-0.00684	-0.00438	-0.0137	-0.00415	-0.000986	-0.00805	-0.0106
	(0.000137)	(0.000576)	(0.000282)	(0.00218)	(0.000358)	(0.00224)	(0.000229)	(0.000589)
Inconsistency (loads / week)	0.0173	0.0341	0.0235	-0.0534	0.0122	0.00311	0.0154	0.0328
	(0.000679)	(0.00411)	(0.00176)	(0.0281)	(0.00194)	(0.00713)	(0.001)	(0.00274)
Inconsistency (day of week)	-0.00984	-0.0108	-0.00522	-0.0154	-0.00656	0.00406	-0.0122	-0.0207
	(0.000442)	(0.00134)	(0.000863)	(0.00485)	(0.0012)	(0.00364)	(0.000749)	(0.00177)
Spot rate - contract rate	-0.00224	-0.00947	-0.00346	-0.00264	-0.00362	-0.014	-0.00302	0.00958
	(0.000369)	(0.00199)	(0.00066)	(0.00587)	(0.000845)	(0.00859)	(0.000769)	(0.0066)
Rejection rate × Volume	-0.00759	-0.00747	-0.0105	0.00927	-0.00803	-0.0192	-0.0089	0.00788
	(0.000285)	(0.00163)	(0.000478)	(0.00451)	(0.000717)	(0.00637)	(0.000596)	(0.00223)
Rejection rate × Inconsistency (loads / week)	-0.0141	-0.0504	-0.0171	0.0746	-0.00792	0.00527	-0.0147	-0.0621
	(0.00109)	(0.00929)	(0.00219)	(0.037)	(0.00313)	(0.0109)	(0.00189)	(0.00875)
Rejection rate × Inconsistency (day of week)	-0.00709	0.00279	-0.0116	-0.0103	-0.0093	-0.0478	-0.0064	0.0257
	(0.00101)	(0.00464)	(0.00169)	(0.0129)	(0.00287)	(0.0128)	(0.00196)	(0.00809)
Rejection rate × (Spot rate - contract rate)	0.000754	0.0208	0.00195	-0.0287	0.00347	0.0252	0.00358	-0.0333
	(0.000769)	(0.00712)	(0.00112)	(0.0191)	(0.00199)	(0.0249)	(0.00187)	(0.0225)
$\hat{\alpha}$	-	1.000	-	0.974	-	0.999	-	0.985
N	680,229	680,229	173,787	173,787	67,508	67,508	250,197	250,197

Notes: Standard errors are in parentheses. For ease of interpretation, covariates that are interacted with rejection rate (other lanes rejection rate, volume, inconsistency, and spot rate minus contract rate) are normalized to have mean zero. The GMM specification jointly estimates α and the linear coefficients presented in this table. An outer loop searches over values of α, while, for a given α, an inner step estimates the linear coefficients by 2SLS and computes the 2SLS objective function. For each carrier type, the OLS specification takes the value of α estimated in the GMM specification as given and estimates the linear coefficients by OLS.

use a single-lane incentive mechanism for this type of carrier. Example 1 suggests a possible explanation for this difference across carrier types. Since an ABC's costs and match-specific benefits on a lane depend in large part on the lane's alignment with the rest of the carrier's network, we would expect large ABCs to have significantly more heterogeneity across lanes than brokers.[50] Such heterogeneity makes it harder to effectively combine incentives across lanes using simple incentive schemes, such as conditioning relationship continuation across relationships on a joint scorecard.

For all carrier types, the estimated coefficients suggest that the shipper's punishment scheme is soft, rather than harsh. For large ABCs, for instance, our estimate $\hat{\gamma}_{\mathrm{Rej}(\ell)}$ is

[50]The presence or absence of multi-lane punishment might also relate to divides within the shipper's organization. Shippers may divide responsibility for different lanes in various ways, for example, having separate teams managing inbound lanes (shipments of inputs from suppliers) versus outbound lanes (shipments of outputs to customers). Divides like these (combined with a lack of communication between these separate teams) may offer an operational explanation for the shipper's failure to implement multi-lane punishment. Yet such organizational divides arise endogenously; by choosing to separate responsibility for different lanes, shippers forgo the potential benefits of multi-lane punishment.

positive and very statistically significant, indicating that shippers punish rejections with an increased probability of demotion. At first glance, however, this coefficient may appear very small, as it indicates that an increase in the rejection rate from 0% to 100% increases the probability of demotion between load t and load $t+1$ by only 8.4 percentage points. However, this coefficient should be interpreted in light of the fact that Rejection rate$_{sct}^{\ell}$ is a persistent state variable; since $\hat{\alpha} \gg 0$, a rejection of one load results in a sustained increase in the probability of demotion for many periods to come.

To get a sense of the economic significance of the estimated degree of punishment, we run a simple simulation to illustrate the effect of a rejection on the expected duration of a relationship.[51] The results show that if a large asset-based carrier rejects the first offer for the relationship, the expected relationship duration is 63.5 loads, as compared to 58.5 loads if he accepts the first offer. This 5 load effect difference is economically large. The mean per-load payment from shipper to carrier is $1,129, so a rejection early in the relationship may cost the carrier as much as $5,645 in revenue. The prospect of such a loss from a single rejection is likely to create meaningful incentives for carrier cooperation. Nevertheless, we conclude that punishment is soft, not harsh.

5.3 Empirical Evidence: Carriers' Acceptance

To conclude our empirical evidence, we use carrier behavior throughout the relationship to build upon the evidence presented in Section 5.1, bolstering the claim that carriers respond to dynamic incentives and quantifying the magnitude of this response. To do this, we estimate the response of carriers' tendency to accept loads to relationship characteristics. We show that carriers respond strongly to lane volume, which would not be true of a carrier playing static best response.

As we did for shippers' strategies in the previous subsection, we estimate a linear probability model,

$$\begin{aligned}\text{Accepted}_{sct}^{\ell} = \beta_0 \text{ controls}_s^{\ell} + &\ \beta_{\text{volume}} \text{volume}_s^{\ell} \\ &+ \beta_{\text{inc.}} \text{inconsistency}_s^{\ell} + \beta_{\text{volume}\times\text{inc.}} \text{volume}_s^{\ell} \times \text{inconsistency}_s^{\ell} \\ &+ \beta_{\text{spot}} (\text{spot - contract})_t^{\ell} + \beta_{\text{volume}\times\text{spot}} \text{volume}_s^{\ell} \times (\text{spot - contract})_t^{\ell} + \epsilon_{sct}^{\ell},\end{aligned} \tag{4}$$

regressing an indicator for carrier c accepting load t from shipper s on lane ℓ on controls, along with a set of lane and relationship characteristics. Our choice of functional form—in

[51]See Appendix B.2 for the details of our simulation exercise.

particular, the inclusion of interactions between volume and other characteristics—reflects insights from Proposition 1.

Identification strategy In estimating (4), we face two identification challenges stemming from the fact that a component of match-specific gain (η_2), as well as the carrier's cost distribution (F), is unobserved and thus omitted.

First, as in the previous subsection, we face the problem that the contract rate is an endogenous object likely correlated with the omitted variable.[52] We again address this issue by instrumenting for the contract rate using the spot rate at the time of the RFP in which the contract rate was established.

Second, while our use of average volume as a proxy for the frequency of future interactions (and thus for the discount factor δ) is in keeping with the empirical literature on relational contracting (e.g. Gil and Marion (2013)), we would face a potential identification challenge if this measure of volume were correlated with the unobserved component of carrier's match-specific gains, η_2. Such correlation might arise either because of a *selection effect* or an *investment effect*. The selection effect would result if carriers with better match-specific value or cost were systematically more likely to be primary carriers on higher-volume lanes.[53] The investment effect would result if the carrier could, for example, adjust its network of truck movements to better serve the contracted lane; he would have greater incentive to carry out such adjustments on a higher-volume lane.[54]

In either case, a positive correlation between volume and the unobserved match-specific value would induce bias in our key parameter of interest, β_{volume}. To address the potential bias resulting from the selection effect, we include two sets of fixed effects that absorb variation in the carrier's unobserved match-specific gain and cost: the first are shipper-carrier fixed effects; the second are or carrier-origin-destination-year fixed effects, where origin and destination are defined as Census regions. By including the shipper-carrier fixed effects, we absorb a variety of potential shipper-carrier specific components of the unobserved value, including, for instance, relationship-specific knowledge and integration of payment or communication systems.[55] By including the carrier-origin-destination-year fixed effects, we absorb

[52]Since contract rates are established through an RFP process, a carrier with high match-specific value would tend to submit a lower bid, so we expect upward bias in the OLS estimates of β_{spot} and $\beta_{\text{volume}\times\text{spot}}$.

[53]There are several ways this selection could arise. First, carriers with high η could be more likely to submit bids for RFPs on high-volume lanes. Second, in choosing winners of RFPs, shippers might be more likely to select high-η carriers from among the bidders on high-volume lanes.

[54]Note that this investment effect story involves a violation of the timing assumption of the model, which imposes that (η_2, F) are fixed and known at time $t = 0$.

[55]Note, however, that when these fixed effects are included, we estimate the parameters by exploiting variation in characteristics *across* different lanes of the same shipper-carrier pair. This approach would not produce meaningful estimates of the effects of these characteristics if shippers employed multilane punish-

key geographic components of the carrier's match-specific value, namely the compatibility of a particular (broadly defined) route with the rest of the carrier's network.

For each carrier type, the first column in Table 3 reports the OLS estimates of Equation (4) while the second column reports the IV estimates. The fifth column reports estimates for the main specification, IV with carrier-origin-destination-year and shipper-carrier fixed effects included. For comparison, the third and fourth columns each report IV estimates for a specification including only one of these sets of fixed effects.

Response to dynamic incentives The results for the main specification indicate that β_{volume} is positive and signficant for all carrier types, which would not be true of a carrier playing static best responses.[56] We interpret this result as strong evidence that carriers respond to dynamic incentives. We also observe striking differences in the magnitude of this response across carrier types: large ABCs exhibit a very strong dynamic response (doubling volume increases acceptance probability by 6.2pp), whereas brokers (2.0pp) and small ABCs (1.3pp) show weaker responses. Notably, this ordering of carrier types by responsiveness to dynamic incentives aligns precisely with the relative strength of same-lane dynamic incentives created by the incentive schemes each type faces.[57]

As described above, the appropriateness of our average volume measure as a proxy for the discount factor δ is potentially threatened by the selection and investment effects, which, if present, would result in an upward bias in $\hat{\beta}_{\text{volume}}$. However, comparing the IV estimates to the IV-FE1 and IV-FE2 specifications, we see that, for ABCs, the inclusion of fixed effects actually *increases* $\hat{\beta}_{\text{volume}}$. This speaks against the hypothesized selection and/or investment effects. However, the fact that the inclusion of fixed effects changes the estimates for ABCs so substantially suggests that there is substantial heterogeneity in costs (F) and match-specific values (η_2) which is absorbed by these fixed effects. For brokers, in contrast, the inclusion of fixed effects changes the estimates very little; this suggests that brokers are much less heterogeneous in costs and match-specific values than ABCs, which accords with the fact that brokers do not need to undertake network planning.

While we rule out the potential endogeneity issue caused by a positive correlation between volume and relationship-specific investments, these investments—which are likely to take the form of a carrier adjusting his network of truck movements—could still directly affect carriers' behavior. To assess the role of network adjustments, we include as a regressor an indicator

ment strategies; if that were the case, a carrier's acceptance/rejection decisions on lane ℓ might respond to characteristics of other lanes. However, our analysis indicates that this is not a concern for ABCs.

[56] For such a carrier, acceptances would depend only on the current spot rate and threshold $\underline{p} = \eta_1 + \eta_2 + p$.

[57] Recall that results in the previous subsection show that large ABCs face a relatively harsh own-lane punishment scheme, while brokers face a multi-lane scheme and small ABCs face a negligible degree of punishment.

Table 3: Estimation of carriers' acceptance

	All carriers					Large asset-based carriers				
	(OLS)	(IV)	(IV-FE1)	(IV-FE2)	(IV-FE3)	(OLS)	(IV)	(IV-FE1)	(IV-FE2)	(IV-FE3)
Volume	0.00792	0.000376	0.00790	0.0188	0.00925	-0.00408	0.0325	0.150	0.0558	0.0893
	(0.00057)	(0.000763)	(0.00167)	(0.0009)	(0.00157)	(0.00142)	(0.0035)	(0.029)	(0.0079)	(0.0169)
Spot rate - contract rate	-0.223	-0.261	-0.150	-0.231	-0.150	-0.142	-0.191	-0.126	-0.133	-0.117
	(0.001)	(0.003)	(0.002)	(0.002)	(0.002)	(0.002)	(0.006)	(0.012)	(0.005)	(0.008)
Inconsistency (loads / week)	-0.0898	-0.0869	-0.0686	-0.0872	-0.0686	-0.264	-0.262	-0.173	-0.177	-0.156
	(0.0024)	(0.0026)	(0.0028)	(0.0027)	(0.0028)	(0.007)	(0.010)	(0.015)	(0.009)	(0.011)
Inconsistency (day of week)	-0.0382	-0.0370	-0.0346	-0.0427	-0.0282	-0.0576	0.00157	0.0995	-0.0170	0.0460
	(0.0015)	(0.0016)	(0.0021)	(0.0020)	(0.0019)	(0.0035)	(0.00663)	(0.0277)	(0.0091)	(0.0159)
Volume × (Spot rate - contract rate)	0.0356	0.362	0.388	0.362	0.334	0.0352	-0.772	-1.460	-0.774	-0.902
	(0.0009)	(0.019)	(0.030)	(0.028)	(0.028)	(0.0023)	(0.064)	(0.311)	(0.111)	(0.189)
Volume × Inconsistency (loads / week)	-0.0330	-0.0178	-0.0270	-0.0219	-0.0264	-0.0717	-0.0924	-0.0592	-0.125	-0.0592
	(0.0022)	(0.0025)	(0.0024)	(0.0024)	(0.0023)	(0.0065)	(0.0087)	(0.0113)	(0.010)	(0.0085)
Volume × Inconsistency (day of week)	-0.0589	-0.0332	-0.0432	-0.0300	-0.0368	-0.0791	-0.109	0.0302	-0.0561	0.0104
	(0.0016)	(0.0023)	(0.0021)	(0.0017)	(0.0019)	(0.0038)	(0.006)	(0.0104)	(0.0048)	(0.0058)
< 7 days since promtion	-0.0848	-0.0919	-0.0330	-0.0441	-0.0296	-0.143	-0.124	0.0127	-0.0289	-0.000743
	(0.0020)	(0.0022)	(0.0019)	(0.0020)	(0.0018)	(0.005)	(0.007)	(0.0150)	(0.0072)	(0.009930)
Fixed effects										
Carrier × region × region × year			X		X			X		X
Shipper × carrier				X	X				X	X
N	796346	796346	796167	796346	796167	167095	167095	167065	167095	167065

	Small asset-based carriers					Brokers				
	(OLS)	(IV)	(IV-FE1)	(IV-FE2)	(IV-FE3)	(OLS)	(IV)	(IV-FE1)	(IV-FE2)	(IV-FE3)
Volume	-0.0374	-0.0352	0.0223	-0.0422	0.0194	0.0230	0.0299	0.0277	0.0473	0.0292
	(0.0019)	(0.0019)	(0.0031)	(0.0027)	(0.0031)	(0.0009)	(0.0013)	(0.0012)	(0.0013)	(0.0013)
Spot rate - contract rate	-0.323	-0.159	-0.0950	-0.111	-0.0969	-0.338	-0.356	-0.248	-0.337	-0.243
	(0.004)	(0.008)	(0.0081)	(0.008)	(0.0081)	(0.02)	(0.004)	(0.005)	(0.004)	(0.005)
Inconsistency (loads / week)	-0.183	-0.135	-0.123	-0.191	-0.123	-0.0339	-0.0278	-0.0208	-0.0483	-0.0248
	(0.012)	(0.014)	(0.016)	(0.015)	(0.016)	(0.0030)	(0.0035)	(0.0043)	(0.0041)	(0.0043)
Inconsistency (day of week)	-0.113	-0.0850	-0.00348	0.00453	-0.00132	0.00305	-0.0638	-0.0159	-0.0293	-0.0179
	(0.006)	(0.0070)	(0.01170)	(0.009630)	(0.01190)	(0.00218)	(0.0070)	(0.0035)	(0.0041)	(0.0038)
Volume × (Spot rate - contract rate)	-0.0648	-0.0743	0.0254	0.0641	0.0293	0.0771	0.576	0.248	0.385	0.261
	(0.0028)	(0.0129)	(0.0162)	(0.0139)	(0.0161)	(0.0017)	(0.049)	(0.030)	(0.030)	(0.032)
Volume × Inconsistency (loads / week)	-0.0605	-0.0166	-0.0742	-0.0698	-0.0726	-0.0255	-0.0457	-0.00803	-0.0205	-0.0108
	(0.0076)	(0.0086)	(0.0097)	(0.0091)	(0.0097)	(0.0030)	(0.0040)	(0.00357)	(0.0038)	(0.0036)
Volume × Inconsistency (day of week)	-0.0633	-0.0257	-0.0299	0.00580	-0.0224	-0.0304	-0.0635	-0.0297	-0.0199	-0.0289
	(0.0056)	(0.0066)	(0.0088)	(0.00832)	(0.0090)	(0.0024)	(0.0042)	(0.00293)	(0.0029)	(0.0030)
< 7 days since promtion	-0.113	-0.106	-0.0525	-0.0625	-0.0544	-0.0271	-0.0361	-0.00188	-0.00873	-0.00165
	(0.008)	(0.008)	(0.0063)	(0.0067)	(0.0063)	(0.0030)	(0.0035)	(0.00266)	(0.00301)	(0.00267)
Fixed effects										
Carrier × region × region × year			X		X			X		X
Shipper × carrier				X	X				X	X
N	64478	64478	64458	64478	64458	255686	255686	255615	255686	255615

Notes: Standard errors in parentheses. Controls include distance and distance squared. Standard errors are in parentheses. For ease of interpretation, the covariates that are interacted with volume (inconsistency and (spot rate - contract rate)) are normalized to have mean zero. Specifications IV-FE1 and IV-FE3 include carrier × origin region × destination region × year fixed effects; specifications IV-FE2 and IV-FE3 include shipper × carrier fixed effects. For the former set of fixed effects, regions are determined using US Census regions of loads' origin and destination locations. The small difference in sample size across the specifications with and without fixed effects reflects the fact that singleton observations are dropped in the specifications with fixed effects.

for whether a load t occurs within 7 days after carrier c's promotion to primary status on lane ℓ. Since network adjustments likely take some time to complete, we would expect the carrier's tendency to accept to be lower when it is freshly promoted. This is the case only for small ABCs. We find that a small ABC is 5.4pp less likely to accept a load in the first week after promotion.

Responses to inconsistency and spot rates In addition to shedding light on responses to dynamic incentives, the estimates in Table 3 also provide insight into two other key aspects of carrier behavior and how these differ by carrier type.

First, we see that inconsistency (a component of η_1) does indeed affect carriers' acceptance decisions. In particular, lanes with more inconsistency in the number of loads per week have lower acceptance probabilities, with this effect being stronger on higher-volume lanes (matching the predictions of Proposition 1). The measure of inconsistency in the timing of loads within the week is seemingly of lesser importance.

Second, responses to variation in spot rates differ substantially across carrier types, with brokers being far more sensitive than ABCs. This is intuitive, as the profit margins of brokers, who subcontract loads to the spot market, are much more closely tied to spot rates than those of ABCs, who transport loads using their own physical assets. Note that since spot rates in reality exhibit autocorrelation, lower acceptance in response to a higher current spot rate captures two effects: (i) contemporaneous temptation and (ii) a decrease in the carrier's continuation value from the relationship (since future spot rates are also likely to be higher).

6 Discussion

In this section, we discuss alternative mechanisms of the interactions between shippers and carriers. These mechanisms will complement the punishment mechanism modeled in Section 4 to explain the richness of the empirical findings in Section 5.

Table 4 provides a summary of our key empirical findings.

On the one hand, the punishment mechanism can explain most of these findings. First, brokers face multi-lane punishment and thus respond moderately to own-lane volume. Second, large ABCs face single-lane punishment and thus respond strongly to own-lane volume. Third, small ABCs face no punishment and thus have the weakest response to volume. Finally, the differences in the spatial scope of relationships for brokers and asset-based carriers could be explained by the fact that brokers face less heterogeneity in gains and costs across lanes, making effective incentive pooling easier to achieve for brokers.

Table 4: A recap of empirical findings

Evidence	own-lane	multi-lane	response to volume	heterogeneity	network adjustment
Brokers	-	✓	moderate	-	-
Large asset-based carriers	✓	-	strong	✓	-
Small asset-based carriers	-	-	mild	✓	✓

On the other hand, the heterogeneity across relationships of asset-based carriers could give rise to another dynamic mechanism—learning. We argue that while learning cannot be the only mechanism, it serves as a complementary mechanism in explaining our evidence on asset-based carriers, particularly small ones.

A learning model for asset-based carriers Suppose carriers have unobserved characteristics, such as idiosyncratic gains or costs from the relationship. In this case, past rejections would be indicative of future rejections, thereby affecting the shipper's expected value from maintaining the carrier's primary status. Notice that being ranked first on the routing guide ensures the primary carrier greater consistency in the timing of offered loads, facilitating network planning and load fulfillment. Thus, the opportunity cost of maintaining a primary carrier who is likely to reject offers is the higher acceptance probability that the first backup carrier *would have* were he to be primary. This is one reason why the shipper might prefer a primary carrier with higher acceptance probability. From the perspective of a primary carrier, learning by the shipper that conditions on past rejections would also create dynamic incentives for the carrier to accept more loads.

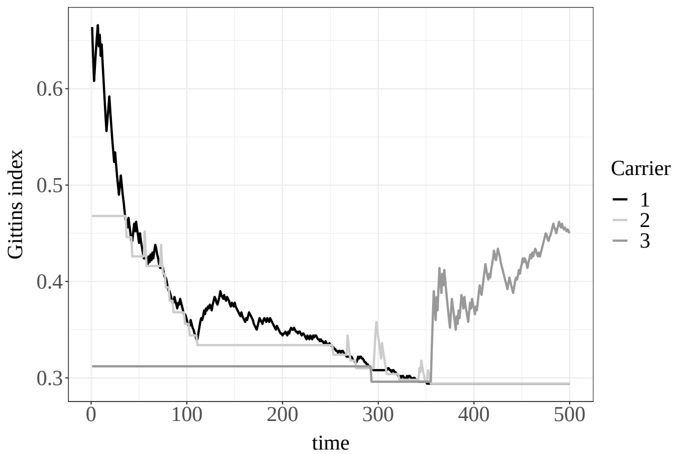
Assumption 1. *Each carrier has a permanent tendency to accept loads and the shipper holds independent priors over carriers' acceptance tendencies.*

Learning cannot be the only mechanism However, we observe that demotions are generally permanent, a fact which rules out learning as the only mechanism. Under Assumption 1, the potential learning mechanism described above can be simplified to a bandit problem with independent arms, a common model in the literature on learning. By choosing a carrier to be the primary carrier and receive the first offer in each period, the shipper gradually learns the carrier's tendency to accept loads as a primary carrier. The solution of the shipper's dynamic optimization problem is as follows: each period, she chooses as primary the carrier with the highest Gittins index, which captures both the exploitation and exploration value of choosing a carrier over the outside option.[58] Given that the tendency to accept

[58]The exploitation value of an option refers to the expected payoff of that option given the current beliefs. The exploration value refers to the informational value of an additional observation of that option. See Whittle (1980) and Weber et al. (1992) for details.

loads should be independent across carriers once conditioned on observed characteristics, a carrier's Gittins index evolves only when he is chosen as primary.

Figure 5: A simulated learning path



Notes for Figure 5: In this example, the shipper's prior is overly optimistic about Carriers 1 and 2, and overly pessimistic about Carrier 3. Thus, the Gittins indices of the former two carriers are generally decreasing, both because of the initial overoptimism and because the decrease in informational values as they are chosen. The shipper makes many switches between these two carriers before she starts to experiment with Carrier 3, at which time the Gittins index of Carrier 3 evolves.

Table 5: Probability of repromotion

	Returns to primary	*Never again primary*	*Return probability*
Panel A: All carriers			
Full sample (2015-2019)	405	8701	4.45%
First half (2015-2017)	246	4044	5.73%
Panel B: Large asset-based carriers			
Full sample (2015-2019)	106	1378	7.14%
First half (2015-2017)	65	655	9.03%
Panel C: Small asset-based carriers			
Full sample (2015-2019)	38	647	5.55%
First half (2015-2017)	18	180	9.09%
Panel D: Brokers			
Full sample (2015-2019)	116	4125	2.74%
First half (2015-2017)	61	2035	2.91%

Notes for Table 5: This table presents evidence on the frequency of a carrier who is demoted from the primary position on a lane ℓ ever returning to the primary position on lane ℓ. The first column list the number of times in the given subsample that we observe such a return. The second column lists the number of instances where the demoted carrier never again regains the primary position on the lane. One concern with this exercise is that—since our sample is finite—we might not observe a carrier regaining the primary position because this occurs after the end of our sample period. To ameliorate this concern, the second row restricts the sample to carriers that are demoted in the first half of the sample period. In both rows, we also restrict the sample to the set of lanes that only ever have a single primary carrier at any point in time.

Figure 5 provides an example illustrating the evolution of Gittins indices in a learning problem with three carriers. Initially, Carrier 1 is primary. The shipper continues to choose Carrier 1 until her belief about this carrier's tendency to accept loads drops just below that of Carrier 2, at which point she switches to the latter carrier. While the shipper chooses Carrier 2, the Gittins index of Carrier 1 remains the same and well above that of Carrier 3. Thus, the next time the shipper needs to make a switch, she switches back to Carrier 1 rather than switching to Carrier 3. This intuition generalizes: when learning steps are small, we expect to see the shippers switching from Carrier 1 to Carrier 2 and then back to Carrier 1 ("switch-back pattern") much more often than we see her switching from Carrier 1 to Carrier 2 and then to Carrier 3.

While the learning story predicts the prevalence of the switch-back pattern, a carrier returning to the primary position on a lane after being demoted ("repromotion") is rare in our TMS microdata. Table 5 breaks down the probabilities of repromotion by carrier type and timing of demotion. For an asset-based carrier who is demoted, the probability of ever being repromoted is about 9% for demotions occurring in 2015-2017 and about 6-7% for demotions throughout the sample. The difference between these two statistics further suggests that if a carrier does get repromoted, this occurs only after a long period of time.[59] Note also that the repromotion probability of brokers is even lower, at less than 3%. This is consistent with the lack of heterogeneity in brokers' performance; there is little to learn about individuals drawn from such a homogeneous population.

Shippers' commitment to punish large asset-based carriers Evidence that shippers are on average worse off after replacing a primary carrier would strongly support the hypothesis that shippers commit to punishing carriers who reject loads. We find such evidence for large ABCs. In comparison to demoted carriers, promoted carriers of this group have substantially worse acceptance rates (15 pp lower) and no discounts in contract rates.[60] We thus conclude that while learning might offer another reason for not combining the performance across all lanes for large ABCs, its role in explaining these carriers' behavior is small. Instead, the leading mechanism is likely that shippers use a punishment scheme to discipline large ABCs' behavior.

Learning as a complementary mechanism Though our data rejects learning as the only mechanism, a combination of learning and relationship-specific investments could explain the behavior of small ABCs. In particular, suppose that small ABCs undo their relationship-specific investments after being demoted. Such undoing would invalidate Assumption 1 and thus, the prevalence of switch-back patterns predicted by our learning model.[61] Since we do not observe carriers' networks, this paper does not explore this potential learning-investment mechanism for small ABCs in detail; this would, however, be an interesting avenue for future research.

[59]Conditional on being repromoted, the average time between demotion and repromotion is about 200 days.

[60]See Table 8 in Appendix C for results for all carrier types.

[61]Consider our example in Figure 5 but with a new assumption that being a primary carrier requires network adjustments. In this case, Carrier 1 would undo his investment after the shipper first switches away from him to Carrier 2. This means that when the shipper then contemplates switching away from Carrier 2, her value from switching to Carrier 1 is potentially lower than the value of continuing using Carrier 1 at the time the shipper made her first switch. Thus, switch-back patterns need not be prevalent in models that combine learning and adaptations.

7 Conclusion

In this paper, we ask how informal interfirm relationships work in an economically important setting: the US truckload freight industry. We use a novel transaction-level data set uniquely well-suited to studying informal relationships to provide evidence on the mechanism governing these relationships, as well as the scope of that mechanism.

We begin by presenting evidence of endgame effects, a phenomenon that suggests that the temporal scope of the incentive mechanism is within contract periods. Next, we estimate the shipper's demotion strategy. The results indicate that while the spatial scope of the demotion strategy is limited to a single lane for large asset-based carriers, shippers employ multi-lane punishment for brokers. Third, we quantify carriers' responses to the dynamic incentives generated by the incentive scheme. Finally, we address alternative, non-punishment mechanisms. Our evidence suggests that, while punishment seems to be the primary mechanism at play, other mechanisms—in particular, specific investment—likely play a role for relationships involving small asset-based carriers.

Taken together, these results provide valuable insight for future empirical work both on long-term relationships and on the trucking industry. First, we use rich microdata to empirically test the common assumption that relationship scope is at the firm-to-firm level. Our findings demonstrate the value of such a test, which we demonstrate to be feasible with increasingly available microdata. Second, by providing a detailed empirical description of the shipper-carrier relationships around which the trucking industry is organized, this paper serves as a key stepping stone to studying other important questions about this macroeconomically vital, yet understudied, industry.

References

Acocella, A. and C. Caplice (2023). Research on truckload transportation procurement: A review, framework, and future research agenda. *Journal of Business Logistics*.

Acocella, A., C. Caplice, and Y. Sheffi (2020). Elephants or goldfish?: An empirical analysis of carrier reciprocity in dynamic freight markets. *Transportation Research Part E: Logistics and Transportation Review 142*, 102073.

Baker, G., R. Gibbons, and K. J. Murphy (2002). Relational contracts and the theory of the firm. *The Quarterly Journal of Economics 117*(1), 39–84.

Baker, G. P. and T. N. Hubbard (2003). Make versus buy in trucking: Asset ownership, job design, and information. *American Economic Review 93*(3), 551–572.

Baker, G. P. and T. N. Hubbard (2004). Contractibility and asset ownership: On-board computers and governance in us trucking. *The Quarterly Journal of Economics 119*(4), 1443–1479.

Barron, D., R. Gibbons, R. Gil, and K. J. Murphy (2020). Relational adaptation under reel authority. *Management Science 66*(5), 1868–1889.

Bernheim, B. D. and M. D. Whinston (1990). Multimarket contact and collusive behavior. *The RAND Journal of Economics*, 1–26.

Brugues, F. (2023). Take the goods and run: Contracting frictions and market power in supply chains. Technical report, Working Paper.

Caplice, C. (2007). Electronic markets for truckload transportation. *Production and Operations Management 16*(4), 423–436.

Gil, R., M. Kim, and G. Zanarone (2021). Relationships under stress: Relational outsourcing in the us airline industry after the 2008 financial crisis. *Management Science*.

Gil, R. and J. Marion (2013). Self-enforcing agreements and relational contracting: evidence from california highway procurement. *The Journal of Law, Economics, & Organization 29*(2), 239–277.

Green, E. J. and R. H. Porter (1984). Noncooperative collusion under imperfect price information. *Econometrica: Journal of the Econometric Society*, 87–100.

Harris, A. and T. M. A. Nguyen (2023). Long-term relationships and the spot market: Evidence from us trucking.

Hubbard, T. N. (2001). Contractual form and market thickness in trucking. *RAND Journal of Economics*, 369–386.

Levin, J. (2003). Relational incentive contracts. *American Economic Review 93*(3), 835–857.

Li, J. and N. Matouschek (2013). Managing conflicts in relational contracts. *American Economic Review 103*(6), 2328–51.

Macaulay, S. (1963). Non-contractual relations in business: A preliminary study. *American sociological review*, 55–67.

Macchiavello, R. and A. Morjaria (2015). The value of relationships: evidence from a supply shock to kenyan rose exports. *American Economic Review 105*(9), 2911–45.

Macchiavello, R. and A. Morjaria (2023). Relational contracts: Recent empirical advancements and open questions. Technical report, National Bureau of Economic Research.

MacLeod, W. B. and J. M. Malcomson (1989). Implicit contracts, incentive compatibility, and involuntary unemployment. *Econometrica: Journal of the Econometric Society*, 447–480.

Malcomson, J. M. (2010). Relational incentive contracts. *Handbook of Organizational Economics*.

Masten, S. E. (2009). Long-term contracts and short-term commitment: Price determination for heterogeneous freight transactions. *American Law and Economics Review 11*(1), 79–111.

Pavan, A., I. Segal, and J. Toikka (2014). Dynamic mechanism design: A myersonian approach. *Econometrica 82*(2), 601–653.

Rose, N. L. (1985). The incidence of regulatory rents in the motor carrier industry. *The RAND Journal of Economics*, 299–318.

Rose, N. L. (1987). Labor rent sharing and regulation: Evidence from the trucking industry. *Journal of Political Economy 95*(6), 1146–1178.

Scott, A., C. Parker, and C. W. Craighead (2017). Service refusals in supply chains: Drivers and deterrents of freight rejection. *Transportation Science 51*(4), 1086–1101.

Taylor, D. (2011). Trucking industry updates (03/22/2011).

Weber, R. et al. (1992). On the gittins index for multiarmed bandits. *The Annals of Applied Probability 2*(4), 1024–1033.

Whittle, P. (1980). Multi-armed bandits and the gittins index. *Journal of the Royal Statistical Society: Series B (Methodological) 42*(2), 143–149.

A Proofs

A.1 Carriers' acceptance

Denote by V the average discounted expected utility of the carrier from the relationship. Denote by $V(A)$ and $V(R)$ the average discounted expected utilities of the carrier at the beginning of period t following $d_{t-1} = A$ and $d_{t-1} = R$, respectively. We have

$$\begin{aligned} V &= \mathbf{E}_{\tilde{p}_t, c_t} \left[\max \left\{ (1-\delta)(\eta + p - c_t) + \delta V(A), (1-\delta)(\tilde{p}_t - c_t) + \delta V(R), \delta V(R) \right\} \right] \\ &= \delta V(R) + (1-\delta) \mathbf{E}_{\tilde{p}_t, c_t} \left[\max \left\{ \eta + p - c_t + \frac{\delta}{1-\delta}(V(A) - V(R)), \tilde{p}_t - c_t, 0 \right\} \right], \end{aligned} \tag{5}$$

where

$$V(A) = \sigma_0(A)V + (1 - \sigma_0(A))\underline{V}, \tag{6}$$

$$V(R) = \sigma_0(R)V + (1 - \sigma_0(R))\underline{V}. \tag{7}$$

Let $\overline{p} = \eta + p + \frac{\delta}{1-\delta}(V(A) - V(R))$ and $h(\overline{p}) = \mathbf{E}_{\tilde{p}_t, c_t}[\max\{\overline{p} - c_t, \tilde{p}_t - c_t, 0\}]$. When $\tilde{p}_t < \overline{p}$, the carrier's optimal strategy is to accept whenever $c_t < \overline{p}$. When $\tilde{p}_t > \overline{p}$, the carrier optimally rejects regardless of the cost draw. Thus, the probability of acceptance at each level of spot rate is $\Pr(d_t = A|\tilde{p}_t) = \mathbf{1}(\tilde{p}_t < \overline{p})F(\overline{p})$, increasing in the acceptance threshold $\overline{p}$. Manipulating Equations (5), (6) and (7) yields the following fixed point equation of $\overline{p}$,

$$\frac{1 - \delta\sigma_0(R)}{\sigma_0(A) - \sigma_0(R)}(\overline{p} - \eta - p) = \delta(h(\overline{p}) - \underline{V}). \tag{8}$$

Lemma 1. *$h' \in [0, 1]$ and $h'' \geq 0$.*

Proof. By the independence of spot rates and cost draws,

$$h(\overline{p}) = G(\overline{p}) \int_0^{\overline{p}} (\overline{p} - c)f(c)dc + \int_{\overline{p}}^{\infty} \int_0^{\tilde{p}} (\tilde{p} - c)f(c)g(\tilde{p})dcd\tilde{p}.$$

Thus $h'(\overline{p}) = G(\overline{p})F(\overline{p}) \in [0, 1]$ and $h''(\overline{p}) = g(\overline{p})F(\overline{p}) + G(\overline{p})f(\overline{p}) \geq 0$. □

Proof of Proposition 1 Notice that if $\eta + p \in \text{Supp}(G)$, there is an option value to having contracted offers. That is, $V > \underline{V}$. This means that when $\sigma_0(A) > \sigma_0(R)$, we have $V(A) > V(R)$, and thus $\bar{p} > \eta + p$. This completes the proof of part (i).

Next, we exploit Equation (8) to generate predictions on how relationship characteristics and the reward-punishment scheme affect the likelihood of the carrier accepting, as captured by threshold $\overline{p}$. Referring to the left-hand-side and the right-hand-side of Equation (8) as LHS and RHS, we have

$$\frac{\partial(LHS - RHS)}{\partial \overline{p}} = \frac{1 - \delta\sigma_0(R)}{\sigma_0(A) - \sigma_0(R)} - \delta h'(\overline{p}) > 1 - h'(\overline{p}) \geq 0,$$

where the last inequality follows from Lemma 1. Note that since $\frac{\partial(LHS - RHS)}{\partial \overline{p}} > 0$ for all $\overline{p}$, Equation (8) has a unique solution $\overline{p}$. Also,

$$\frac{\partial(LHS - RHS)}{\partial \delta} = \frac{-\sigma_0(R)}{\sigma_0(A) - \sigma_0(R)}(\overline{p} - \eta - p) - (h(\overline{p}) - \underline{V})) < 0.$$

and

$$\frac{\partial(LHS - RHS)}{\partial \eta} = -\frac{1 - \delta\sigma_0(R)}{\sigma_0(A) - \sigma_0(R)} < 0.$$

Thus it follows from the implicit function theorem that $\frac{\partial \overline{p}}{\partial \delta} > 0$ and

$$\frac{\partial \overline{p}}{\partial \eta} = \left[1 - \frac{\delta(\sigma_0(A) - \sigma_0(R))h'(\overline{p})}{1 - \delta\sigma_0(R)}\right]^{-1} \geq 1.$$

Furthermore, notice that $\frac{\partial h'(\overline{p})}{\partial \delta} = h''(\overline{p})\frac{\partial \overline{p}}{\partial \delta} \geq 0$. It follows that $\frac{\partial^2 \overline{p}}{\partial \delta \partial \eta} \geq 0$. This completes the proof of part (ii).

Finally, we prove part (iii) of Proposition 1. Rewrite Equation (8) as follows

$$(1 - \delta\sigma_0(R))(\overline{p} - \eta - p) - \delta(\sigma_0(A) - \sigma_0(R))(h(\overline{p}) - \underline{V}) = 0.$$

Applying the implicit function theorem to the above equation yields

$$\frac{\partial \overline{p}}{\partial \sigma_0(A)} = -\frac{-\delta(h(\overline{p}) - \underline{V})}{1 - \delta\sigma_0(R) - \delta(\sigma_0(A) - \sigma_0(R))h'(\overline{p})} > 0,$$

and

$$\frac{\partial \overline{p}}{\partial \sigma_0(R)} = -\frac{-\delta(\overline{p} - \eta - p - h(\overline{p}) + \underline{V})}{1 - \delta\sigma_0(R) - \delta(\sigma_0(A) - \sigma_0(R))h'(\overline{p})} < 0.$$

The first inequality follows because $h'(\overline{p}) \leq 1$. For the second inequality notice in addition that from Equation (8), $\frac{\overline{p} - \eta - p}{h(\overline{p}) - \underline{V}} = \frac{\delta(\sigma_0(A) - \sigma_0(R))}{1 - \delta\sigma_0(R)} < 1$. □

A.2 Shipper's strategies

First, we derive the shipper's per-period payoff. Each period in a maintained relationship has three possible outcomes: either the carrier accepts the offered load, the carrier rejects because of a high cost draw, or the carrier rejects because of a high spot rate. Thus, the per-period expected utility of the shipper in the relationship equals

$$\begin{aligned} u &= \underbrace{G(\overline{p})F(\overline{p})(\psi - p)}_{\text{accepted}} + \underbrace{G(\overline{p})[1 - F(\overline{p})](-\mathbf{E}[\tilde{p}_t | \tilde{p}_t \leq \overline{p}])}_{\text{rejected because of high cost}} + \underbrace{[1 - G(\overline{p})](-\mathbf{E}[\tilde{p}_t | \tilde{p}_t > \overline{p}])}_{\text{rejected because of high spot rate}} \\ &= -\mathbf{E}[\tilde{p}_t] + G(\overline{p})F(\overline{p})(\psi - p + \mathbf{E}[\tilde{p}_t | \tilde{p}_t \leq \overline{p}]). \end{aligned} \tag{9}$$

Notice that the term $\mathbf{E}[\tilde{p}_t|\tilde{p}_t \leq \overline{p}] \leq \mathbf{E}[\tilde{p}_t]$ is the shipper's expected payment were she to be served by the spot market conditional on the carrier being willing to accept the offered load. This term represents a selection effect: the carrier has the largest temptation to reject exactly when his acceptance is most valuable to the shipper.[62] This means that even when $\psi - p > -\mathbf{E}[\tilde{p}_t]$, a relationship that cannot induce sufficiently high level of cooperation may not be worth sustaining for the shipper. The following lemma provides a sufficient condition for the relationship to be worth sustaining for any incentive scheme with $0 \leq \sigma_0(R) < \sigma_0(A) \leq 1$.

Lemma 2. *If $\psi - p + \mathbf{E}[\tilde{p}_t|\tilde{p}_t \leq p] \geq 0$, then $u \geq -\mathbf{E}[\tilde{p}_t]$, that is, the shipper is better off offering to the carrier first than going directly to the spot market.*

Proof. Recall that

$$u = -\mathbf{E}[\tilde{p}_t] + G(\overline{p})F(\overline{p})(\psi - p + \mathbf{E}[\tilde{p}_t|\tilde{p}_t \leq \overline{p}]),$$

where $\mathbf{E}[\tilde{p}_t|\tilde{p}_t \leq \overline{p}]$ represents a selection effect. Define

$$\hat{p} = \inf\{p' \in \text{supp } G : \psi - p + \mathbf{E}[\tilde{p}_t|\tilde{p}_t \leq p'] \geq 0\}.$$

Then $u > \underline{U}$ and $\frac{\partial u}{\partial \overline{p}} > 0$ for all $\overline{p} > \hat{p}$. The shipper should opt out of the relationship if and only if the sustained level of cooperation satisfies that $\overline{p} < \hat{p}$. Under Condition 1 that $\psi - p + \mathbf{E}[\tilde{p}|\tilde{p} \leq p] \geq 0$, we have $\hat{p} \leq p \leq \overline{p}$, so the relationship is worth sustaining. □

We now derive the shipper's average discounted expected utility. Let

$$q = G(\overline{p})F(\overline{p})\sigma_0(A) + (1 - G(\overline{p})F(\overline{p}))\sigma_0(R) \tag{10}$$

denote the probability of maintaining the relationship next period calculated at the beginning of the current period's stage game. The average discounted expected utility U of the shipper in a maintained relationship is

$$U = (1 - \delta)u + \delta(qU + (1 - q)\underline{U}), \tag{11}$$

where $\underline{U} = \mathbf{E}[-\tilde{p}_t]$ is the expected payoff of the shipper going directly to the spot market. Thus,

$$U - \underline{U} = \frac{(1 - \delta)(u - \underline{U})}{1 - \delta q}. \tag{12}$$

[62]Note that acceptances require both a low spot rate and a low cost draw. However, under the assumption that cost draws are independent of spot rates, as in our model, the hypothetical expected payment in the spot market of an accepted load does not depend on the cost draw being low.

For $x \in \{\sigma_0(A), \sigma_0(R)\}$,

$$\frac{dU}{dx} = \underbrace{\left(\frac{\partial U}{\partial u}\frac{\partial u}{\partial \overline{p}} + \frac{\partial U}{\partial q}\frac{\partial q}{\partial \overline{p}}\right)\frac{\partial \overline{p}}{\partial x}}_{\text{incentive-inducing effect}} + \underbrace{\frac{\partial U}{\partial q}\frac{\partial q}{\partial x}}_{\text{regime-switching effect}}, \tag{13}$$

where $\partial U/\partial u, \partial U/\partial q, \partial u/\partial \overline{p}$ and $\partial q/\overline{p}$ are all positive.

A.2.1 Proof of Proposition 2

From Equation (13), it follows from $\partial \bar{p}/\partial \sigma_0(A) \geq 0$ and $\partial q/\partial \sigma_0(A) \geq 0$ that $\partial U/\partial \sigma_0(A) \geq 0$. Thus, $\sigma_0^*(A) = 1$. To see that the optimal punishment could be soft, refer to the single-lane relationships in Example 1.

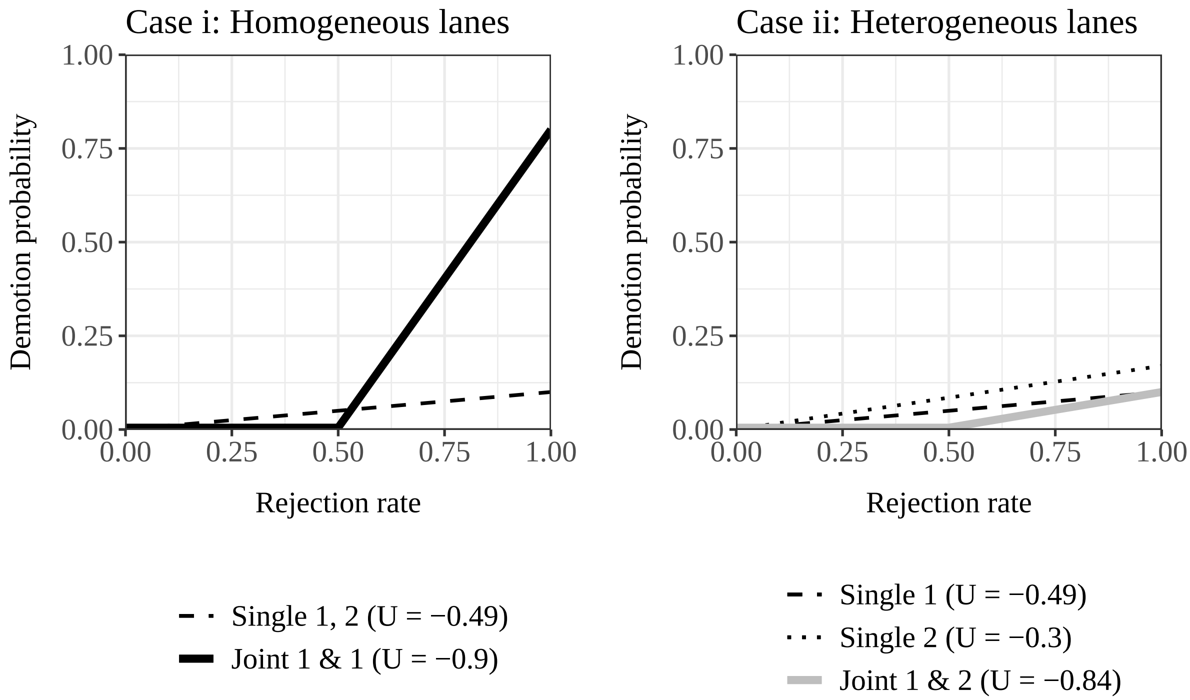
A.2.2 Details of Example 1

Under the chosen parameter values, the optimal single-lane punishment on lane 1 and lane 2 are soft. Specifically, when $\eta_1^\ell + \eta_2^\ell + p^\ell = 0.65$, the optimal demotion probability following a rejection is 0.1, and when $\eta_1^\ell + \eta_2^\ell + p^\ell = 0.8$, the optimal demotion probability following a rejection is 0.13. Define a multi-lane rejection rate as the average rejections across all lanes last period, $\frac{1}{2}\sum_{\ell=1}^{2} \mathbf{1}\{d_{t-1} = R\}$, and consider the shipper's incentive scheme that maps last-period average rejections to a probability of maintaining (or ending) both relationships.

The left panel of Figure 6 plots the optimal single-lane and multi-lane incentive schemes for case i. In this case, the optimal multi-lane incentive scheme forgives non-concurrent rejections but punishes concurrent rejections harshly. Specifically, if the carrier rejects on both lanes, the probability of demotion in the next period is 0.8. This multi-lane incentive scheme benefits the shipper over the optimal single-lane incentive scheme, $U = -0.9 > 2(-0.49)$, for two reasons. First, by forgiving non-concurrent rejections, the shipper allows the carrier to attain more allocative efficiency across the two lanes. This, in turn, increases the continuation value of the relationship for the carrier. Second, combining incentives with harsher punishments on joint rejections make these joint rejections low-probability events, and thus, clear signals of noncooperation.

The right panel of Figure 6 plots the optimal single-lane and multi-lane incentive schemes for case ii. The optimal multi-lane incentive scheme in this case takes a similar convex shape as that in case i, but punishes concurrent rejections softly. Moreover importantly, the shipper is strictly worse off by this scheme as compared to using the single-lane optimal incentive scheme, $U = -0.84 < (-0.49) + (-0.3)$. This example shows that using a simple incentive

Figure 6: Gains from pooling homogeneous lanes and losses from pooling heterogeneous lanes



scheme that conditions on average rejections (i.e., a common scorecard) across heterogeneous lanes might hurt the shipper.

B Additional empirical details

B.1 Construction of empirical variables

In this appendix, we explain the construction of right-hand side variables used in our analysis but the details of whose construction is omitted from the main text.

Inconsistency (loads / week) Our first empirical measure of the inconsistency of load timing captures a notion of how the number of loads varies from week to week within a month.

Let n^{ℓ}_{mw} denote the number of loads on lane ℓ in week w of month m. We then construct a measure of how much this count varies from week to week within a month by computing CV^{ℓ}_{m}, the coefficient of variation among $(n^{\ell}_{m1}, n^{\ell}_{m2}, n^{\ell}_{m3}, n^{\ell}_{m4})$. We then get a lane-level measure by averaging CV_m^ℓ over all active months on the lane:

$$\text{Inconsistency (loads / week)}^\ell = \frac{1}{M}\sum_m \text{CV}_m^\ell.$$

Inconsistency (day of week) Our second empirical measure of the inconsistency of load timing captures a notion how the timing of loads within a week varies from week to week (within a month). For instance, if a lane has 50% of its weekly loads on Monday and 50% of its weekly loads on Wednesday every week, the lane's loads would be perfectly consistent according to this measure. If, on the other hand, a lane's weekly loads were randomly allocated across days within each week, this lane's loads would be highly inconsistent according to this measure.

Our construction of this measure is motivated by a Chi-squared Goodness of Fit test of the null hypothesis that the lane's distribution of loads across days of the week is the same every week.

For any week w of month m and any day of the week d, the observed fraction of weekly volume on day of the week d is

$$O_{mwd}^\ell = \frac{n_{mwd}^\ell}{\sum_{d'} n_{mwd'}^\ell}.$$

Across all weeks within month m, the fraction of volume on day of the week d is

$$E_{md}^\ell = \frac{\sum_{w'} n_{mw'd}^\ell}{\sum_{w'}\sum_{d'} n_{mw'd'}^\ell}.$$

Under the null hypothesis, $O_{mwd}^\ell = E_d^\ell$ for all weeks w and days of the week d. The Chi-square statistic measures the deviation from this this null:

$$\chi_w^\ell = \sum_d \frac{\left(O_{mwd}^\ell - E_{md}^\ell\right)^2}{E_{md}^\ell}.$$

Then, we get our lane-level inconsistency measure by averaging across all active months and weeks:

$$\text{Inconsistency (day of week)}^\ell = \frac{1}{M}\sum_m \frac{1}{W}\sum_w \chi_w^\ell.$$

B.2 Simulation exercise

Using the estimated shipper's strategy and the mean tender acceptance probability, we simulate relationships under two scenarios. In the first scenario, the carrier accepts the first load of the relationship; in the second, he rejects the first load of the relationship. Each load $t > 1$ is accepted with probability 0.71, the average primary carrier acceptance rate. Based on each acceptance/rejection, we update the rejection rate and compute the probability of demotion between each load t and $t+1$. At each t, the relationship continues if the shipper is not demoted (which happens with probability $1 - \sigma_0(R_t)$, according to the estimated demotion strategy and the current the rejection index state) and if an RFP does not occur (which happens exogenously with constant probability $\frac{1}{83+1}$ (this RFP probability is chosen to match the average number of loads (83) between RFPs observed in the data). For each scenario, we run 10 million simulations.

C Additional empirical evidence

Shippers' (non)response to spot rates While carriers are tempted to defect from relationships to service the spot market when spot rates are high (Figure 1), we find no evidence that shippers face a similar temptation when spot rates are low. To show this, we run the following regression

$$\log(\text{Volume}^{\ell}_{scm}) = \beta_0^{\text{shipper}} + \beta_1^{\text{shipper}} \log(\text{Spot}^{\ell}_{m}) + \gamma^{\ell}_{sc} + \epsilon^{\ell}_{scm},$$

where $\text{Volume}^{\ell}_{scm}$ is the number of offers that shipper s sent to carrier c for service on lane ℓ within month m, and Spot^{ℓ}_{m} is the average spot rate on lane ℓ in month m. We include shipper-carrier-lane fixed effects γ^{ℓ}_{sc} to absorb differences in contract rates, spot rates, and volumes across relationships. If shippers tend to skip the routing guide and go directly to the spot market when spot rates are low, we should expect $\beta_1^{\text{shipper}} > 0$. In contrast, we find that $\hat{\beta}_1^{\text{shipper}} = -0.0363$ with a standard error of 0.0119. That is, our estimate of β_1^{shipper} is economically very small. We thus conclude that unlike carriers, shippers appear non-responsive to spot temptation.

Relationship characteristics across carrier types Our empirical evidence, as summarized by Table 4, shows substantial heterogeneity by carrier type in the nature of within-relationship interactions between shippers and carriers. Table 6 shows that key relationship characteristics determined at the formation stage of relationships also differ systematically by carrier type. First, large asset-based carriers have lower contract rates than small asset-based carriers and brokers, suggesting potential cost advantages of large asset-based carriers. Second, compared to an asset-based carrier, a broker tends to serve as the primary carrier for the same shipper on more lanes, but these lanes tend to have lower volume. This tendency helps explain why shippers use multi-lane punishments with brokers: First, low-volume lanes creates the need to pool incentives across lanes. Second, a shipper and broker interacting on many lanes means that the scope for incentive pooling is larger.

Table 6: Relationship characteristics by carrier type

	Large ABCs	Small ABCs	Brokers
Monthly lane volume	9.14	8.73	6.35
Contract rate - spot rate	-0.125	-0.0878	-0.0432
Inconsistency (loads / week)	1.04	1.12	1.17
Inconsistency (day of week)	0.615	0.546	0.555
Number of lanes	6.91	4.13	11.61

Notes: This table shows differences in relationship characteristics across carrier types. The first four rows give the means of four key shipper-carrier-lane characteristics. The last row indicates the average number of lanes comprising a shipper-carrier relationship.

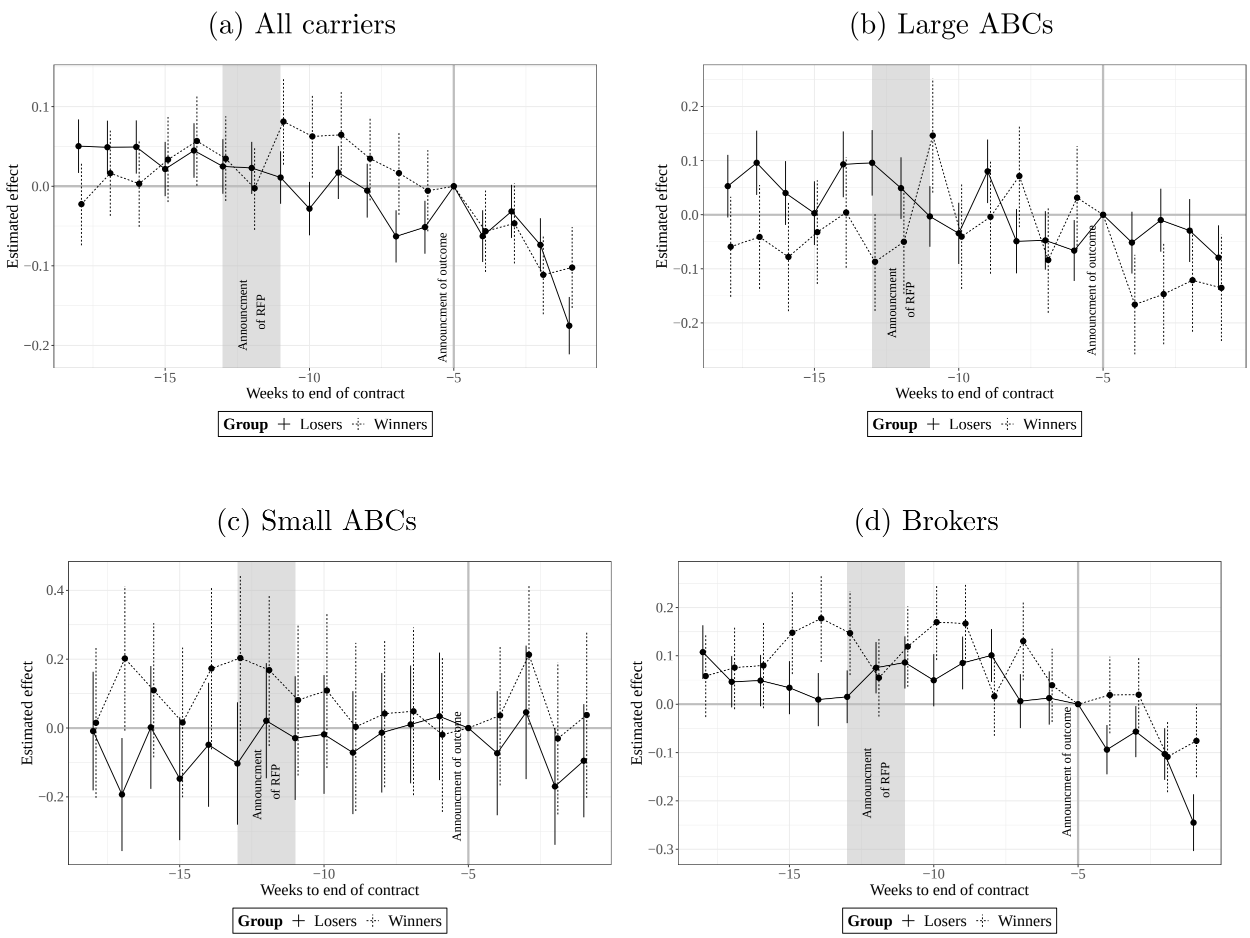
Role of relationship-specific investments In the truckload freight setting, relationship-specific investments could take the form of a carrier adjusting his network of truck movements to best service a contracted lane. For example, a carrier can reduce empty miles by lining up a backhaul, thus making accepting forehaul loads less costly.[63] Among the three groups of carriers, we expect that such adjustments matter most for small ABCs; large ABCs have more flexibility in network adjustments due to their larger fleets, and brokers, who find carriers on the spot market, do not have similar concerns for empty miles because they do not incur the cost of such empty miles. Our analysis on carriers' acceptance (Table 3) shows evidence consistent with this hypothesis. Only small ABCs have lower acceptance rates in the first seven days after being promoted from backup to primary status.

A breakdown of our endgame effect analysis (Section 5.1) by carrier type further supports the hypothesis that relationship-specific investments do not play an important role in the behavior of large ABCs and brokers. Figure 7 shows that following the announcement of auction outcomes, the acceptance rate of both large ABCs and brokers, whether they won

[63]The terminology of "forehaul" and "backhaul" is widely used in the industry. If a carrier has a contract for the lane from A to B (the forehaul), he would ideally also have a contract providing regular demand on the backhaul from B back to A.

or lost the auction, decreases significantly. This behavior cannot be driven by relationship-specific investments, including network adjustments. The reason is that winning carriers—who continue to be primary carrier the same lane—should have no incentives to undo their investments toward the end of the current contract period.

Figure 7: End-of-contract effects by carrier type



Notes: Like Figure 4, this figure plots the estimated coefficients $\{\alpha_k\}$ from equation (1) with the normalization $\alpha_5 = 0$ is imposed. Panel (a) is same as Figure 4, while panels (b)-(d) present separate estimates by carrier type.

Table 7: Response of auction outcomes to carrier behavior

	Wins RFP	New contract premium
Rejection rate	-0.0676	-0.430
	(0.0239)	(0.0798)
Observations	1673	373

Notes: Standard errors in parentheses. This table reports estimated coefficients for regressions of the form $y^{\ell}_{scr} = \delta_0 + \delta_1 \text{Rejection rate}^{\ell}_{scr} + \delta_2 \text{Miles}^{\ell} + \epsilon^{\ell}_{scr}$ where Rejection rate$^{\ell}_{scr}$ is the proportion of load offers rejected by carrier c from shipper s on lane ℓ during the contract period and Miles$^{\ell}$ is the length of lane ℓ. In the first column, the outcome variable y^{ℓ}_{scr} is an indicator for whether the primary carrier c wins RFP r and therefore maintains the primary position on lane ℓ. In the second column, the outcome variable y^{ℓ}_{scr} is the new contract rate of carrier c after the RFP (conditional on winning the RFP and maintaining the primary position). In keeping with the event study regressions above, both regression samples are limited to mass RFP events. The sample for the second column is further limited to primary carriers who win the RFP.

Table 8: Performance of promoted carrier relative to demoted carrier

	All types		Large ABCs	
	Acceptance	Contract rate	Acceptance	Contract rate
	-0.0122	0.0911	-0.150	0.0448
	(0.00244)	(0.0275)	(0.00654)	(0.0813))
N	1009352	1009352	177030	177030

	Small ABCs		Brokers	
	Acceptance	Contract rate	Acceptance	Contract rate
	-0.184	-0.185	0.115	-0.000292
	(0.0105)	(0.0903)	(0.00451)	(0.0159
N	80686	80686	277878	277878

Notes: This table presents estimates from the following regression:

$$y^{\ell}_{sct} = \sum_{g \in \mathcal{G}} \beta_g \mathbb{1}\{g^{\ell}_{sct} = g\} + \underbrace{\gamma \left(\text{Spot rate}^{\ell}_{t} - \text{Contract rate}^{\ell}_{sct} \right)}_{\text{only included for } y=\text{Acceptance}} + \epsilon^{\ell}_{sct}$$

where $\mathcal{G}$ represents a set of three groups of primary carriers: (a) RFP winners who are never demoted, (b) RFP winners who are eventually demoted, and (c) non-RFP winners who are promoted. The estimates listed in the table state the difference $\beta_{(c)} - \beta_{(b)}$. They therefore have the interpretation of how much the expectation of the outcome y changes when the shipper demotes the RFP winner and promotes his replacement. Outcomes are (1) an indicator for carrier c accepting load t and (2) carrier c's contract rate. The results for all carrier types indicate that demoting the primary carrier tends to result a small (1.2pp) decrease in the acceptance rate while resulting in a sizeable (9 cent per mile) increase in the contract rate.

Chapter 2

Long-Term Relationships and the Spot Market: Evidence from US Trucking*

Adam Harris† Thi Mai Anh Nguyen‡

1 Introduction

Long-term relationships are a ubiquitous feature of the economy (Macaulay, 1963; Macchiavello, 2022), and in many settings, these relationships coexist and interact with other forms of transactions (Allen & Wittwer, 2021; Macchiavello & Morjaria, 2015). While the ubiquity of relationships suggests that they benefit the participating parties, it does not rule out the possibility that relationships exert negative externalities on the rest of the market. Long-term relationships are individually rational, but are they socially efficient?

We develop an empirical framework to quantify the market-level value of long-term relationships and alternative institutions in the setting of the US for-hire truckload freight industry, one in which long-term relationships coexist with spot marketplaces. Combining

*This research was made possible by the many people in the truckload freight industry who generously shared their time, insights, and data with us. We are also very grateful to Angi Acocella, Nikhil Agarwal, Chris Caplice, Glenn Ellison, Bob Gibbons, Stephen Morris, Nancy Rose, Tobias Salz, and Mike Whinston for their advice and comments. We also thank all participants in the MIT Industrial Organization Lunch and Seminar, the MIT Organizational Economics Lunch and Seminar, and the MIT Freight Lab for their comments and feedback. We acknowledge the support of the George and Obie Shultz Fund. While this paper makes use of data provided by CH Robinson and DAT, the opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of CH Robinson or DAT.
Harris: This material is also based upon work supported by the National Science Foundation Graduate Research Fellowship under Grant No. 1745302. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.
Nguyen: This material is based upon work supported by the Jerry A. Hausman Graduate Dissertation Fellowship.

†MIT Department of Economics. Email address: asharris@mit.edu.
‡Corresponding author. MIT Department of Economics. Email address: anhng@mit.edu.

elements from the auction and dynamic discrete choice literature, our model captures both the formation of and interactions within long-term relationships. The demand side (shippers) and the supply side (carriers) form relationships via auctions, facilitating the formation of relationships with high match quality. Within relationships, carriers face a temptation to defect to the spot market when spot rates are high, and shippers use a relational incentive scheme to mitigate this problem. At the market level, we allow for two-way crowding-out effects between long-term relationships and the spot market (Kranton, 1996), finding strong effects in both directions. On the one hand, the spot market creates a moral hazard problem in long-term relationships, crowding out low-value relationships. On the other hand, the formation of long-term relationships results, in aggregate, in a thin and substantially less efficient spot market. We quantify this tension in two counterfactual exercises: (i) centralizing all transactions into a spot market for maximal spot market thickness and (ii) using index-priced contracts in long-term relationships to resolve the moral hazard problem.

The US for-hire truckload freight industry is an important economic setting and one in which long-term relationships play a central role. In 2019, this industry generated revenues equivalent to 0.8% of US GDP and transported 72% of domestic shipments by value, playing an integral part in US domestic trade. Consequently, the outcomes in this industry have implications for supply chains and the goods economy as a whole. In this market, 80% of total transacted volume is accounted for by long-term relationships. Relative to spot transactions, long-term relationships may allow their participants to enjoy a variety of possible benefits: more reliable service; more seamless loading, docking, insurance, and payment processing; better planning; or lower transaction costs.[1] While our empirical framework does not decompose relationship benefits into specific channels, we quantify the total benefits of relationships, the extent to which these benefits are realized, and how the realized benefits are shaped by market structure.

The US for-hire truckload freight industry offers an ideal empirical setting for studying long-term relationships and their interactions with the spot market. In this setting, shippers are firms that demand transportation service on some origin-destination pair ("lane"). Carriers are transportation firms that are hired to provide such service. Within a long-term relationship, a contract fixes the price ("rate") but not the volume. This leaves room for relational incentives to govern transactions. In many settings, a lack of data on interactions within informal relationships poses a significant obstacle to studying such relationships. Our setting, however, is one in which such interactions between shippers and carriers leave a

[1] In this setting, shippers (the demand side) may care about reliability of service, efficiency at loading and docking, or ease of communication; carriers (the supply side) may care about the ability to find backhauls, familiarity with facilities, or promptness of payment. See Hubbard (2001) and Masten (2009) for a discussion on non-price factors that matter in this industry.

digital record. Specifically, shippers use transportation management systems (TMS) to automate many aspects of their relationships with carriers; such systems record shippers' offers, carriers' responses, and the status of relationships at each point in time. Our study uses an anonymized panel of relationships of shippers that use one of the TMSs. To capture the outside option of long-term relationships, we obtain data on spot rates and volumes at fine spatial and temporal granularity. The combined data set allows us to have both a microscopic view of individual relationships and a bird's-eye view of their aggregate effects on the spot market.

We start in Section 4 with key patterns in the data. First, we find large heterogeneity in shippers' and carriers' behaviors across relationships after conditioning on contract and spot rates. This suggests the role of non-price factors in generating match-specific gains in long-term relationships. Second, consistent with Hubbard (2001), we find that spot arrangements take a larger share of total market volume on lanes with higher total demand.[2] A potential explanation for this result is that lanes with higher total market demand have the potential for achieving higher spot market thickness, which, in turn, increases the relative attractiveness of spot arrangements. The latter link means that long-term relationships crowd out the spot market by making it thinner and less attractive.

In Section 5, we develop a model that allows us to quantify the tension between realizing relationships' match-specific gains and maintaining spot market efficiency. The modeling of individual relationships combines elements of models from the auction and dynamic discrete choice literature. In our model, each relationship consists of two stages. In the first stage, the shipper holds an auction to select a carrier with whom to form a relationship. In the second stage, the shipper and the winning carrier interact in a repeated game. In each period of this game, the shipper decides whether to terminate the relationship or maintain the relationship and offer a load; the carrier decides whether to accept the load, reject it for a spot offer, or reject it to remain idle. Motivated by the data patterns established in Section 4, we allow for two-sided match-specific gains from relationships and capture the link between spot market thickness and efficiency via a search cost that the carrier incurs from servicing the spot market. Motivated by evidence from Harris and Nguyen (2021), we model the shipper as using an incentive scheme that conditions the probabilistic termination of relationships on carriers' past rejections. The carrier responds optimally in each period, taking into account the current and future compensation for an accepted offer, the compensation and search cost in the spot market, and the operational cost for the current period. At the market

[2]While Hubbard (2001) exploits equilibrium supply-side variation, we exploit the predicted trade flows between different states of the US (Caliendo, Parro, Rossi-Hansberg, & Sarte, 2018) as an exogenous source of demand-side variation.

level, long-term relationships and the spot market interact in two ways. First, spot rates are determined in equilibrium, absorbing both direct spot demand and rejected offers from long term relationships. Second, search costs on the spot market are determined endogenously by the equilibrium spot volume.

In Section 6, we show that shippers' and carriers' primitives are nonparametrically identified from their behaviors in the auction and repeated game. Our identification strategy illustrates how insights into both the formation of and interactions within relationships help recover a rich set of model primitives. In three sequential steps, we identify *carriers' primitives* from their dynamic play in the repeated game. The first step identifies the distribution of the sum of search and operational costs. The argument for the identification of this distribution is motivated by the following thought experiment: If a carrier could only decide between "spot" and "idle", the probability that this carrier chose "spot" over "idle" at different spot rates would trace out the distribution of the sum of search and operational costs. However, the empirical challenge in our setting is that we only observe whether the carriers in long-term relationships—who decide between "accept", "spot", and "idle"—accept or reject. To overcome this challenge, we develop a support-based argument that relates the unobserved decision margin (between "spot" and "idle") to the observed decision margin (between "accept" and "reject"), thereby locally recreating the hypothetical scenario. The second step decomposes search and operational costs from their sum. Here we establish causality between search costs and spot market thickness by exploiting the predicted trade flows between different states of the US (Caliendo, Parro, Rossi-Hansberg, & Sarte, 2018) as a demand shifter. The third step identifies carriers' match-specific gains from their acceptance probability, observed prices, and the cost parameters identified in the previous steps. Intuitively, a carrier's match-specific gain can be inferred from the level of spot rate at which this carrier's acceptance becomes responsive to spot rates.

Next, we exploit the fact that relationships are formed via auctions to identify the distribution of *shippers' match-specific gains*. Intuitively, the way that carriers' match-specific gains are reflected in their bidding is determined by how contract rates split total match-specific gains into carriers' rents and shippers' rents. Building on this observation, we derive, under empirically plausible conditions, equilibrium conditions that give rise to a monotone mapping between carriers' rents (whose components are either observed or already identified) and shippers' rents (which we need to identify). In the spirit of Guerre, Perrigne, and Vuong (2000), we pin down this monotone mapping from the first-order condition of carriers' bidding in the space of carriers' rents. Finally, identified rents and observed contract rates recover match-specific gains.

In Section 7, we present our estimates of the model primitives, showing that the key tension in our setting is between the large benefits of long-term relationships to the participating parties and their substantial negative externalities on the spot market. On the one hand, we find that each shipper and carrier in long-term relationships enjoys an average premium over a spot transaction of 58% and 10%, respectively, for each realized transaction. Moreover, current fixed-rate contracts and relational incentive schemes capture these potential premiums fairly well. Specifically, the current relationships achieve, on average, 44% of the relationship-level first-best surplus, with large heterogeneity across relationships with different match quality. Underlying this heterogeneity is the fact that, in relationships with lower match quality, carriers' moral hazard is more severe, and relational incentives are less effective at mitigating this problem. In other words, the spot market crowds out in particular long-term relationships with low match quality. On the other hand, relationships' formation and high performance result in a thin spot market with substantially higher search costs. We estimate that doubling the thickness of the spot market on a lane reduces search costs by an amount equivalent to reducing operational costs by 29%.

Motivated by these findings, Section 8 evaluates two counterfactual institutions: (i) a centralized spot market for optimal spot market efficiency and (ii) individually first-best contracts for optimal performance in long-term relationships. Comparing the current institution to the first counterfactual institution suggests that the current dominance of long-term relationships is not due to a coordination failure to form a thick spot market but instead due to the considerable benefits of long-term relationships. Specifically, we find that centralizing all transactions into a spot market results in substantial welfare loss, equivalent to about 25% of the median operational cost. In theory, a centralized spot market has two potential sources of gains: (i) reduction in search costs and (ii) improvements in allocative cost efficiency. However, our estimates suggest that both of these gains are small. One reason is that while substantially reducing search costs on the spot market, a centralized spot market increases search costs for those who would otherwise be in relationships and not incur any search costs. Overall, cost reductions from centralizing all transactions are not nearly enough to compensate for the complete loss of match-specific gains from long-term relationships.

In the second counterfactual exercise, we replace fixed-rate contracts with index-priced contracts designed to achieve the first-best welfare for individual relationships. Comparing the market-level performance of these contracts highlights the key tradeoff in our setting: any attempt to improve the performance of long-term relationships would worsen their negative externalities on the spot market. Specifically, while these contracts increase the realized relationship benefits by 11% to 28%, such gains are roughly offset by a substantial increase in search costs in the spot market and a reduction in allocative cost efficiency. Only in periods of high demand, when fixed-rate contracts face serious moral hazard, does the former effect dominate, leading to welfare gains from index-priced contracts. Overall, both fixed-rate and index-priced contracts perform fairly well at the market level, achieving at least 40% of the market-level first-best surplus for medium trips of around five hundred miles and at least 60% of the market-level first-best surplus for long trips of around a thousand miles.

The paper proceeds as follows. Section 2 reviews the related literature. Section 3 presents institutional details. Section 4 describes our data and presents the data patterns that motivate our model. Section 5 describes our model. Section 6 explains our identification argument and the estimation procedure. Section 7 presents our estimates of key model primitives, and Section 8 presents counterfactual results.

2 Literature review

Our paper contributes to two empirical literatures—the literature on long-term informal relationships and the literature on trucking—with an empirical framework that combines tools from the auction and dynamic discrete choice literatures. We divide our literature review into three subsections relating to literatures on long-term relationships, trucking, and spot market efficiency.

Long-term informal relationships. The empirical literature on long-term informal relationships has developed a rich set of insights into the mechanisms through which relationships create value for participating parties and respond to external factors.[3] We contribute to this literature by quantifying both the value of relationships to participating parties and the negative externalities of relationships on the spot market. Conceptually, our paper is most closely related to Kranton (1996), who uses a market equilibrium model to theorize two-way crowding-out effects between long-term relationships and the spot market. The first direction is, as argued by Baker, Gibbons, and Murphy (1994), that the spot market is the outside option of relationships, crowding out long-term informal relationships by making relational incentives harder to enforce. The other direction takes place at the market level. As relationships are formed, the spot market becomes thinner and less efficient; this, in turn, reinforces the relative attractiveness of relationships.[4] To the best of our knowledge, we are

[3] For example, value creation in long-term informal relationships can arise from supply reliability (Adhvaryu, Bassi, Nyshadham, & Tamayo, 2020; Cajal-Grossi, Macchiavello, & Noguera, 2022), reputation building (Macchiavello & Morjaria, 2015), or relational adaptations (Barron, Gibbons, Gil, & Murphy, 2020). Relationships may terminate (Macchiavello & Morjaria, 2015) or restructure (Gil, Kim, & Zanarone, 2021) in the face of large shocks, and can be hampered by competition (Macchiavello & Morjaria, 2021).

[4] Tunca and Zenios (2006) make a similar theoretical argument by examining the competition between procurement auctions and long-term relationships. The broad idea that market thickness can be self-fulfilling is examined in other settings. For example, Ngai and Tenreyro (2014) show that thick-market effects can

the first to quantify this second direction.

Methodologically, we contribute an empirical framework that takes advantage of both the formation of and interactions within long-term relationships to recover a rich set of primitives. Specifically, we build on techniques in the dynamic discrete choice literature for both long panels (Rust, 1994) and short panels (Kasahara & Shimotsu, 2009) to recover model primitives on the carriers' side. Our problem is not standard in this literature but closely related to the literature on contracting with moral hazard (Perrigne & Vuong, 2011), in that payoff-relevant actions are not fully observed. Rather than relying on a mapping between unobserved actions and observables (Gayle & Miller, 2015) or the state transition process (Hu & Xin, 2021), we develop a support-based argument that relates the unobserved decision margin to the observed decision margin. We then adapt techniques from the empirical auction literature (Guerre, Perrigne, & Vuong, 2000) to recover primitives on the shipper' side, factoring in the equilibrium path of play in each potential relationship.

A few other papers quantify the value of long-term relationships, but they differ from our paper both methodologically and conceptually. Macchiavello and Morjaria (2015) exploit temporal variation in spot rates to bound the value of trading relationships in the Kenyan rose market. Our empirical strategy builds on their idea that variation in spot rates helps trace relationship value but differs in that we recover primitives that can be used for counterfactual analysis. A recent paper that recovers the primitives of relationships and performs counterfactual analysis is Brugues (2020), which studies the trading relationships in the manufacturing supply chain of Ecuador. This paper exploits the optimality conditions of dynamic contracting with flexible monetary transfers to identify the distribution of buyer's type. Since relationships in our setting use fixed-rate contracts, we cannot apply similar methods. Instead, we exploit the formation of relationships via auctions and the rich dynamics within relationships to identify the distribution of two-sided match-specific gains from relationships. Startz (2021) is among the few papers, including ours, that take a market equilibrium approach and capture both the formation of and interactions within relationships. This paper quantifies the search and contracting frictions faced by Nigerian importers of consumer goods by exploiting importers' decisions to travel to the source country as a way of reducing both frictions. A key difference between our paper and Startz (2021) is that we identify and quantify the thickness-externality of long-term relationships on the spot market.

amplify the seasonality of the housing markets; in the setting of labor markets with match-specific quality, Elliott (2014) argues that thick-market effects give rise to multiplicity of search equilibria, subobtimal entry and market fragility.

Trucking. Exploration of long-term contracts and spot arrangements in trucking dates back to Hubbard (2001), who finds that selection into spot transactions increases with market thickness, and Masten (2009), who argues that savings on transaction costs are an important driver for long-term contracts.[5] Search costs in our model have a similar interpretation as transaction costs in Masten (2009); both increase the relative attractiveness of relationships. One additional insight from our paper is that these costs increase endogenously when more relationships are formed and the spot market becomes thinner. This insight also offers an additional explanation for the link between market thickness and the use of spot arrangements found in Hubbard (2001). On thicker lanes, there is more potential for spot market thickness, so search costs tend to be lower, which in turn increases the relative attractiveness of spot arrangements.

Since these papers, there have been significant improvements in how the trucking industry organizes shipper-carrier matching and how firms keep track of their interactions within relationships. These improvements have generated rich transaction-level data, from which our paper benefits. In our previous paper, Harris and Nguyen (2021), we establish an understanding of the nature and effects of dynamic incentives in long-term relationships in the US trucking industry. Our current work builds on such understanding, but the goal is to quantify the market-level tradeoff between long-term relationships and the spot market. We examine this tradeoff in two counterfactual exercises: (i) a centralized spot market for maximal spot market thickness and (ii) first-best contracts for optimal performance within long-term relationships. Our paper is the first to study two-way market-level interactions between long-term relationships and the spot market in a transportation setting. A series of papers in the transportation and logistics literature also uses the same data as our paper to study the effects of relationships on participating parties, examining reciprocity (Acocella, Caplice, & Sheffi, 2020), factors that affect the value of relationships to carriers (Acocella, Caplice, & Sheffi, 2022b), and potential Pareto improvements for participating parties from index pricing (Acocella, Caplice, & Sheffi, 2022a). In the economics literature, Yang (2021) studies the home bias of truck drivers using a market equilibrium model but focuses exclusively on the spot market.

Spot market efficiency. The benefits and, to a lesser extent, the tradeoffs of centralizing all transactions into a spot platform have been examined in other settings. In other

[5] A related strand of literature studies asset ownership in the trucking industry. For example, Baker and Hubbard (2003) study shippers' choices over private fleets versus for-hire carriers; Baker and Hubbard (2004) and Nickerson and Silverman (2003) study drivers' ownership of trucks. Another strand of the literature (Marcus, 1987; Rose, 1985, 1987; Ying, 1990) studies the effects of deregulation on the trucking industry. These papers provide an important historical context for the evolution of this industry.

transportation markets, the benefits include the reduction of search costs from economies of density (Frechette, Lizzeri, & Salz, 2019) and the correction of spatial misallocation via platform pricing (Buchholz, 2022; Lagos, 2000, 2003). Tradeoffs include the platform exploiting its market power to extract surplus from both sides of the market (Brancaccio, Kalouptsidi, Papageorgiou, & Rosaia, 2020; Rosaia, 2020). The key difference between these papers and ours is the central role of long-term relationships in our setting. Specifically, we model economies of density as the channel through which long-term relationships exert externalities on the spot market. We abstract from spatial misallocation, focusing instead on the (mis)allocation of transactions between long-term relationships and the spot market.

Multiple modes of transaction. The coexistence of multiple modes of transaction has been examined in various settings. Such coexistence could generate quality dispersion (Galenianos & Gavazza, 2017), and market frictions could shift market modality (Gavazza, 2010, 2011). In the setting of the liquefied natural gas industry, Zahur (2022) shows that formal long-term contracts increase investment incentives but reduce firms' ability to respond to demand shocks, overall hurting allocative efficiency. Our paper is similar to Zahur (2022) in that we quantify the tradeoffs between the benefits of long-term relationships and their negative externalities on allocative efficiency, but differs in that we highlight an additional channel of such externalities. That is, the formation and high performance of long-term relationships increase spot market frictions by making the spot market thinner.

In financial settings, liquidity externalities—the fact that entry decisions by some participants increase the depth and liquidity of a market, thereby making that market more attractive to other participants—mirror the market-thickness externalities in the truckload setting. An implication of these liquidity externalities is that the prevalence of over-the-counter markets could be self-fulfilling while socially inefficient (Admati & Pfleiderer, 1988; Biais & Green, 2019; Pagano, 1989). Other arguments for the prevalence of over-the-counter markets include information asymmetry (Collin-Dufresne, Hoffmann, & Vogel, 2019; Lee & Wang, 2018), dealer heterogeneity (Dugast, Üslü, & Weill, 2019), and barriers to entry and insufficient competition on existing platforms (Allen & Wittwer, 2021). Our paper differs from these papers in the importance of match-specific gains from long-term relationships in our setting. In fact, we find that centralizing all transactions for truckload service results in substantial welfare loss, precisely because of the complete loss of these match-specific gains.[6]

[6]In the market for Canadian government bonds, Allen and Wittwer (2021) study the coexistence of a centralized platform and investor-dealer relationships. However, their analysis largely abstracts from match-specific gains from these relationships.

3 Institutional details

The US for-hire truckload freight industry offers an ideal setting to study the functioning of long-term relationships and their market impact. It is an economically important industry and one in which long-term relationships and spot arrangements coexist. The former are a central feature, with sophisticated institutions built around both the formation and management of relationships. This section provides institutional background and describes the nature of long-term relationships in this setting.

3.1 The US for-hire truckload freight industry

Trucking is the most important mode of transportation of US domestic freight. In 2019, trucks carried 64% of domestic shipments by weight and 72% of domestic shipments by value.[7] There are four main segments within the US trucking industry, separated by governance structure and size of shipments: truckload for-hire fleets, truckload private fleets, less-than-truckload, and parcel.

Our focus is on the for-hire truckload segment of the US freight trucking industry. In this segment, a *shipper* (e.g., manufacturer, wholesaler, or retailer) with a *load* (shipment) to be transported on a *lane* (an origin-destination pair) on a specified date needs to hire a *carrier* (e.g., trucking company) for that service. This is in contrast to private fleets, which are vertically integrated carriers serving a single shipper.[8] In terms of shipment size, a shipment of a *truckload* carrier fills the entire truck. These carriers are concerned about reducing miles traveled empty, and thus unpaid, but not about how to optimally combine shipments to fill up their trucks. The latter is a key concern of less-than-truckload and parcel carriers. This means that truckload carriers face simpler routing decisions and rely less on economies of scale, a difference partially responsible for why the truckload segment is more fragmented than other segments (Ostria, 2003). The top 50 truckload fleets account for only about 10% of the segment's total revenue, and about 90% of truckload fleets have fewer than six trucks.[9]

Within the for-hire truckload segment of the US trucking industry, shipments are also separated by distance and trailer type. Long-hauls are shipments on lanes greater than 250 miles long. On such lanes, the ability of carriers to find backhauls is important (Hubbard,

[7] These statistics are calculated using data from the Bureau of Transportation statistics.

[8] Such vertical contractual arrangements tend to be chosen by companies that prioritize quality and reliability of service, and typically have a dense network of truck movements that allow for efficient routing. In fact, shippers and carriers may maintain a portfolio of contractual arrangements (Acocella, Caplice, & Sheffi, 2022a). For example, shippers with private fleets can sell excess capacity on their "backhauls", and for-hire carriers can dedicate some of their capacity to some shippers.

[9] See https://medium.com/@sambokher/segments-of-u-s-trucking-industry-d872b5fca913.

2001). The common types of trailers are dry van, refrigerated, flatbed, and tanker. Our paper focuses exclusively on long-haul dry van truckload services.

Shippers and carriers in the US for-hire truckload industry engage in two main forms of transactions: long-term relationships and spot arrangements. Long-term relationships dominate this market, capturing 80% of total transacted volume; spot arrangements account for the remaining 20%.[10]

It is important to note that transactions on the spot market typically involve search and haggling. For example, shippers and carriers can post and search for available loads and trucks on electronic load boards. These load boards are marketplaces from which both sides can obtain contact information of potential matches, but rate negotiations are conducted offline.[11] Shippers and carriers can also be matched on digital matching platforms, which employ real-time matching and pricing, or via brokers. For our purpose, we treat all of these channels as a single spot market with search costs that potentially vary with spot market thickness.

3.2 Long-term relationships

There is an organized process that forms and manages long-term relationships in this setting, but contracts between shippers and carriers within their relationships are largely incomplete. The remarkably rich available data on how shippers and carriers form relationships and interact within relationships, together with the informal nature of these relationships, makes it an appealing empirical setting for our study.

Relationship formation. Long-term relationships are formed via procurement auctions. These auctions begin with shippers sending requests for proposals to different carriers, detailing their needs.[12] Each carrier then submits a bid on a fixed contract rate to be charged on each load that the carrier transports for the shipper within the contract period, which is typically one or two years. Contract rates are accompanied by a fuel program, typically proposed by shippers, that compensates carriers for changes in fuel costs.[13] The shipper then chooses a *primary carrier* and a set of backup carriers in case transactions with the primary

[10] See https://www.freightwaves.com/news/what-is-the-difference-between-trucking-contract-and-spot-rates.

[11] Figure 23 in Appendix E shows the search interface of DAT load board, the dominant load board for for-hire truckload service.

[12] Typically, shippers send out requests for proposals on multiple lanes simultaneously and carriers are free to bid on a subset of them.

[13] The most common fuel program calculates per-mile fuel surcharge as the per-mile difference between a fuel index and a peg, (index − peg)/escalator, where "escalator" (miles/gallon) is a measure of fuel efficiency. In practice, variation in the choice of the index, the peg and the escalator has little impact on shippers and carriers. For more details, see https://www.supplychainbrain.com/ext/resources.

Table 1: Example load offers: Shipper Z, lane City X - City Y (on June 1, 2018)

Type	Order	Carrier	Rate ($/mile)	Decision
Primary carrier	1	A	1.60	Reject
Backup carriers	2	B	1.44	Reject
	3	C	1.72	Accept
	4	D	1.89	

Notes: The first offer was sent to carrier A at the contracted rate of $1.60/mile; since A rejected, an offer was sent to B at the contracted rate of $1.44/mile; since B also rejected, an offer was sent to C at $1.72/mile; since C accepted, no offer was sent to D.

carrier do not materialize.[14] Our analysis focuses on the relationships between shippers and their primary carriers.

Relationship management. Every interaction between a shipper and her carriers is automated and recorded by a Transportation Management System (TMS). When the shipper needs to transport a load on a lane, she inputs details of the load into her TMS, which automatically sends out offers to the carriers, sequentially in the order of their ranks, until one carrier accepts. Most carriers take less than one hour to respond to an offer.[15] This process of sequential offerings is sometimes referred to as a "waterfall" process in other settings. Table 1 depicts an instance of this process.

While the auction determines an initial ranking of the carriers, the shipper can reorder this ranking at any point within the contract period. A "routing guide" keeps track of carriers' updated ranks. The primary carrier is the top-ranked carrier, receiving most of the offers and typically accepting most of them.[16] While backup carriers do not necessarily know their exact ranks, primary carriers know that they are top-ranked. This is because carriers need to plan ahead if they expect to service a large number of loads.

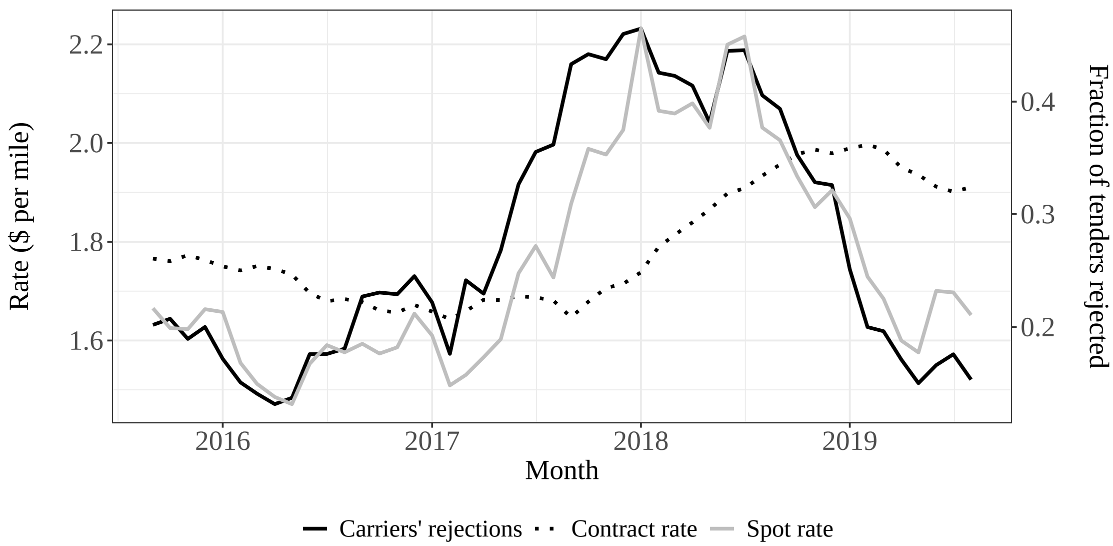
A key and unique feature of this setting is that contracts between shippers and carriers fix rates, but not volume. Shippers can influence the number of offers that each carrier receives using their control over the routing guide, and carriers also face no legal recourse when rejecting loads. Such incompleteness in the contracts between shippers and carriers leaves room for potential opportunistic behaviors and for relational incentives to mitigate

[14]See Caplice (2007) for more details on this procurement process.

[15]The median response time is 41 minutes and 90% of all responses are within two hours. The full waterfall process typically takes less than three hours to complete.

[16]Sometimes, due to capacity constraints and other factors, offers are sent to the backup carriers first. See Appendix A for more details.

Figure 1: Spot market is the outside option of relationships



such behaviors.

Moral hazard and incentive schemes. In Harris and Nguyen (2021), we study the nature of the interactions between shippers and carriers within long-term relationships in this setting. Our two key findings are that (i) carriers can and do reject offers within relationships to take advantage of higher prices offered in the spot market, and (ii) shippers use a relational scheme to mitigate such temptation.

Figure 1 plots the movements of spot rates (in gray), average contract rates (dotted), and carriers' rejections (solid) over the period of our data sample.[17] There is large temporal variation in spot rates. The market started soft with spot rates below contract rates, tightened over 2017 to reach its peak in 2018 with spot rates well above contract rates, and cooled down towards the end of the period. Average contract rates do adjust to spot rates, though only partially and with some lag. This means that there are some periods in which spot rates are much higher than contract rates. In such periods, carriers much more frequently decline shipment requests within relationships.

These data patterns suggest that the spot market is an important outside option for relationships. On the intensive margin, high premiums of spot rates over contract rates create a temptation for carriers to reject loads within long-term relationships. Shippers cannot observe if carriers truly do not have a truck available or if they are opportunistically declining to accept a higher priced offer on the spot market. Since shippers have imperfect monitoring of carriers' reasons for rejections, such a temptation constitutes a moral hazard

[17]This figure is taken from Harris and Nguyen (2021), which has a more detailed discussion on the comovements of spot, contract, and rejection rates.

problem. On the extensive margin, spot rates create an upward pressure on contract rates at the auction stage, affecting which relationships are formed and how relationship surplus is split between shippers and carriers.

To mitigate the moral hazard problem, shippers use their power to reorder the routing guide. That is, shippers can use the threat of demoting the current primary (top-ranked) carrier to a lower position on the routing guide to induce this carrier to accept more loads. In Harris and Nguyen (2021), we find that the shippers' incentive scheme takes a termination form: (i) higher rejection rates increase the likelihood of demotion, and (ii) once demoted, carriers hardly ever regain their primary status. Thus, we will model shippers' incentive scheme as a probabilistic termination strategy that conditions on carriers' past rejections. The estimated strength of this relational scheme will allow us to decompose the effects of relationships' intrinsic benefits from the effects of the dynamic incentives induced by such a scheme.[18]

4 Data

To capture the current market institution, we combine transaction-level data on long-term relationships and market-level data on spot arrangements. This section describes our data and provides empirical facts suggestive of the key tradeoff in our analysis: the match-specific gains from long-term relationships versus the efficiency of the spot market.

4.1 Transaction-level data on long-term relationships

We obtain detailed data on the interactions between shippers and carriers within long-term relationships from the TMS software provided by TMC, a division of C.H. Robinson.[19] For each shipper and each lane of that shipper, we observe the details of all loads, including the origin, destination, distance, and activity date. Furthermore, we observe some aspects of shippers' input into the TMS software, including carriers' ranks and volume constraints, as well as information on the offers made to the carriers through the waterfall process, including their order, timestamps, contract rates, and carriers' decisions. We use these data to identify the primary carriers, when these primary carriers are replaced, and when auctions are held.

[18]The shippers' incentive scheme is described in great details in Harris and Nguyen (2021, p. 30-35). Appendix D.1 presents our estimates of shippers' incentive scheme from a Probit specification, which will be used as an instrumental object in our empirical analysis. Consistent with Harris and Nguyen (2021), our estimates suggest that the shippers' incentive scheme is soft but generates dynamic incentives that are economically significant.

[19]C.H. Robinson is the largest third-party logistics firm in the United States.

We define a relationship as the interactions between a shipper and a primary carrier on a lane, until that carrier is demoted and before the next auction.[20]

Our data cover the period from September 2015 to August 2019. In total, we observe 1,186,413 loads and 2,367,704 offers between 54 shippers and 2020 carriers on 21,336 origin-destination pairs. These are all long hauls, with haul distance of at least 250 miles. We identify a total of 24,601 relationships, of which 13,171 are between a shipper and an asset-owner carrier. For our main analysis, we drop the relationships between shippers and brokers. The reason is that brokers act as intermediaries connecting shippers to carriers in the spot market; their cost structure and the nature of their relationships with shippers are thus very different from those of asset owners.[21]

Within the restricted data set, the average time between two auctions is 320 days, the average duration of a relationship is 33 offers, and the average number of offers is 7 loads per month. The average contract rate of primary carriers is $1.82/mile, with a standard deviation of $0.53/mile. In our sample, the premiums of spot rates over contract rates have mean $0.04/mile, with a standard deviation of $0.53/mile. On average, 70.3% of loads are accepted by primary carriers, 19.8% by backup carriers, and 8.9% of loads are fulfilled in the spot market. To simplify our analysis, we treat loads fulfilled by backup carriers as spot arrangements.

4.2 Market-level data on spot arrangements

We use spot rates data to capture the outside option of shippers and carriers in relationships and spot volume data to quantify the link between the thickness and efficiency of the spot market. These data come from DAT Freight and Analytics, the dominant freight marketplace platform in the US and the leading vendor of spot market data. DAT divides the US into 135 Key Market Areas (KMAs). We observe weekly summary statistics of spot rates and spot volume on each KMA-KMA lane. To merge these data with our transaction-level data on long-term relationships, we redefine origin-destination pairs in observed relationships at the KMA-KMA level. In total, there are 6,287 long-haul KMA-KMA lanes in our data on long-term relationships, out of 17,178 such lanes in the spot market data.

There are persistent differences in spot rates across lanes and large variation in spot rates over time. A regression of spot rates on lane fixed effects has an R^2 of 0.78; the average of these lane fixed-effects is $1.58/mile, with a standard deviation of $0.40/mile.

[20] See Appendix A for details on how we construct indicators of primary status, demotion and auction events, and a graph of how these events are distributed over time in our sample period.

[21] See Appendix A, where we show that compared to asset-owner carriers, brokers have higher tendency to accept loads but are also more responsive to changes in spot rates.

Our analysis will control for time-invariant heterogeneity across lanes and exploit the large temporal variation of spot rates relative to contract rates to trace the value of relationships. The idea is that if a relationship has high value, the carrier's tendency to accept offered loads should be less sensitive to spot rates.

To proxy for spot market thickness on each lane, we use the average weekly number of loads on that lane that shippers post on DAT's marketplace. This marketplace is a "load board" where shippers post their demand and carriers search for available loads. There is large variation in spot market thickness, with 50% of the lanes having less than 20 spot loads per week and the top 1% of lanes having more than 500 spot loads per week. We exploit this variation to pin down the link between spot market thickness and efficiency.

4.3 Descriptive results

The magnitude of the two-way crowding-out effects between long-term relationships and the spot market depends on (i) the match-specific gains within relationships and (ii) the link between thickness and efficiency in the spot market. In one direction, the larger the intrinsic benefits of relationships, the weaker the crowding-out effect of the spot market. In the other direction, the stronger the link between spot market thickness and efficiency, the stronger the crowding-out effect of long-term relationships. This subsection provides evidence that relationships generate match-specific gains to participating parties but that they could also exert substantial negative externalities on the spot market by reducing spot market thickness.

Match-specificity. Patterns in the data suggest that match-specific gains are an important concern for both shippers and carriers.

On the shippers' side, two patterns in shippers' requests for shipment within relationships suggest that their gains from relationships are match-specific and potentially large. First, when selecting a primary carrier (the first carrier to receive load offers) in an auction, the shipper does not necessarily select the carrier that proposes the lowest contract rate. On the contrary, we see in Figure 2, which plots the distribution of the difference between the rates of the primary carrier and the lowest-rate backup carrier, that the primary carrier has the lowest contract rate in only two-thirds of the auctions. Among the remaining auctions, the median primary-backup price gap is 17 cents/mile. Such non-monotonicity in contract rates of shippers' ranking over carriers suggests that shippers also care about factors other than prices.

Second, shippers continue offering loads to their primary carriers even when lower-rate alternatives are available in the spot market. This pattern suggests that shippers' gains from

Figure 2: Distribution of primary-backup price gaps

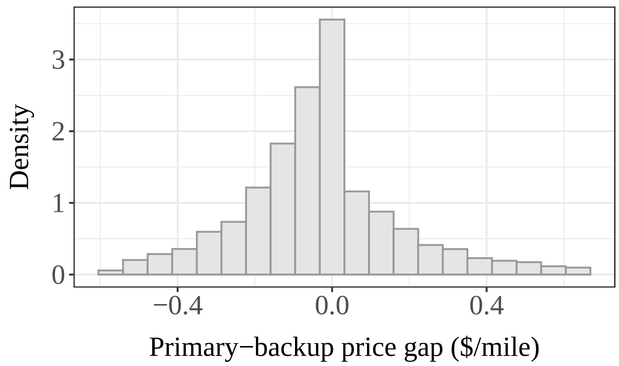


Table 2: Estimation results of Equation (1)

Sensitivity of shippers' offers to spot	
β_1^{shipper}	-0.055
	(0.011)
# Observations	197,749

transactions within relationships are potentially large. To show this, we run a regression of the variation in shippers' requests on variation in spot rates within the same contract period (auction) on the same lane,[22]

$$\ln\left(\frac{\text{Volume}^{\text{LT}}_{ia\ell,\text{month}}}{\text{Volume}^{\text{LT}}{}_{ia\ell}}\right) = \beta_0^{\text{shipper}} + \beta_1^{\text{shipper}} \ln\left(\frac{\tilde{p}_{ia\ell,\text{month}}}{\text{Rate}_{ia\ell}}\right) + \tilde{\epsilon}_{ia\ell,\text{month}}. \tag{1}$$

Table 2 shows that the estimate of β_1^{shipper} is negligible. This suggests that, unlike carriers, shippers do not defect from relationships when faced with spot temptation.

On the carriers' side, differences in the tendency of a carrier to accept offers in different relationships suggest that its gains from relationships are match-specific but potentially small. We consider the carrier with the largest number of relationships in our data set and approximate its acceptance tendency by a relationship-specific function of the normalized gap between spot and contract rates,[23]

$$\Pr(d_{ij\ell t} = \text{accept}) = \text{Logit}\left(\beta^{\text{carrier}}_{ij\ell,0} + \beta^{\text{carrier}}_{ij\ell,1}\left(\frac{\tilde{p}_{\ell t} - p_{ij\ell}}{\text{Std\_Rate}_\ell}\right)\right). \tag{2}$$

Table 3 presents the estimates of Equation (2) by mixed Logit. The likelihood ratio test of the mixed Logit specification against the pooled Logit specification has a Chi-square value of 1365.9, showing strong evidence of heterogeneity in the acceptance tendency of this single carrier across different relationships. Figure 3 demonstrates this heterogeneity in the

[22] $\text{Volume}^{\text{LT}}_{ia\ell,\text{month}}$ is the monthly number of requests that shipper i sends to her routing guide within the contract period of auction a on lane ℓ; $\tilde{p}_{ia\ell,\text{month}}$ is the median spot rate on lane ℓ in that month. We normalize these measures by their averages across all months of the contract period, $\text{Volume}^{\text{LT}}_{ia\ell}$ and $\text{Rate}_{ia\ell}$, to control for volume and rate differences across shippers and lanes.

[23] $d_{ij\ell t}$ is carrier j's decision in period t in its relationship with shipper i on lane ℓ; $\tilde{p}_{\ell t}$ is the spot rate on lane ℓ in period t; $p_{ij\ell}$ is the contract rate; Std_Rate_ℓ is the standard deviation of spot rates on lane ℓ.

Figure 3: Carriers' acceptance tendency

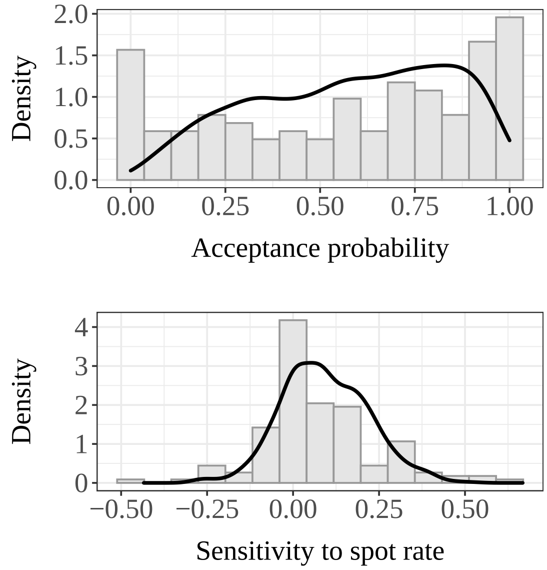


Table 3: Estimation results of Equation (2)

Sensitivity of a carrier's acceptances to spot	
$\begin{pmatrix}\beta^{\text{carrier}}_{ij\ell,0}\\ \beta^{\text{carrier}}_{ij\ell,1}\end{pmatrix}$	Normal $\left(\begin{pmatrix}0.46\\ -0.53\end{pmatrix}, \begin{pmatrix}1.84 & 0.28\\ 0.28 & 0.54\end{pmatrix}\right)$
Heterogeneity	$\chi^2 = 1365.9$
# Relationships	143
# Observations	7,306

Notes: **Figure 3**: The histograms are constructed by estimating Equation (2) separately for each relationship. The density curves are constructed by estimating Equation (2) with mixed Logit. **Table 3**: Mixed Logit results.

distribution of the carrier's acceptance probabilities and sensitivities to spot rate. Here, acceptance probabilities are predicted for spot rates equal to contract rates; sensitivities to spot rates are measured by how much predicted acceptance probabilities decrease when spot rates increase relative to contract rates by one standard deviation. In addition to showing large heterogeneity across relationships, the predicted sensitivities of this carrier's acceptance to spot rates suggest that its match-specific gains are potentially small. In two-thirds of its relationships, this carrier accepts less frequently as soon as spot rates exceed contract rates. In the median relationship, a one-standard-deviation increase in spot rates beyond this carrier's contract rate reduces its acceptance probability by 13 percentage points.

Spot market thickness and efficiency. At the market level, we find evidence suggesting a link between the thickness and the efficiency of the spot market. Intuitively, if the spot market becomes more efficient as it becomes thicker, then there is an equilibrium force that—all else equal—results in a higher share of transactions taking place in the spot markets on lanes with higher potential for total market volume. We test this hypothesis by running the following regression on the difference in the growth rates of spot volume and long-term relationship volume as the total market volume increases,

$$\ln(\text{Volume}^{\text{spot}}_{ss'}) - \ln(\text{Volume}^{\text{LT}}_{ss'}) = \beta_0 + \beta_1 \ln(\text{Volume}^{\text{total}}_{ss'}) + \text{controls} + \epsilon_{ss'}. \tag{3}$$

For this regression, lanes are defined at the state level. On each state-to-state lane, $\text{Volume}_{ss'}^{\text{spot}}$ is the average weekly load posts in the spot market, $\text{Volume}_{ss'}^{\text{LT}}$ is the average weekly loads accepted within long-term relationships, and $\text{Volume}_{ss'}^{\text{total}}$ is the total for-hire truckload volume in 2017 taken from the Commodity Flow Survey. We run a log-regression to avoid scaling issues between different data sets.[24] If $\beta_1 > 0$, the spot market takes a larger share of the total market volume as that total market grows, supporting our hypothesis.

An endogeneity concern in estimating this regression is that unobserved demand and supply factors could affect both total volume and the split between spot transactions and long-term relationships. First, demand patterns across industries may exhibit a correlation between total volume and preferences over forms of transactions. For example, shippers in some industries may have large total demand, but their lane-specific demand is infrequent and irregular; the latter prevents them from establishing long-term relationships. Second, unobserved cost factors may also make relationships easier or harder to establish. We mitigate the first concern by controlling for the frequency and consistency of load timing within observed relationships.[25] To address the second concern, we instrument for $\text{Volume}_{ss'}^{\text{total}}$ with a demand shifter, the predicted trade flows between different states of the US from Caliendo, Parro, Rossi-Hansberg, and Sarte (2018). To construct these predicted flows, the authors first build a state-of-the-art trade model of the US economy that captures input-output linkages between different sectors, labor mobility, and heterogeneous productivities, but not the split between spot transactions and long-term relationships. They calibrate this model using 2012 data. The resulting predicted trade flows are for all modes of transportation, not just trucking.

Table 4 presents our estimates of Equation (3), showing in all specifications that spot volume increases faster than long-term relationship volume when there is greater demand for transportation service.[26] With regard to demand factors, the coefficient estimates of (OLS2) confirm that shipper- and lane-specific frequency and consistency of load timing make long-term relationships more desirable. However, the inclusion of these variables gives an estimate of β_1 similar to that of (OLS1); this suggests that demand factors, while important, do not create serious endogeneity concern in Equation (3). With regard to cost factors, the (IV) specification estimates a stronger link between spot market share and total market volume. This suggests that unobserved costs, which tend to reduce total market volume, may favor

[24] Relative to market-level data on long-term relationships, our microdata overrepresent the American Midwest. We control for this overrepresentation by including indicators of a lane's origin or destination being in the Midwest.

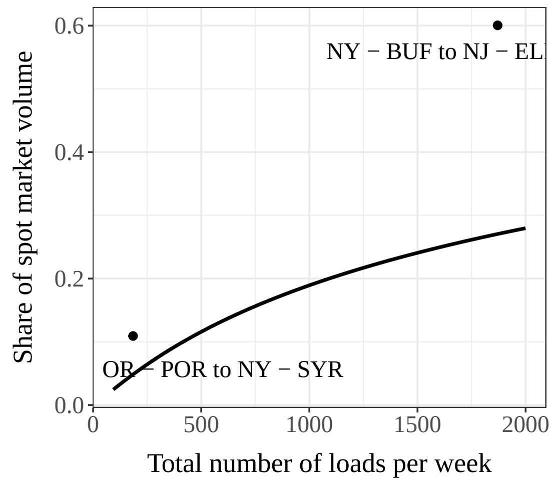
[25] Ideally, we would also control for the lane-specific frequency and consistency of load timing of shippers who use spot arrangements. However, we do not have access to such data.

[26] See Table 9 in Appendix D.2 for robustness checks.

Table 4: Estimation results of Equation 3

	$\ln(\text{Volume}^{\text{spot}}) - \ln(\text{Volume}^{\text{LT}})$		
	(OLS1)	(OLS2)	(IV)
$\ln(\text{Volume}^{\text{total}})$	0.285	0.269	0.345
	(0.040)	(0.037)	(0.045)
ln(distance)	0.117	−0.028	0.032
	(0.080)	(0.075)	(0.077)
Frequency		−0.156	−0.159
		(0.076)	(0.076)
Inconsistency		1.682	1.653
		(0.319)	(0.321)
# Observations	588	588	588

Figure 4: Shares of spot market volume



Notes: **Table 2:** Frequency is the median average monthly volume on a lane; Inconsistency is the median coefficient of variation of loads in a week over the four weeks of a month. This regression aggregates spot and long-term relationship volumes to the state-to-state level and restricts to lanes with at least 10 relationships. Standard errors are in parentheses. **Figure 3:** The fitted line is constructed from $\hat{\beta}_1 = 0.345$ from the IV specification. Two examples are included: Portland, Oregon to Syracuse, New York is a thin lane (200 loads/week); Buffalo, New York to Elizabeth, New Jersey is a thick lane (2000 loads/week). See Appendix C.3 for more details.

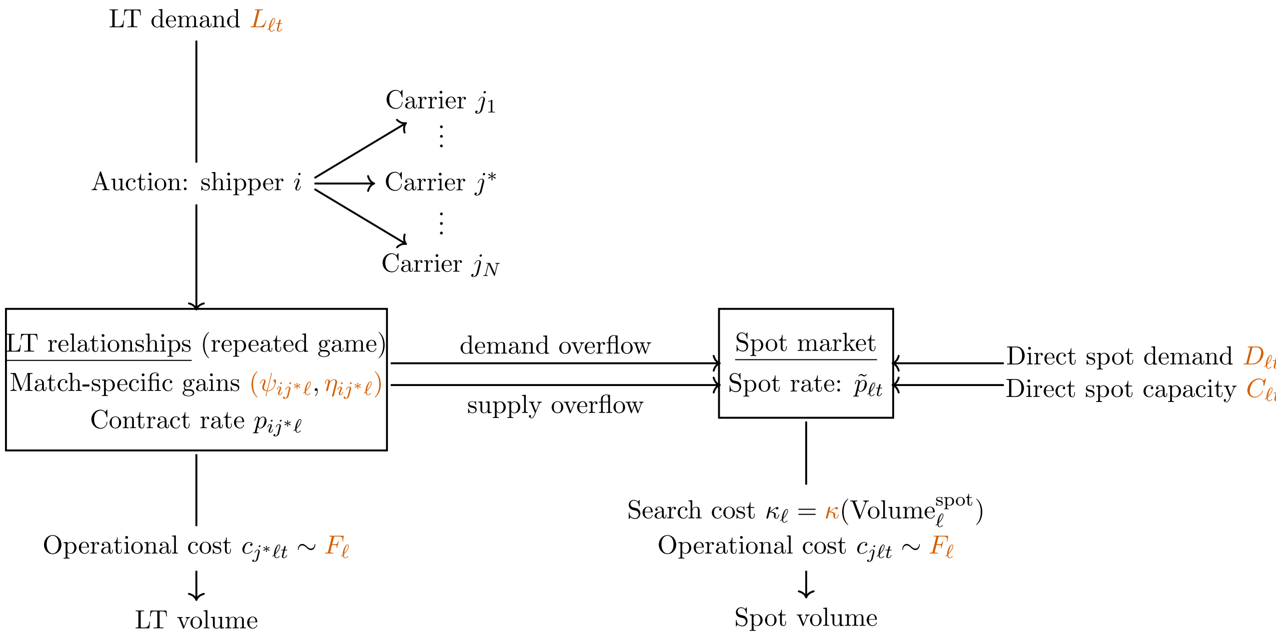
spot arrangements relative to long-term relationships.

To interpret the strength of the link between spot market share and total market thickness, we use the coefficient estimates in the (IV) specification to calibrate the shares of spot volume across all lanes. Figure 4 plots these shares against the total market volume. The fitted relationship shows that increasing the total market volume from 500 to 1000 loads per week increases the share of the spot market from 10% to 20%. This finding suggests a potentially strong link between spot market thickness and the desirability of the spot market. Additionally, it is consistent with the finding in Hubbard (2001), that the share of spot relative to contractual arrangements in freight trucking increases with market thickness.

5 Model

To quantify the benefits of long-term relationships and their aggregate effects on the spot market, we need a model that captures: (i) the levels and heterogeneity of gains from long-term relationships, (ii) how these relationships interact with the spot market, and (iii) how spot market thickness is linked to spot market efficiency. To capture the match-specificity of relationships, we model the potential gains from the relationship between a shipper i and a primary carrier j on lane ℓ as match-specific gains $(\psi_{ij\ell}, \eta_{ij\ell})$ to the shipper and

Figure 5: Model overview



carrier respectively over spot transactions. To allow for a potential link between spot market thickness and efficiency, we model per-load search cost for spot loads on lane ℓ as a function of spot volume on that lane, $\kappa_\ell = \kappa(\text{Volume}_\ell^{\text{spot}})$. Long-term relationships and the spot market interact in two ways. First, the spot market serves as a clearing mechanism, fulfilling loads rejected within relationships. Second, the equilibrium volume split between long-term relationships and spot arrangements endogenously determines search costs on the spot market.

Figure 5 provides an overview of our model. Each long-term relationship goes through two stages. In the first stage, the shipper holds an auction to select a primary carrier. In the second stage, the shipper and primary carrier interact repeatedly under the fixed contract rate that was established by the auction. For each offer in the relationship, the carrier may reject because either spot rate or operational cost that period is high. This creates an overflow in both demand and supply from long-term relationships to the spot market. In equilibrium, such overflow, direct spot demand, spot capacity, search and operational costs pin down spot rate. We will use the observed equilibrium behaviors of individual shippers and carriers to recover key model primitives: the distribution of match-specific gains $(\psi_{ij\ell}, \eta_{ij\ell})$, the distribution F_ℓ of operational costs, and function κ, which links search costs to spot market thickness. The market equilibrium condition will be used to recover the underlying supply and demand shocks, which are an input into our counterfactual analysis.

5.1 Timing and primitives of an individual relationship

We introduce the elements of an individual relationship in three layers: (i) shippers and carriers' per-period payoffs, (ii) the dynamics within a relationship, and (iii) the formation of that relationship.

Per-period payoffs. The relationship between a shipper i and a carrier j on lane ℓ is characterized by a tuple $(\psi_{ij\ell}, \eta_{ij\ell}, p_{ij\ell}, \delta_{i\ell})$ of relationship characteristics and a tuple $(\mathcal{P}_\ell, F_\ell, \kappa_\ell)$ of lane characteristics. Here, $\psi_{ij\ell}$ is the match-specific gain of the shipper from transacting with the carrier on lane ℓ; $\eta_{ij\ell}$ is the match-specific gain of the carrier from transacting with the shipper on lane ℓ; $p_{ij\ell}$ is the contract rate; $\delta_{i\ell}$ is the discount factor, reflecting the frequency of interactions; $\mathcal{P}_\ell$ is the spot process and F_ℓ is the distribution of the carrier's operational costs on lane ℓ; $\kappa_\ell = \kappa(\text{Volume}_\ell^{\text{spot}})$ is the cost to a carrier of searching for a spot load on lane ℓ, which depends on the average spot volume on that lane. Denote by $\tilde{p}_{\ell t}$ and $c_{j\ell t}$, respectively, the spot rate and the operational cost draw of the carrier on lane ℓ in period t. The shipper's period-t payoff is $u_{i\ell t} = \psi_{ij\ell} - p_{ij\ell}$ if she is served by the contracted carrier and $u_{i\ell t} = -\tilde{p}_{\ell t}$ if she is served by the spot market. The carrier receives the period-t payoff of $v_{j\ell t} = \eta_{ij\ell} + p_{ij\ell} - c_{j\ell t}$ when delivering a load for the contracted shipper and $v_{j\ell t} = \tilde{p}_{\ell t} - \kappa_\ell - c_{j\ell t}$ when serving the spot market. That is, a contracted load, if accepted, yields a premium (over a spot load) of $\psi_{ij\ell}$ to the shipper and a premium of $\eta_{ij\ell} + \kappa_\ell$ to the carrier, including the carrier's savings on search costs.

Repeated game (dynamics within the relationship). In each period ($t \geq 1$) of a relationship between shipper i and carrier j on lane ℓ, the shipper decides whether to terminate the relationship or offer a load to the carrier, and the carrier decides, if the shipper offers a load, whether to accept or reject it.

We allow the shipper to condition relationship termination on past decisions of the carrier. Let d_t denote the carrier's decision in period t: $d_t =$ accept if the carrier accepts the offered load; $d_t =$ spot if the carrier rejects the offered load to serve the spot market; $d_t =$ idle if the carrier rejects and remains idle. Denote by R_t an index summarizing carrier rejections in every period up to t. It is defined recursively by $I_R : (R_{t-1}, d_t) \mapsto R_t = \alpha R_{t-1} + (1-\alpha)\mathbf{1}\{d_t \neq \text{accept}\}$, where the weight α and the initial state R_0 are known. Let the spot rate follow an AR(1) process. The rejection index at the beginning of a period and the spot rate in the last period form a public (Markov) state $(R_{t-1}, \tilde{p}_{\ell t-1})$. If the relationship terminates in some period, both the shipper and the carrier resort to the spot market for all future transactions; otherwise, the relationship continues to the next period starting at a new public state.

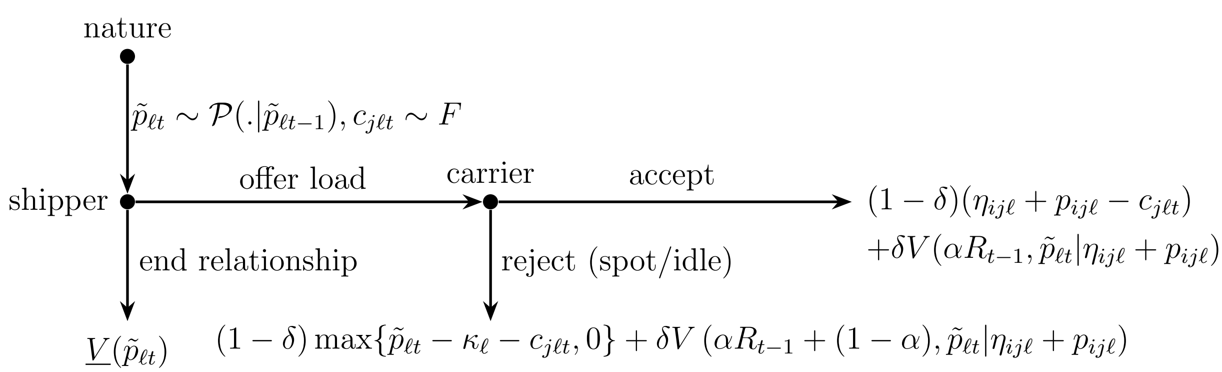
Assume that the shipper uses an incentive scheme $\sigma_s : (R_{t-1}, \tilde{p}_{\ell t}) \mapsto [0, 1]$, which specifies for each rejection index and current spot rate the probability that the shipper maintains the carrier's primary status and offers it a load. Assume that the carrier strategy, $\sigma_c : (R_{t-1}, \tilde{p}_{\ell t}) \mapsto \{\text{accept}, \text{spot}, \text{idle}\}$, specifies the carrier's optimal action for each rejection index and spot rate.

At the beginning of an auction, the seller announces an incentive scheme σ_s that will apply to whomever wins the auction. Although it would seem restrictive to assume that the shipper does not condition the incentive scheme on the auction outcome, in our framework, σ_s can be interpreted as the (average) incentive scheme perceived by all bidding carriers.

Assumption 1. *The shipper's incentive scheme $\sigma_s : (R_{t-1}, \tilde{p}_{\ell t}) \mapsto [0,1]$ does not depend on the outcome of the auction. That is, σ_s depends only on the characteristics of the lane and the discount factor δ.*

Under Assumption 1, the expected payoff of the winning carrier j depends on the auction outcome only through its per-transaction rent $\eta_{ij\ell} + p_{ij\ell}$. Write $V(R_{t-1}, \tilde{p}_{\ell t-1} | \eta_{ij\ell} + p_{ij\ell})$ for the expected payoff of the carrier in period t conditional on the rejection index R_{t-1} at the beginning of that period and the spot rate $\tilde{p}_{\ell t-1}$ last period, if this carrier's per-transaction rent is $\eta_{ij\ell} + p_{ij\ell}$. Write $\underline{V}(\tilde{p}_{\ell t})$ for the carrier's expected payoff from always going to the spot market, starting with $\tilde{p}_{\ell t}$ as the current spot rate. Figure 6 plots the timing of the stage game at state $(R_{t-1}, \tilde{p}_{\ell t-1})$ and the carrier's payoffs in different outcomes.

Figure 6: The stage game at $(R_{t-1}, \tilde{p}_{\ell t-1})$ and the carrier's discounted expected payoffs



In contrast, the expected payoff of shipper i depends on both her per-transaction rent $\psi_{ij\ell} - p_{ij\ell}$ and the carrier's per-transaction rent $\eta_{ij\ell} + p_{ij\ell}$, since the latter affects the carrier's tendency to accept loads within their relationship. This also means that the shipper might prefer a higher contract rate to a lower contract rate if the former induces significantly higher acceptance probability by the carrier, an idea similar to an efficiency wage. Thus, we write $U(R_{t-1}, \tilde{p}_{\ell t-1} | \psi_{ij\ell} - p_{ij\ell}, \eta_{ij\ell} + p_{ij\ell})$ for the shipper's expected payoff in state $(R_{t-1}, \tilde{p}_{\ell t-1})$ if the relationship induces per-transaction rents $(\psi_{ij\ell} - p_{ij\ell}, \eta_{ij\ell} + p_{ij\ell})$, and $\underline{U}(\tilde{p}_{\ell t})$ for the shipper's expected payoff from always going to the spot market given the current spot rate $\tilde{p}_{\ell t}$.

Auction (formation of the relationships). At the auction stage ($t = 0$), a set of carriers propose contract rates and the shipper chooses a carrier with whom to form a relationship. We use subscript a to denote auction-specific variables. The timing of an auction is as follows:

(i) Shipper i announces the expected frequency of interaction ($\delta_{i\ell}$), other characteristics of the lane, and incentive scheme σ_s.

(ii) A set J_a of N carriers arrive.

(iii) Pairs of shipper-carrier match-specific gains ($\psi_{ij\ell}, \eta_{ij\ell}$) are drawn i.i.d from the distribution $G_\ell^{\psi,\eta}$. The shipper's match-specific gain $\psi_{ij\ell}$ is observable to both her and carrier j, while the carrier's match-specific gain $\eta_{ij\ell}$ is privately known to carrier j alone.

(iv) Each carrier j proposes a contract rate $p_{ij\ell}$.

(v) Shipper i chooses the carrier that maximizes her expected payoff as long as it is not lower than her expected payoff from the outside option of always going to the spot market.

A key assumption is that all carriers on the same lane have the same cost distribution and search costs, but they differ in the match-specific gains ($\psi_{ij\ell}, \eta_{ij\ell}$) that they will generate for each interaction with the shipper. Moreover, carriers' match-specific gains are privately known to carriers, whereas the shipper's match-specific gain with each carrier is known between the pair. In addition to providing tractability, these informational assumptions match certain features of the communication process between shippers and carriers. In her requests for proposals, a shipper details her preferences for the service on a lane, and carriers respond with proposals explaining how they can meet such preferences. It is harder for the shipper to know how much carriers value their relationships, since this further depends on carriers' internal operations.

The informational assumptions above will allow us to transform carriers' bidding problem into the space of per-transaction rents. Match-specific gains affect shippers and carriers' expected payoffs only through these rents. First, since $\psi_{ij\ell}$ is known between shipper i and carrier j, by proposing a contract rate $p_{ij\ell}$, the carrier essentially proposes a per-transaction rent $\psi_{ij\ell} - p_{ij\ell}$ to the shipper. That is, the proposed shipper's rent is the carrier's effective bid. Second, the shipper forms her expected payoff in each relationship from each carrier's effective bid and the carrier's rent; the latter is inferred in equilibrium. Finally, the assumption that carriers do not know the match-specific gains potentially generated by other carriers allows us to use empirical tools from the literature on independent private value auctions.

5.2 Equilibrium behaviors of individual shippers and carriers

This section derives three equilibrium conditions of individual relationships that will be used for identification. First, the winning carriers play optimally within their relationships with shippers. Second, given their expected payoff from a relationship, carriers bid optimally at the auction stage. Third, shippers select the relationships that yield them the highest expected payoffs, unless these payoffs are lower than what they would get from the spot market. Restricting our analysis to the class of symmetric monotone equilibria, we will write the last two conditions in the space of shippers and carriers' rents.

We start with the optimal dynamic play of the winning carrier j in the repeated game. Define the "full compensation" for this carrier by

$$\begin{aligned}&\bar{p}(R_{t-1}, \tilde{p}_{\ell t}|\eta_{ij\ell} + p_{ij\ell})\\ =&\eta_{ij\ell} + p_{ij\ell} + \kappa_\ell + \frac{\delta}{1-\delta}(V(\alpha R_{t-1}, \tilde{p}_{\ell t}|\eta_{ij\ell} + p_{ij\ell}) - V(\alpha R_{t-1} + (1-\alpha), \tilde{p}_{\ell t}|\eta_{ij\ell} + p_{ij\ell}))\end{aligned} \quad (4)$$

and "transformed cost" by $\tilde{c}_{j\ell t} = c_{j\ell t} + \kappa_\ell$. The optimal strategy σ_c of this carrier at each $(R_{t-1}, \tilde{p}_{\ell t})$ depends on the relative ranking of $\bar{p}_{\ell t} = \bar{p}(R_{t-1}, \tilde{p}_{\ell t}|\eta_{ij\ell} + p_{ij\ell})$, $\tilde{p}_{\ell t}$ and $\tilde{c}_{j\ell t}$ as follows:

$$\sigma_c(R_{t-1}, \tilde{p}_{\ell t}|\eta_{ij\ell} + p_{ij\ell}) = \begin{cases} \text{accept} & \text{, if } \bar{p}_{\ell t} \geq \max\{\tilde{p}_{\ell t}, \tilde{c}_{j\ell t}\} \\ \text{spot} & \text{, if } \tilde{p}_{\ell t} > \max\{\bar{p}_{\ell t}, \tilde{c}_{j\ell t}\} \\ \text{idle} & \text{, if } \tilde{c}_{j\ell t} > \max\{\bar{p}_{\ell t}, \tilde{p}_{\ell t}\}. \end{cases} \quad (5)$$

When a carrier decides whether to accept a load, it takes into account the contract rate $p_{ij\ell}$, the match-specific gain $\eta_{ij\ell}$, savings on search cost, and the effect of an acceptance on its continuation value. These components constitute the carrier's benefits from accepting a load within the relationship at each Markov state of the relationship, captured by the full compensation $\bar{p}$. Intuitively, when the carrier's per-period rent is higher, it is compensated more for an acceptance today both directly through higher compensation today and indirectly through higher compensation in the future.

Next, we focus on the class of symmetric monotone equilibria, which transform the bidding problem into the space of (per-transaction) rents.

Definition 1. (Symmetric monotone equilibria) An equilibrium of the auction and repeated game is a symmetric monotone equilibrium if there exists a strictly increasing and differentiable function $\mathbf{b} : \mathbb{R} \mapsto \mathbb{R}$, referred to as the "effective bidding function", such that

(i) (Single indexing) The equilibrium (per-transaction) rents of shipper i and carrier j depend only on their total match quality $\theta_{ij\ell} \equiv \psi_{ij\ell} + \eta_{ij\ell}$,

$$\text{shipper } i\text{'s rent: } \psi_{ij\ell} - p_{ij\ell} = \mathbf{b}(\theta_{ij\ell}),$$
$$\text{carrier } j\text{'s rent: } \eta_{ij\ell} + p_{ij\ell} = \theta_{ij\ell} - \mathbf{b}(\theta_{ij\ell}).$$

(ii) (Monotone bidding) The shipper's rent $\mathbf{b}(\theta_{ij\ell})$ and the carrier's rent $\theta_{ij\ell} - \mathbf{b}(\theta_{ij\ell}) \equiv \mathbf{r}(\theta_{ij\ell})$ are both strictly increasing in $\theta_{ij\ell}$.

(iii) (Optimal selection) The shipper chooses the carrier j^* that maximizes her per-transaction rent subject to her expected payoff being no less than her outside option of always going to the spot market,

$$j^* \in \arg\max_{j \in J_a} \mathbf{b}(\theta_{ij\ell}) \quad \text{s.t. } U(R_0, \tilde{p}_{\ell 0} | \mathbf{b}(\theta_{ij\ell}), \theta_{ij\ell} - \mathbf{b}(\theta_{ij\ell})) \geq \mathbf{E}[\underline{U}(\tilde{p}_{\ell 1}) | \tilde{p}_{\ell 0}].$$

The first condition says that only the total match quality, rather than the relative magnitude of the match-specific gains $\psi_{ij\ell}$ and $\eta_{ij\ell}$, affects which relationship is formed and how the gain from each transaction is split between the shipper and the carrier. That is, the observed price $p_{ij\ell}$ will adjust to reflect the relative magnitude of the shipper's versus the carrier's match-specific gains. The second condition implies that there is a one-to-one mapping between carriers' rents and effective bids, or shippers' rents. This condition is crucial to our identification argument; it will allow us to recover the distribution of shipper's rents from the distribution of carriers' rents. Finally, the third condition reduces a shipper's selection rule to choosing the carrier with the highest effective bid subject to her individual rationality constraint. The intuition is that in a symmetric monotone equilibrium, carriers that have higher effective bids are also those with higher rents, and thus would accept more frequently in a relationship. By choosing the carrier with the highest effective bid, a shipper thus maximizes both her per-transaction rent and the likelihood that such rent realizes. Appendix B.3 provides sufficient conditions for the existence of a symmetric monotone equilibrium.

Recall that $G_\ell^{\psi,\eta}$ denotes the distribution of match-specific gains in random matches of shippers and carriers. Let $G_\ell^\theta = G_\ell^{\psi+\eta}$ be the distribution of total match quality induced by $G_\ell^{\psi,\eta}$. Similarly, write $G_\ell^{\eta+p}$ for the distribution of carriers' rents and $G_\ell^{\psi-p}$ for the distribution of shippers' rents induced by $G_\ell^{\psi,\eta}$ and bidding function $\mathbf{b}$. Key to our welfare analysis is the distributions of carriers and shippers' rents in relationships selected by the auction process.

Denote these distributions by $[G_\ell^{\eta+p}]^{1:N}$ and $[G_\ell^{\psi-p}]^{1:N}$ respectively. The following proposition summarizes all equilibrium conditions on shippers and carriers' behaviors that will be used for identification.

Lemma 1. *(Equilibrium behaviors) Consider a symmetric monotone equilibrium with initial spot rate $\tilde{p}_{\ell 0}$ and effective bidding function $\boldsymbol{b}$. The following hold:*

(i) (Optimal dynamic play) The winning carrier with per-transaction rent $\eta_{ij\ell} + p_{ij\ell}$ uses the optimal accept/reject strategy $\sigma_c(\cdot, \cdot|\eta_{ij\ell}+p_{ij\ell})$ defined in Equation (5) when offered loads.

(ii) (Optimal symmetric monotone bidding) For all type $\theta_{ij\ell}$,

$$\boldsymbol{b}(\theta_{ij\ell}) = \arg\max_b G_\ell^\theta(\boldsymbol{b}^{-1}(b))^{N-1}(V(R_0, \tilde{p}_{\ell 0}|\theta_{ij\ell} - b) - \boldsymbol{E}[\underline{V}(\tilde{p}_{\ell 1})|\tilde{p}_{\ell 0}]), \tag{6}$$

with $\boldsymbol{b}(\theta_{ij\ell})$ and $\theta_{ij\ell} - \boldsymbol{b}(\theta_{ij\ell})$ both strictly increasing in $\theta_{ij\ell}$.

(iii) (Binding shipper's IR constraint) Let $\underline{\theta}_\ell$ be the lowest match quality in a relationship. Then,

$$U(R_0, \tilde{p}_{\ell 0}|\boldsymbol{b}(\underline{\theta}_\ell), \underline{\theta}_\ell - \boldsymbol{b}(\underline{\theta}_\ell)) = \boldsymbol{E}[\underline{U}(\tilde{p}_{\ell 1})|\tilde{p}_{\ell 0}]. \tag{7}$$

At every state and period of the repeated game, the carrier accepts only if its full compensation is higher than the compensation from the spot market and also higher than the sum of its operational and search costs. At the bidding stage, the bidding problem is as if carriers bid on the shipper's rent in a first-price auction with a reserved price determined by the shipper's individual rationality constraint. Notice that we do not impose optimality on the shipper's incentive scheme. One could in principle assume also that the shipper commits to an ex ante optimal incentive scheme or chooses a self-enforcing strategy, which is optimal period by period. Given that shipping is only a small component of shippers' business, we are less confident that such conditions would hold in reality, and did not want to impose it.

5.3 Market equilibrium condition

Next, we derive the market equilibrium condition that pins down the allocation of loads between long-term relationships and the spot market. Denote by $L_{\ell t}$ the measure of shippers who want to establish long-term relationships on lane ℓ in period t, by $D_{\ell t}$ the measure of direct spot demand by shippers who do not want to establish relationships and by $C_{\ell t}$ the spot capacity of carriers. Let $\mu_{\ell t}(.|\tilde{p}_{\ell t})$ denote the distribution over full compensations $\bar{p}$ for potential relationships, conditional on the current spot rate. That is, $\mu_{\ell t}(.|\tilde{p}_{\ell t})$ captures both the extensive and intensive margins of volume in long-term relationships. Specifically on the extensive margin, $\mu_{\ell t}(0|\tilde{p}_{\ell t})$ measures potential relationships that are not formed and relationships that have ended. On the intensive margin, higher $\mu_{\ell t}(.|\tilde{p}_{\ell t})$ in the FOSD-sense means higher aggregate acceptance probability. Moreover, the status of each relationship as captured by the rejection index, is embedded in the measure $\mu_{\ell t}(.|\tilde{p}_{\ell t})$. The market equilibrium condition is

$$L_{\ell t} + D_{\ell t} = \underbrace{L_{\ell t} \int_{\tilde{p}_{\ell t}}^{\infty} F(\bar{p} - \kappa_\ell) d\mu_{\ell t}(\bar{p}|\tilde{p}_{\ell t})}_{\text{LT-relationship volume}} + \underbrace{[L_{\ell t}\mu_{\ell t}(\tilde{p}_{\ell t}|\tilde{p}_{\ell t}) + C_{\ell t}]F(\tilde{p}_{\ell t} - \kappa_\ell)}_{\text{spot volume}}. \tag{8}$$

Within long-term relationships, loads are accepted if the full compensations $\bar{p}$ of carriers are higher than the current spot rate and higher than the sums of carriers' search and operational costs. Loads offered but rejected in long-term relationships will be fulfilled in the spot market, either by carriers in relationships with low full compensations or by those not in relationships.[27] A positive aggregate demand shock, either from an increase in demand for long-term relationships or from an increase in direct spot demand, increases the equilibrium spot rate.

For the welfare analysis of different market institutions, we keep fixed the long-term demand $L_{\ell t}$, the short-term demand $D_{\ell t}$, and the total capacity of carriers, including the spot capacity $C_{\ell t}$ and those that form relationships $L_{\ell t}$. Additionally, we normalize the value of spot interactions for shippers who want to establish long-term relationships to zero. Thus, the market-level welfare of the current institution on lane ℓ in period t is

$$\begin{aligned} W^0_{\ell t} =& L_{\ell t} \int_{\tilde{p}_{\ell t}}^{\infty} (\mathbf{E}[\theta|\bar{p}, \tilde{p}_{\ell t}] - \mathbf{E}[c_{\ell t}|c_{\ell t} \leq \bar{p} - \kappa_\ell]) F(\bar{p} - \kappa_\ell) d\mu_{\ell t}(\bar{p}|\tilde{p}_{\ell t}) \\ &+ [C_{\ell t} + L_{\ell t}\mu_{\ell t}(\tilde{p}_{\ell t}|\tilde{p}_{\ell t})](-\kappa_\ell - \mathbf{E}[c_{\ell t}|c_{\ell t} \leq \tilde{p}_{\ell t} - \kappa_\ell]) F(\tilde{p}_{\ell t} - \kappa_\ell). \end{aligned}$$

Alternative institutions that change the dynamics of long-term relationships modify the measure $\mu_{\ell t}(.|\tilde{p}_{\ell t})$ and the conditional expected match quality $\mathbf{E}[\theta|\bar{p}, \tilde{p}_{\ell t}]$. For the extreme case with no long-term relationships, we set $\mu_{\ell t}(0|\tilde{p}_{\ell t}) = 1$.

6 Identification and estimation

In this section, we discuss the identification of our model. We then specify parametric assumptions in an empirical model and explain our estimation procedure, which follows the

[27] For simplicity, we count backup carriers as spot carriers.

steps in the identification argument closely.

6.1 Identification

Suppose that in each relationship we observe the contract rate $p_{ij\ell}$, the duration of the relationship $T_{ij\ell}$, and in each period $t \leq T_{ij\ell}$, the spot rate $\tilde{p}_{\ell t}$ and whether the carrier accepts or rejects, that is, whether $d_{j\ell t} = \text{accept}$ or $d_{j\ell t} \in \{\text{spot}, \text{idle}\}$. The observed relationships have unobserved match-specific gains $(\psi_{ij\ell}, \eta_{ij\ell})$, lane-specific distribution F_ℓ of operational costs, search cost κ_ℓ, and incentive scheme σ_s. Furthermore, suppose that we observe the number n_a of bidders in each auction who pass the shipper's individual rationality constraint and become either primary or backup carriers. For simplicity, assume that the discount factor is $\delta_{i\ell} = \delta$. Our identification argument relies on the following assumptions, which will be maintained throughout our analysis.

Assumption 2. *(Regularity) Assume the following regularity conditions:*

(i) The spot process $\mathcal{P}_\ell$ is AR(1) and has $supp(\tilde{p}_{\ell t}|\tilde{p}_{\ell t-1}) = \mathbb{R}^+$ for every $\tilde{p}_{\ell t-1}$.

(ii) The shipper's strategy satisfies that $\sigma_s(R_{t-1}, \tilde{p}_{\ell t}) > 0$ for all $(R_{t-1}, \tilde{p}_{\ell t})$.

(iii) The underlying distribution of match-specific gains $G_\ell^{\psi,\eta}$ has full support in $\mathbb{R}^2$. Moreover, it induces an underlying distribution of match quality $G_\ell^\theta = G_\ell^{\psi+\eta}$ that has a strictly decreasing hazard rate, $g_\ell^\theta / G_\ell^\theta$.[28]

(iv) The cost distribution F_ℓ is Normal(μ_ℓ^c, σ^c).

Assumption 3. *(Properties of full compensation schedules) Under the spot process $\mathcal{P}_\ell$ and the shipper incentive scheme σ_s, the full compensation schedule $\bar{p}(R_{t-1}, \tilde{p}_{\ell t}|\eta + p)$ is:*

(i) strictly increasing in the carrier's rent $\eta + p$ for all $(R_{t-1}, \tilde{p}_{\ell t})$,

(ii) continuous in $\tilde{p}_{\ell t}$ for all R_{t-1} and $\eta + p$,

(iii) bounded below by $\eta + p + \kappa_\ell$ for all $(R_{t-1}, \tilde{p}_{\ell t})$.

Note that Assumption 3 is essentially an assumption on the underlying spot process and the shipper's incentive scheme. The most substantive assumption is (i).[29] Assumption (ii) holds under mild regularity conditions, and Assumption (iii) is satisfied if the incentive scheme σ_s is decreasing in R_{t-1}, that is, if the shipper punishes the carrier's rejections with a higher probability of demotion.

[28] For example, Normal distributions have strictly decreasing hazard rates.

[29] The left panel of Figure 22 in Appendix E shows that this assumption is numerically verified under our estimates of the spot process and the shipper's incentive scheme.

Proposition 1. *(Full identification) Under Assumption 2 and Assumption 3, the model is fully identified within the class of symmetric monotone equilibria.*

The shipper's incentive scheme σ_s is identified, since under Assumption 2, every Markov state $(R_{t-1}, \tilde{p}_{\ell t})$ is observed.[30] The spot process $\mathcal{P}_\ell$ is identified from the realized path of spot rates. This section provides an identification argument for the following key primitives: the distribution F_ℓ of operational costs, the search cost κ_ℓ, and the joint distribution $G_\ell^{\psi,\eta}$ of the match-specific gains of shippers and carriers on each lane ℓ.

These primitives are identified sequentially in four steps. First, we identify the distribution of carriers' operational and search costs. In this step, we exploit how carriers' tendency to accept offered loads varies across relationships and across lanes. The variation of such a tendency across different relationships on the same lane gives us different draws of the common cost distribution, and its variation across lanes with different spot market thickness pins down search costs. Given operational and search costs, the second step identifies the distribution of carriers' rents from the distribution of observed acceptances. The idea is that, all else equal, carriers with higher rents accept more. Third, we recover the distribution of shippers' rents by exploiting the optimality of carriers' bidding and shippers' selection in a symmetric monotone equilibrium. This step uses objects recovered from the previous steps to construct the carriers' probability of and expected payoff from winning an auction, which are key inputs to the optimal bidding condition. Finally, we exploit observed prices to map the distribution of shippers' and carriers' rents to the joint distribution of match-specific gains.

6.1.1 Identification of the distribution of operational costs and search costs

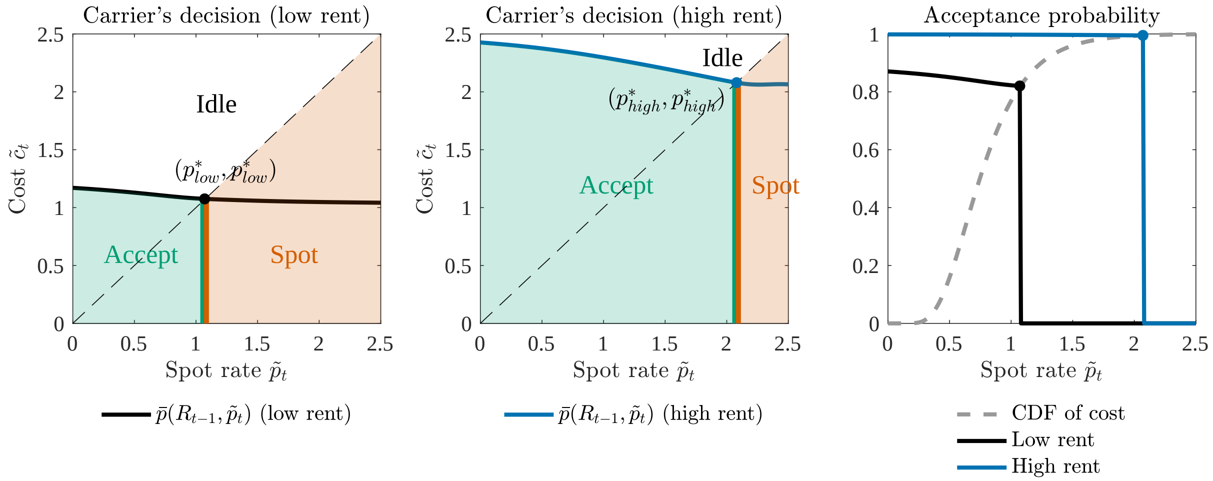
The key observables in our identification of carriers' primitives are carriers' "acceptance schedules", or their tendency to accept a load at each Markov state of their relationship. Following from carriers' optimal dynamic play in Equation (5), a carrier's acceptance schedule depends only on its transformed rent $\eta_{ij\ell} + p_{ij\ell} + \kappa_\ell$ and the distribution $\tilde{F}_\ell$ of transformed costs, $\tilde{c}_{\ell t} = c_{\ell t} + \kappa_\ell \sim \text{Normal}(\mu_\ell^c + \kappa_\ell, \sigma^c)$.

Definition 2. (Acceptance schedules) Fix the shipper's strategy σ_s, carrier's rent $\eta_{ij\ell} + p_{ij\ell}$ and lane characteristics $(\mathcal{P}_\ell, F_\ell, \kappa_\ell)$. Carrier j's acceptance schedule is that carrier's tendency to accept a load at each Markov state $(R_{t-1}, \tilde{p}_{\ell t})$ of its relationship with the shipper,

$$\Pr(d_t = \text{accept} | R_{t-1}, \tilde{p}_{\ell t}) = \mathbf{1}\{\bar{p}(R_{t-1}, \tilde{p}_{\ell t}) \geq \tilde{p}_{\ell t}\} \tilde{F}_\ell(\bar{p}(R_t, \tilde{p}_{\ell t})).$$

[30] Specifically, Assumption 2(i) ensures that every level of spot rate is observed, regardless of the current rejection state; (ii) ensures that every Markov state is non-absorbing; (iv) ensures a strictly positive probability of both acceptance and rejection in any Markov state.

Figure 7: Acceptance schedules and acceptance thresholds for a fixed rejection index



Notice that our assumption on the form of shippers' incentive schemes has two implications: (i) relationships evolve with a Markov state, and (ii) relationship terminations can occur on-path. The first implication means that what we observe about a carrier's dynamic play is precisely its tendency to accept loads at each Markov state of the relationship. Moreover, this tendency is fully observed in long-lasting relationships ($T \to \infty$). However, the second implication means that a significant proportion of relationships are relatively short-lived. Thus, we develop a two-step procedure. The first step takes advantage of long-lasting relationships to identify cost parameters. The second step identifies the distribution of carriers' rents, pooling relationships of all length. This section focuses on the first step.[31]

In this step, we separately identify cost parameters from carriers' match-specific gains by building on a thought experiment. Consider a hypothetical scenario when carriers only decide between "spot" and "idle". In this scenario, the probability that a carrier chose "spot" over "idle" at some spot rate would equal the CDF of the sum of this carrier's search and operational costs evaluated at this spot rate. Thus, the variation in the probability that a carrier chose "spot" over "idle" as spot rates vary would trace out the distribution of the sum of its search and operational costs. We build on this hypothetical scenario to pin down lane-specific cost parameters. In each relationship, we recreate this scenario locally by exploiting the observed level and sensitivity-to-spot-rate of the carrier's acceptance. Then, variation in carriers' acceptance across different relationships, with different contract rates or match-specific gains, on the same lane recreates the hypothetical scenario globally, pinning down cost parameters on this lane.

[31]Note that long-lasting relationships are selected in match-specific gains or in cost draws. Our two-step procedure handles selection in match-specific gains. Appendix C.1.2 discusses how to correct for selection in cost draws in the first step.

Figure 7 illustrates how the variation in the acceptance schedule induced by the variation in the level of carrier rent across relationships traces the distribution of transformed costs. The first two panels of Figure 7 plot the full compensation (black and blue lines) and optimal decisions of two carriers with different levels of carrier rent, fixing a rejection index and in the space of spot rate (x axis) and transformed cost (y axis). For each rent level, as spot rate increases, the full compensation starts from being higher than spot rate to eventually being lower than spot rate, crossing the 45-degree line at a critical point p^*. The carrier decides between "accept" and "idle" when spot rate is lower than p^*, between "spot" and "idle" when spot rate is higher than p^*, and is different between "accept" and "idle" exactly at p^*. This means that the acceptance probability at p^*, which is the observed probability mass on the green vertical line, is the probability that the carrier chooses "spot" over "idle" at this level of spot rate were "accept" to not be an option, which is the unobserved probability mass on the red vertical line. That is, $(p^*, \tilde{F}_\ell(p^*))$ gives us one point on the distribution of transformed costs. Moreover, increasing the carrier rent shifts the acceptance schedule outwards. Such a shift results in a higher critical point, giving us another point on the distribution of transformed costs.

The last panel of Figure 7 illustrates how these critical points manifest as "jump" points in the observed acceptance schedules and are thus identified. At spot rates lower than the critical points, the carriers accept if their costs are below the full compensation, which happens with a positive probability. Acceptance probabilities jump to zero for any higher level of spot rate, because carriers would prefer spot to contracted loads even when cost draws are low. A caveat of this identification strategy is that it is empirically harder to pin down the left tail of the cost distribution, since "jump points" tend to be quite high, and more likely so in long-lasting relationships. However, acceptance schedules provide information about the cost distribution not only through their jump points, but also through the portion of these schedules to the left of the jump points. This argument provides intuition for Lemma 2 below. We delegate the formal definition of the "jump points" to Appendix B.1 and the identification proof to Appendix B.2.

Lemma 2. *(Identification of the distribution of transformed costs) Fix lane characteristics. A carrier's acceptance schedule on lane ℓ identifies at least one point on the distribution $\tilde{F}_\ell$ of transformed costs, and the variation in carriers' rents identifies more points on $\tilde{F}_\ell$.*

Lemma 3. *(Identification of search costs and distribution of operational costs) If there is a demand shifter z_ℓ independent of operational costs, then the variation in z_ℓ and in the average spot volume, $Volume_\ell^{spot}$, identifies search costs $\kappa_\ell = \kappa(Volume_\ell^{spot})$ and the cost distribution*

F_ℓ, up to a constant.[32]

Proof. By Lemma 2, the distribution $\tilde{F}_\ell$ of transformed costs $\tilde{c}_{j\ell t} = c_{j\ell t} + \kappa(\text{Volume}_\ell^{\text{spot}})$ is identified. Since z_ℓ is independent of $c_{j\ell t}$ and correlates with $\text{Volume}_\ell^{\text{spot}}$, it can be used as an instrument to non-parametrically identify κ in the range of spot volume. □

6.1.2 Identification of the distribution of carriers' rents

Given the common cost distribution, we now identify the distribution of carrier rent. It is important to include all relationships in this step since excluding short-lived relationships would result in an upward bias of the recovered distribution of carriers' rents. To build intuition, consider the case where there are finitely many levels of carrier rent. The key idea is that the acceptance schedules associated with different levels of carrier rent are linearly independent, a property that ensures identification of finite mixtures of carrier rent.[33] Suppose that linear independence fails, that is, some acceptance schedule can be written as a linear combination of other acceptance schedules. Then none of these schedules can involve the highest acceptance schedule, since it is the only one in which a strictly positive probability of acceptance is observed near the highest "jump point". Applying this argument iteratively from the highest to the lowest acceptance schedules yields a contradiction. Thus, the mixture of carriers' rents, if finite, is identified. Lemma 4 generalizes this intuition to a continuum of rent level. We present a direct proof of this lemma in Appendix B.2.

Lemma 4. *(Identification of the distribution of carrier rent) Suppose that the shipper's incentive scheme σ_s, search cost κ_ℓ, and the distribution F_ℓ of operational costs are identified. Then the distribution $[G_\ell^{\eta+p}]^{1:N}$ of winning carriers' rents is identified.*

6.1.3 Identification of the distribution of shippers' rents

To identify the distribution $[G_\ell^{\psi-p}]^{1:N}$ of shipper rent, we take advantage of the fact that in a symmetric monotone equilibrium, there is a one-to-one mapping between carrier rent $(\eta + p)$ and shipper rent $(\psi - p)$. To identify this mapping, we adapt the identification strategy in Guerre, Perrigne, and Vuong (2000). First, we transform the bidding problem into the space of carrier rent, the distribution of which is identified in the previous step. Second, we identify the monotone mapping between carriers' rents and their effective bids,

[32]This constant has no bearing on our comparison of aggregate welfare between the current and alternative institutions. In the estimation, we pin down this constant by normalizing the median operational cost across lanes to industry estimates.

[33]Kasahara and Shimotsu (2009) show, in general dynamic discrete choice models with Markov states, that linear independence of response functions is sufficient for identification of finite mixtures.

or shippers' rents, from the first-order condition of optimal bidding in the space of carrier rent.

Lemma 5. *(Identification of the distribution of shippers' rents) Suppose that the incentive scheme σ_s, the distribution F_ℓ of operational costs, search cost κ_ℓ, the distribution $[G_\ell^{\eta+p}]^{1:N}$ of winning carriers' rents and the distribution of the number of bidders that pass the shipper's individual rationality constraint are identified. Then the distribution $[G_\ell^{\psi-p}]^{1:N}$ of shipper rent is nonparametrically identified.*

Proof. Consider an auction of shipper i with an initial spot rate $\tilde{p}_{\ell 0}$. In a symmetric monotone equilibrium with an effective bidding function $\mathbf{b}$ and a rent function $\mathbf{r} : \theta \mapsto r = \theta - \mathbf{b}(\theta)$, there exists a unique monotone mapping $\mathbf{b}_r : r \mapsto b$ defined by $\mathbf{b}_r(r) = \mathbf{b}(\mathbf{r}^{-1}(r))$. Thus, for carrier j with $\theta_{ij\ell} = \theta \geq \underline{\theta}$, the optimal bidding condition reduces to choosing r that solves

$$\max_r [G_\ell^{\eta+p}(r)]^{N-1}(V(R_0, \tilde{p}_{\ell 0}|\theta - \mathbf{b}_r(r)) - \mathbf{E}[\underline{V}(\tilde{p}_{\ell 1})|\tilde{p}_{\ell 0}]).$$

The first-order condition gives

$$(N-1)\frac{g_\ell^{\eta+p}(r)}{G_\ell^{\eta+p}(r)} = \frac{\frac{\partial}{\partial r}V(R_0, \tilde{p}_{\ell 0}|r)}{V(R_0, \tilde{p}_{\ell 0}|r) - \mathbf{E}[\underline{V}(\tilde{p}_{\ell 1})|\tilde{p}_{\ell 0}]}\mathbf{b}_r'(r). \tag{9}$$

Let $\underline{r}_\ell$ denote the lowest level of carrier's rent in the support of $[G_\ell^{\eta+p}]^{1:N}$, the shipper's IR constraint can be rewritten as

$$U(R_0, \tilde{p}_{\ell 0}|\mathbf{b}(\underline{r}_\ell), \underline{r}_\ell) = \mathbf{E}[\underline{U}(\tilde{p}_{\ell 1})|\tilde{p}_{\ell 0}]. \tag{10}$$

Notice that the carrier's expected payoff as a function of carrier's rent r is identified from the incentive scheme σ_s, cost distribution F_ℓ and search cost κ_ℓ. To pin down $\mathbf{b}_r'$ from Equation (9), it remains to show that $G_\ell^{\eta+p}$ is identified on $[\underline{r}_\ell, \infty)$. We have for any rent level $r > \underline{r}_\ell$, the distribution of winning carriers' rents satisfies

$$[G_\ell^{\eta+p}]^{1:N}(r) = \frac{[G_\ell^{\eta+p}(r)]^N - [G_\ell^{\eta+p}(\underline{r}_\ell)]^N}{1 - [G_\ell^{\eta+p}(\underline{r}_\ell)]^N}.$$

Moreover, the distribution of the number of effective bidders, who pass the shipper's individual rationality constraint, is Binomial$(N, 1 - G_\ell^{\eta+p}(\underline{r}_\ell))$. This means that $G_\ell^{\eta+p}(\underline{r}_\ell)$ is identified, which in turn identifies $G_\ell^{\eta+p}(r)$ on $[\underline{r}_\ell, \infty)$ from the distribution of winning carriers' rents.

Finally, since the left hand side of Equation (10) is strictly increasing in its first argument, this equation pins down $\mathbf{b}(\underline{r}_\ell)$. Thus, $\mathbf{b}_r$ is identified, which in turn identifies the distribution of shippers' rents from the distribution of carriers' rents. □

6.1.4 Identification of the distribution of match-specific gains

Finally, exploiting the fact that contract rates are observed, we identify the distribution of carriers' rents conditional on contract rates and thus, the joint distribution of shippers and carriers' match-specific gains.

Lemma 6. *(Identification of the distribution of match-specific gains) Given the shipper's punishment scheme σ_s, the common cost distribution F_ℓ, search cost κ_ℓ, the number of effective bidders and the contract rates in all relationships, the distribution $[G_\ell^{\eta+p|p}]^{1:N}$ of winning carriers' rents conditional on contract rates is identified. It follows that the joint distribution $[G_\ell^{\psi,\eta}]^{1:N}$ of shippers and carriers' match-specific gains is identified.*

Proof. That $[G_\ell^{\eta+p|p}]^{1:N}$ is identified is an extension of Lemma 4 by conditioning on observed contract rates. Since the distribution $[G_p]^{1:N}$ of contract rates is observed, it follows that the joint distribution $[G_\ell^{\eta,p}]^{1:N}$ of carriers' match-specific gains and contract rates is identified. Furthermore, notice that Lemma 5 identifies the monotone equilibrium mapping $\mathbf{b}_r$ from carrier rent to shipper rent, that is, $\mathbf{b}_r(\eta+p)=\psi-p$. This establishes a one-to-one mapping from (η,p) to (η,ψ), completing the proof that $[G_\ell^{\psi,\eta}]^{1:N}$ is identified. This means that the fundamental distribution of match-specific gains $G_\ell^{\psi,\eta}$ is identified but only for $\psi+\eta\geq\underline{\theta}_\ell$, since we do not observe potential relationships that fail shippers' IR constraints. □

6.2 Empirical model

Three features of the data require adaptations of our model. First, different shippers and carriers, through negotiations in the spot market, settle on different spot rates, but we only observe summary statistics of spot rates. To address this, our empirical model allows for idiosyncratic noise in the spot rate observed by a carrier at the time it decides whether to accept a load. Second, there is large persistent heterogeneity across lanes, which need to be controlled for in a pooled estimation. We use three variables to control for such heterogeneity: $\text{Rate}_\ell^{\text{spot}}$, $\text{Volume}_\ell^{\text{spot}}$, and Distance_ℓ, which are respectively the average spot rate across time, the average spot volume across time and the average distance of a trip on a KMA-KMA lane.[34] For ease of interpretation, our estimation treats match-specific gains, rates and costs on a per-mile basis. Third, the truckload freight market goes through phases, as macroeconomic conditions change. A "soft" ("tight") market is one in which demand for

[34]We take Distance_ℓ to be the practical mileage on a KMA-KMA lane, which is the industry's estimate of the most likely distance of a trip from a KMA-origin to a KMA-destination.

truckload service is lower (higher) than truckload capacity. Our empirical model takes the unit of a relationship to be between a shipper and a carrier on a lane in an auction period. We assume that the market phase can differ across auctions, but does not change within auctions.[35]

Spot variance and spot process. Let $\tilde{p}_{\ell t}$ denote the mean spot rate on lane ℓ in period t. Assume that at the time of decision making, carrier j faces spot rate $\tilde{p}_{\ell t} + \zeta_{j\ell t}$, where $\zeta_{j\ell t} \sim \text{Normal}(0, \sigma_\ell^\zeta)$. Assume further that future mean spot rates depend on the current mean spot rate through the following AR(1) process,

$$\frac{\tilde{p}_{\ell\tau}}{\text{Rate}_\ell} = \rho_0 + \rho_1 \frac{\tilde{p}_{\ell\tau-1}}{\text{Rate}_\ell} + \epsilon_{\ell\tau},$$

where τ denotes calendar day. The spot process as perceived in a relationship adjusts this calendar-based spot process to the frequency of the shipper and carrier's interactions. Under these two assumptions, the Markov state of a relationship includes the rejection index and mean spot rate, $(R_{t-1}, \tilde{p}_{\ell t})$, and the observed acceptance schedule is "smoothed" out,

$$\Pr(d_t = A | R_{t-1}, \tilde{p}_{\ell t}) = \Phi\left(\frac{\bar{p}(R_{t-1}, \tilde{p}_{\ell t}) - \tilde{p}_{\ell t}}{\sigma_\ell^\zeta}\right) \tilde{F}_{ia\ell}(\bar{p}(R_{t-1}, \tilde{p}_{\ell t})),$$

where $\tilde{F}_{ia\ell}$ is the auction- and lane-specific distribution of transformed costs.

Shippers' strategies. We specify a Probit model for demotion probability,

$$1 - \sigma_s(R_{t-1}, \tilde{p}_{\ell t}) \sim \Phi\left(\alpha_0 + \alpha_1 R_{t-1} + \alpha_2 \mathbf{X}_{i\ell t} + \alpha_3 R_{t-1} \mathbf{X}_{i\ell t}\right), \tag{11}$$

where $\mathbf{X}_{i\ell t}$ is a tuple including the normalized spot rate $\tilde{p}_{\ell t}/\text{Rate}_\ell$, the log of average monthly volume and the average coefficient of variation of weekly volume on lane ℓ of shipper i across all months of the auction period. As argued in Harris and Nguyen (2021), the latter two variables affect the desirability of a shipper-lane: (i) the frequency of interactions affects the continuation values of the relationship of both the shipper and carrier, and (ii) the consistency of offers captures the extent to which the carrier can benefit from planning.

[35] Acocella, Caplice, and Sheffi (2020) identify breaks in the time series of rejection rates to define market phases. They identify the period from April 15, 2016 to April 14, 2017 as a full year in a soft market, and the period from October 1, 2017 to September 30, 2018 as full year a tight market. Since we assume that each relationship belongs to a market phase, we use the time of the auction to classify a relationship's market phase. Specifically, relationships that start in 2017 or 2018 belong to a tight market.

Operational and search costs. Let (per-mile) operational costs of carrier j on lane ℓ of shipper i in auction a be distributed as a Normal distribution with mean $\mu^c_{ia\ell}$ and common standard deviation σ^c,

$$c_{j\ell t} \sim \text{Normal}(\mu^c_{ia\ell}, \sigma^c).$$

Thus, the transformed cost is distributed as $\text{Normal}(\tilde{\mu}^c_{ia\ell}, \sigma^c)$, where $\tilde{\mu}^c_{ia\ell} = \mu^c_{ia\ell} + \kappa_\ell$.

To decompose the transformed cost into operational and search costs, we make two further parametric assumptions. First, per-load search costs are linked to spot market thickness through a scale efficiency parameter γ_1, giving per-mile search costs[36]

$$\kappa_\ell = \frac{\gamma_0 + \gamma_1 \ln(\text{Volume}^{\text{spot}}_\ell)}{\text{Distance}_\ell}. \tag{12}$$

Second, we allow operational costs to vary flexibly with distance, differ across market phases (soft or tight), and have unobserved differences. Specifically,

$$\mu^c_{ia\ell} = \gamma_2 \mathbf{1}^a_{\text{tight}} + h(\text{Distance}_\ell) + \nu^c_\ell + \epsilon^c_{ia\ell}, \tag{13}$$

where h is a flexible function. The regression of the mean of transformed costs $\tilde{\mu}^c_{ia\ell} = \mu^c_{ia\ell} + \kappa_\ell$ as defined by Equations (12) and (13) has a potential endogeneity issue: lanes with higher unobserved cost shifter ν^c_ℓ tend to have lower volume in equilibrium. We resolve this issue by instrumenting for realized spot volume with a demand shifter, the predicted trade flows across US states (Caliendo, Parro, Rossi-Hansberg, & Sarte, 2018).

Number of bidders. We allow for a stochastic number of bidders in each auction, $N_a \sim \text{Binomial}(N, q)$. When bidding, a carrier only knows (N, q) and not N_a.

6.3 Estimation procedure

Our estimation procedure follows the steps of the identification argument closely. First, we obtain instrumental objects, including the discount factor, the distribution of the number of bidders, the spot process, and the shippers' incentive scheme. These objects are input into the remaining steps, in which we estimate the cost parameters, the distribution of rents and the distribution of match-specific gains.

Using relationships of at least 50 offers, we estimate cost parameters in two steps. First,

[36] Our functional-form assumption on per-mile search costs comes from our interpretation of search costs as fixed costs. This interpretation implies that per-mile search costs and per-mile thickness externalities are smaller on longer lanes. We empirically verify this implication by considering alternative functional-form assumptions on search costs. See Appendix 10 for our estimates of search costs under different specifications.

we estimate parameters of transformed costs by maximum likelihood (Rust, 1994). The likelihood contribution of each relationship is the likelihood of the carrier's accept/reject decisions. This likelihood contribution depends on the auction-specific mean $\tilde{\mu}_{ia\ell}$ of transformed costs, the common cost variance σ^c, and the carrier's transformed rent $\eta_{ij\ell}+p_{ij\ell}+\kappa_\ell$. For each relationship and each set of parameter values, we use a fixed-point algorithm to solve for the carrier's value function and optimal strategy. The full likelihood aggregates all relationships and is maximized in two layers. In the outer loop, we search for the value of the common cost variance σ^c on a grid. In the inner loop, we jointly estimate auction-specific means $(\tilde{\mu}_{ia\ell})_{ia\ell}$ of transformed costs, and relationship-specific carriers' transformed rents $(\eta_{ij\ell}+p_{ij\ell}+\kappa_\ell)_{ij\ell}$. Second, we decompose the estimated means of transformed costs into search and operational costs by estimating the following equation by two-stage least squares,

$$\tilde{\mu}^c_{ia\ell} = \frac{\gamma_1 \ln(\text{Volume}^{\text{spot}}_\ell)}{\text{Distance}_\ell} + \gamma_2 \mathbf{1}^a_{\text{tight}} + h(\text{Distance}_\ell) + \nu^c_\ell + \epsilon^c_{ia\ell}, \tag{14}$$

using predicted trade flows (Caliendo, Parro, Rossi-Hansberg, & Sarte, 2018) as an instrument for spot volume. Since we use a flexible function h to control for distance, this regression does not pin down the base per-load search cost γ_0. To obtain an estimate of this parameter, we normalize the median operational costs across all lanes to industry estimates.[37]

Given our cost estimates, we estimate the distribution of rents and the distribution of match-specific gains. We perform this step separately for each of ten clusters of lanes, to ensure a sufficient number of relationships in each lane-specific cluster, and separately for soft and tight market, to allow for different demand and supply factors across the two market phases. Lane-specific clusters are constructed by K-means clustering based on the average spot rate, average spot volume, and distance. The idea is to identify the latent groups of lanes, as defined by fundamental demand and supply factors, by clustering on informative equilibrium fixed effects.[38] Thus, relationships in each of our clusters are potentially different in match-specific gains but similar in other characteristics.

Within each cluster, the distribution of shippers and carriers' rents, and their match-specific gains are estimated in three steps. First, we use an EM-algorithm (Train, 2008) to estimate the distribution of winning carriers' rents. Second, we piece together the monotone

[37]Using an accounting approach, Williams and Murray (2020) estimate the marginal cost of trucking service to be \$1.55/mile, including fuel costs. Since long-term contracts typically separate payment on fuel costs as fuel surcharge, and spot rates in our data subtract fuel surcharge, we also subtract fuel surcharge (\$0.33/mile) from the accounting estimate.

[38]See Bonhomme and Manresa (2015), who propose clustering methods to identify latent groups based on observed patterns of heterogeneity in panel data. See Bester and Hansen (2016), who study grouped effects estimators.

mapping between carriers' rents and shippers' rents from the first-order condition of carriers' bidding evaluated at the percentiles of the distribution of winning carriers' rents. For the initial condition of the bidding function, we take the fifth percentile of this distribution as the lowest carrier rent and pin down the lowest shipper rent via shippers' individual rationality condition.[39] Finally, to obtain the fundamental distribution of match-specific gains, we use simulated methods of moments to match the distribution of carrier rent conditional on different bins of contract rates. See Appendix C for a detailed description of our estimation procedure.

7 Estimates of match-specific gains, search costs, and operational costs

This section presents our estimates of key model primitives. These estimates suggest that long-term relationships generate large expected surplus to participating parties, but they exert substantial negative externalities on the spot market. We find that on a typical lane, doubling the thickness of the spot market reduces search costs by an amount equivalent to reducing operational costs by 29%. High expected surplus from relationships come from high match-specific gains, and the fact that the current fixed rate contracts and relational schemes perform fairly well at realizing these gains. We estimate that the median relationship achieves 44% of the first best surplus for individual relationships, with significantly better performance for relationships with higher match-specific gains.

Our estimates of instrumental objects, including the discount factor, number of bidders and relational scheme can be found in Appendix D.1. Consistent with Harris and Nguyen (2021), we find that shippers punish carriers' rejections by increasing the probability of demotion in future periods. While soft, shippers' incentive scheme generates dynamic incentives that are economically significant.

7.1 Operational and search costs

One noteworthy finding on search costs is that increasing the thickness of the spot market reduces search costs. This supports the idea hypothesized by Kranton (1996), that the formation of long-term relationships can crowd out the spot market, making it thinner and less efficient. Table 5 reports the estimates of the parameter for scale efficiency (γ_1) and

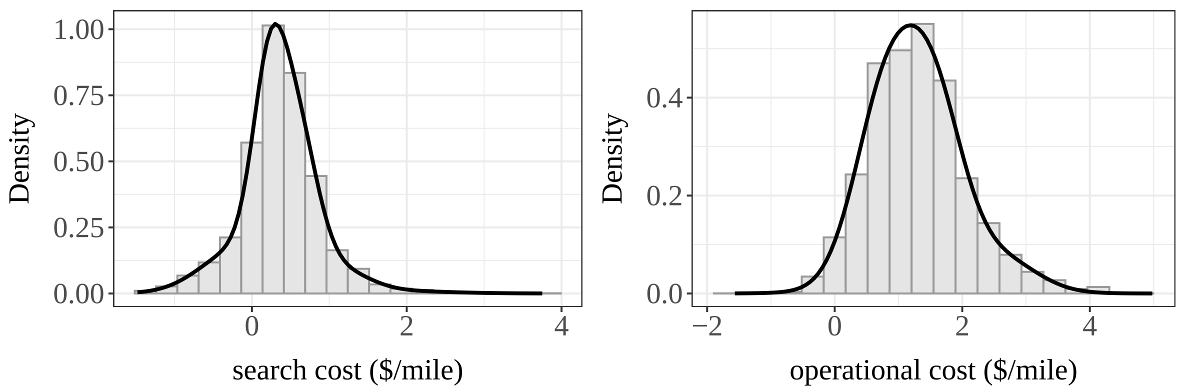
[39]Note that the lower is the carrier's rent, the higher is the shipper's rent required for the relationship to pass the shipper's individual rationality constraint. Moreover, the tail of the estimated distribution of carriers' rents tend to have larger errors. Thus, our choice of the fifth percentile of the distribution of carriers' rents as the lowest level of carrier rent errs on the side of not inflating estimated shipper rents.

Table 5: Estimates of cost determinants

	Estimate	95% CI
Scale efficiency (γ_1)	-255.85	$(-402.97, -146.26)$
Tight market	0.54	$(0.39, 0.64)$

Note: Confidence intervals are constructed by bootstrapping at the auction level.

Figure 8: Estimated distribution of search and operational costs across lanes



the average difference in operational costs between the tight and soft market. Our estimate of γ_1 is negative and economically significant. Specifically, we estimate that doubling the spot volume on a median lane of 500 miles decreases search costs by \$0.35/mile, an amount equivalent to 29% reduction in operational costs.[40]

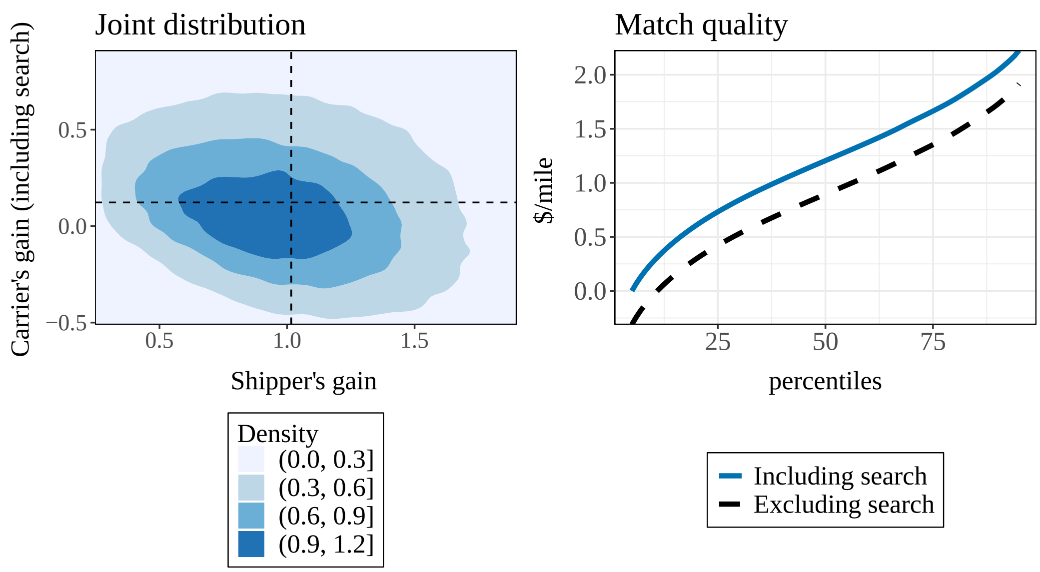
Figure 8 plots the distribution of estimated search costs and operational costs across lanes. The variation in estimated search costs comes from the large variation in spot volumes across lanes, and the residual variation in costs is captured by operational costs. To decompose search and operational costs from their sum, we relied on our estimate of γ_1 and a normalization. Specifically, we calibrated the base per-load search costs (γ_0 in Equation (12)) to match the median of our estimated operational costs to the industry estimate.[41] We estimate that the median search cost is \$0.35/mile, equal to 29% of the median operational costs. Note that while these estimates are sensitive to our normalization method, they are not crucial to our welfare comparison, which relies mostly on our estimate of the scale efficiency γ_1.

Finally, we find that operational costs are \$0.54/mile higher in a tight market than in a soft market. This reflects the capacity crunch reported during the period from early 2017 to

[40] Table 10 in Appendix D.2 presents different specifications of (per-mile) search costs. The results show a strong link between search costs and spot volumes and that this link is weaker on longer lanes.

[41] Normalizing the median operational cost across all lanes to the industry estimate pins down $\hat{\gamma}_0 = 1494.068$. In addition, the standard deviation of operational costs is estimated to be $\hat{\sigma}^c = 1.2$ (\$/mile).

Figure 9: Distribution of match-specific gains and match quality



Note: The dashed lines on the left panel indicate the median levels of carrier's match-specific gain (including savings on search costs) and shipper's match-specific gain in realized relationships.

the end of 2018.[42] This also means that the increase in rejection rates observed in Figure 1 is due to both an incentive effect and an aggregate cost effect.

7.2 Match-specific gains

We find large and heterogeneous match quality in realized relationships, which is accounted for mostly by shippers' match-specific gains. The left panel of Figure 9 plots the density of the joint distribution of carriers' match-specific gains including savings on search costs ($\eta + \kappa$) and shippers' match-specific gains (ψ) in realized relationships. Darker colors demonstrate values of shippers and carrriers' match-specific gains with higher density. We find that in the median relationship, the shipper's match-specific gain is \$1.02/mile, accounting for 89% of the sum of the shipper and carrier's match-specific gains, or their match quality. The median carrier's match-specific gain is relatively small (\$0.12/mile) and in fact smaller than the median savings on search costs (\$0.35/mile). These findings are consistent with our descriptive results in Section 4.3, that (i) shippers do not reduce volume offered within relationships when spot rates are low, while (ii) a fair share of carriers accept much

[42]Industry reports suggest that the capacity crunch during 2017 and 2018 is driven by both rising demand for trucking services and supply-side factors, such as the Electronic Logging Device mandate and the enforcement of Hours of Service Law starting in January 1st, 2018. This capacity crunch is also reflected in the hourly wage of truck drivers. According to data from the American Transportation Research Institute, average hourly wage of truck drivers in the period from 2008 to 2020 peaked in 2018.

less frequently when spot rates are high. Moreover, the slight negative correlation between shippers and carriers' match-specific gains in realized relationships is due to the selection of relationships with high match quality.

The right panel of Figure 9 plots the quantile distribution of match quality, including and excluding savings on search costs. Even when savings on search costs are excluded, the estimated match quality is large and heterogeneous across relationships, increasing from \$0.42/mile to \$1.35/mile from the 25th to the 75th percentiles. Such large heterogeneity in match-quality reflects the importance of auctions in facilitating the formation of high-value matches in the truckload setting. Moreover, excluding savings on search costs, 10% of relationships have negative intrinsic match quality. This means that the dominance of long-term relationships in the current institution is self-fulfilling only to a limited extent.

7.3 Shippers and carriers' rents

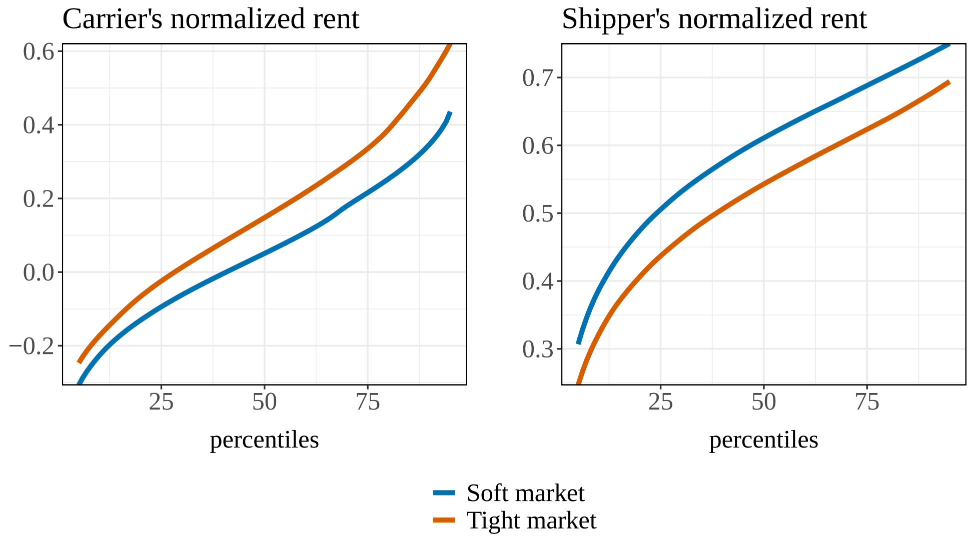
The shipper and carrier's match-specific gains affect welfare only through how their match quality $(\psi+\eta+\kappa)$ is split into the carrier's per-transaction rent $(\eta+p+\kappa)$ and the shipper's per-transaction rent $(\psi-p)$ by the equilibrium contract rate (p). Since these rents are realized only when the carrier accepts an offer of the shipper, the role of the equilibrium contract rate is not merely distributional. When the carrier has a higher rent from acceptance, it accepts more frequently, realizing the match-specific gains for both itself and the shipper.

The equilibrium split of total match quality between shippers and carriers is driven by three forces. First is individual rationality: since carriers have the option to reject shippers' offers as spot rates vary, shippers require large average rents to benefit from relationships, while carriers can benefit from relationships even with small average rents. Second is a competition effect: that carriers bid for relationships in auctions makes the split of match quality into rents more favorable towards shippers. Third is an "efficiency wage" effect: shippers may prefer leaving higher rents to carriers to induce more acceptances.

We estimate large and heterogeneous rents for shippers and carriers from long-term relationships. Figure 10 plots the quantile distribution of average carriers and shippers' rents normalized by the mean spot rates, across soft (blue) and tight (red) markets. In the median relationship, the carrier has a normalized rent of 10% and the shipper has a normalized rent of 58%. From the 25th to the 75th percentiles, the carrier's normalized rent increases from −5% to 39% and the shipper's normalized rent increases from 45% to 65%.

Finally, carriers are compensated more in a tight market than in a soft market. The reason is that the better outside option of carriers in a tight market, when spot rates are high, improves their outcomes in auctions for relationships due to both the competition and

Figure 10: Quantile distribution of per-transaction normalized rents



Note: Carrier's normalized rent: $\frac{(\eta+p+\kappa)-\text{Rate}}{\text{Rate}}$; shipper's normalized rent: $\frac{\text{Rate}-(\psi-p)}{\text{Rate}}$; where Rate is the mean spot rate over the entire sample period. The normalization is performed separately for each cluster of lanes, and the plots are of averaged normalized rents.

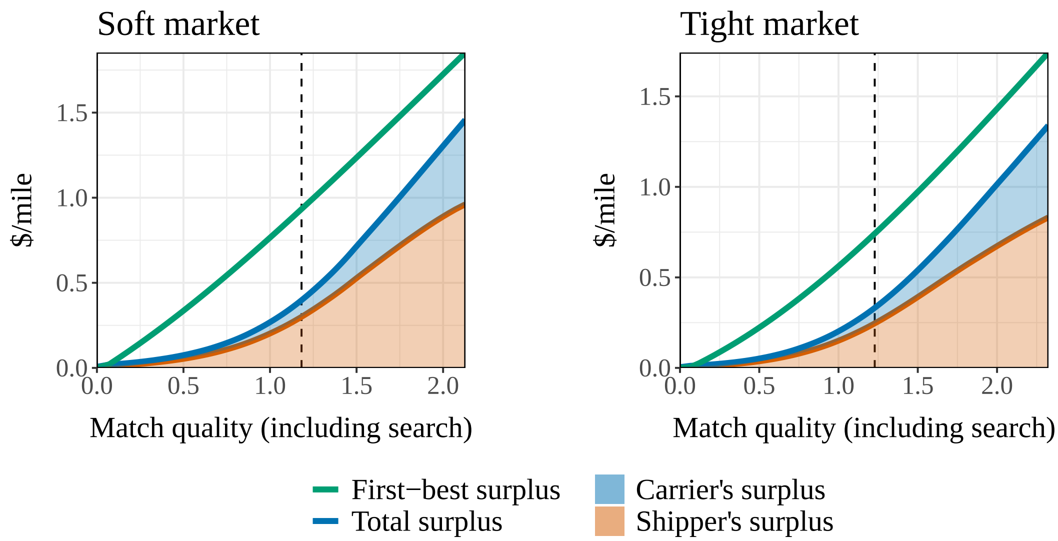
efficiency wage effects. However, the increase of 10% in carriers' normalized rents does not fully compensate carriers for the increase in spot rates from a soft to a tight market.

7.4 Welfare analysis of individual relationships

We transform our estimates of the incentive scheme and per-transaction rents into shippers and carriers' expected surplus from long-term relationships over spot transactions. That is, this exercise takes into account both the intensive margin (equilibrium rejections) and extensive margin (equilibrium demotions) of relationships. We then benchmark the joint expected surplus of an individual relationship in the current institution against the relationship-level first-best surplus. The latter requires that (i) relationships never end, and (ii) the first-best outcome is achieved in each period. Specifically, for a relationship with match quality (inclusive of savings on search costs) $\psi + \eta + \kappa \geq 0$, the carrier should never service the spot market; it should accept when its cost after internalizing the joint match-specific gains is less than the shipper's payment in the spot market, $c_t - (\psi + \eta) \leq \tilde{p}_t$, and remain idle otherwise. Figure 11 plots the expected surplus from long-term relationships (blue line) under current fixed-rate contracts, how it is split between shippers (red area) and carriers (blue area), against the relationship-level first-best surplus (green line).

One of our main findings is that the current fixed-rate contracts and incentive scheme do a fair job at capturing the first-best surplus, with significantly better performance in

Figure 11: Expected surplus from long-term relationships



Note: The dash vertical lines indicate the median match quality (including savings on search cost) in each market.

relationships with higher match quality. Specifically, the median relationship achieves 44% of the relationship-level first-best surplus; this figure increases from 27% to 62% from the 25th to the 75th percentile of match quality. The reason for such heterogeneity is that relationships with lower match quality face a more serious moral hazard problem and have less room to use contract rates as an incentive instrument.

In terms of distributional effects, we find that shippers have a larger share of total expected surplus. At the median relationship, shippers enjoy 75% of total expected surplus. This reflects our estimate that, on average, there are three effective bidders per auction. The share of carriers in total expected surplus slightly increases with their relationships' match quality. This reflects the higher information rents of carriers with higher match quality.

7.5 Key insights

Our findings show that the two-way crowding-out effects between long-term relationships and the spot market, as hypothesized by Kranton (1996), are large in our setting. On the one hand, long-term relationships result in a thinner spot market with significantly higher search costs. Specifically, were we to double the thickness of the spot market, search costs would reduce by about $0.35/mile. On the other hand, the current fixed-rate contracts allow the spot market to crowd out long-term relationships, achieving 44% of the relationship-level first-best surplus in the median relationship. Furthermore, the crowding-out effect of the spot market is stronger in relationships with a lower match quality. Such selectivity tends to be beneficial to market-level welfare. This is because to achieve the same level of spot market thickness, it is generally more efficient to forgo transactions that generate low match quality.

These findings beg the following questions: Is the current "high-relationship" equilibrium socially optimal or just self-fulfilling? What are the welfare effects of increasing the spot market thickness or improving the performance of long-term relationships? The next section will shed light on these questions.

8 Market-level welfare under alternative institutions

At the market level, there is a tradeoff between realizing more match-specific gains generated by long-term relationships and maintaining greater thickness of the spot market. We quantify this tradeoff by comparing the market-level welfare of the current institution to alternative institutions that change the share or performance of long-term relationships. The first counterfactual institution is a centralized spot market for maximal spot market thickness. The second counterfactual replaces all fixed-rate contracts with the individually optimal index-priced contracts. We find that a centralized spot market would result in substantial welfare loss, and index pricing would be beneficial, though only in periods with high demand. We also construct an upper bound on the market-level first-best surplus and find that the current institution achieves 40% of this surplus on medium trips of 500 miles and 60% of this surplus on long trips of 1000 miles.

8.1 Economic tradeoffs

The welfare effects of different market institutions depend on the own benefits of long-term relationships and the spot market, as well as how these two forms of transactions interact. First, long-term relationships generate large match-specific gains to participating parties, while the spot market, by centralizing more transactions, may reduce search costs and improve allocative cost efficiency. Second, there are two-way crowding effects between relationships and the spot market in the US for-hire truckload freight industry. On the one hand, the spot market creates carriers' moral hazard problem, crowding out low-value relationships. On the other hand, the formation of long-term relationships results in higher search costs in a thinner spot market. We quantify the welfare effects of these economic forces in two counterfactual exercises.

A centralized spot market (no relationships). The first counterfactual is a spot platform that centralizes all transactions in the market. This achieves the optimal scale efficiency, thus reducing search cost on the spot market and increases allocative cost efficiency; the tradeoff is the complete loss of match-specific gains in long-term relationships.

Index-priced contracts (optimal relationships). This counterfactual keeps the auctions for relationship formation but replaces all fixed-rate contracts with the individually optimal index-priced contracts. These contracts take the standard idea in the contract literature that to solve a moral hazard problem in a principal-agent relationship, the principal should "sell the firm" to the agent. And to screen out the best relationship, the principal should ask all agents to bid on the contract. In an index-priced contract, it means that the shipper transfers all of the rents from each realized transaction to the carrier, and asks all carriers to bid on a fixed fee to be paid for each offer, regardless of whether the carrier accepts or rejects. While such index-priced contracts eliminate the moral hazard problem, achieving the first-best surplus of individual relationships, they would exacerbate the negative externalities of relationships on the spot market.

Definition 3. (Individually optimal index-priced contracts.) An individually optimal index-priced contract between shipper i and carrier j on lane ℓ is pegged one-to-one to the spot rate and internalizes shipper's match-specific gain in the following way

$$p_{ij\ell}(\tilde{p}_t) = \begin{cases} -\, b^0_{ij\ell} + \overbrace{\psi_{ij\ell} + \tilde{p}_{\ell t}}^{\text{"incentive"}} & \text{if carrier } j \text{ accepts} \\ \underbrace{-b^0_{ij\ell}}_{\text{"screening"}} & \text{if carrier } j \text{ rejects,} \end{cases}$$

where $b^0_{ij\ell}$ is a fixed fee on which carrier j bids in shipper i's auction on lane ℓ.

8.2 Welfare comparison

We calculate the market-level welfare of the current institution and the two alternatives on each cluster of lanes for a full-year period in each market phase. Following Acocella, Caplice, and Sheffi (2020), we take the period from April 15, 2016 to April 14, 2017 as a full year in a tight market, and the period from October 1, 2017 to September 30, 2018 as a full year in a soft market. Our welfare calculation involves two steps. First, we recover the underlying demand and supply factors using the market equilibrium condition in Equation (8). Within each cluster of lanes and each market phase, we recover L relationships, all formed in $t = 0$. For each week t, we recover a direct spot demand D_t and a spot capacity

C_t. Second, we calculate the welfare in the current and counterfactual institutions, keeping fixed the underlying demand and supply factors $(L, D_t, C_t)_t$, distribution of match-specific gains and operational costs, while allowing search costs to vary with the equilibrium thickness of the spot market.[43]

8.2.1 Aggregate welfare

Figure 12 and Table 6 present the per-mile average welfare in three institutions: (i) a centralized spot market for all transactions ("None"), (ii) long-term relationships with fixed-rate contracts coexisting with a spot market ("Fixed-rate"), and (iii) long-term relationships with the individually optimal index-priced contracts coexisting with a spot market ("Index-priced"). To detect the sources of gains and losses, we break down the average welfare into three components: realized match-specific gains, operational costs and search costs. Note that our welfare calculation treats demand for transportation service as inelastic and our welfare measure excludes the benefits to shippers from having their loads transported.[44] That is, we answer the question on which institution is more efficient at fulfilling a fixed number of loads.

First, we find that centralizing all transactions into a spot platform results in substantial welfare loss from the current institution. The welfare loss is 38 cents/mile in a soft market and 39 cents/mile in a tight market, which are equivalent to 31% and 22% of the median operational cost in the respective market phase. Note that while the reduction in search costs, by 36 cents/mile in a soft market and 21 cents/mile in a tight market, is large, it only benefits those serving in the spot market. Moreover, the reduction in operational costs is small. As a result, the reduction in search and operational costs is far from compensating for the complete loss of match-specific gains from long-term relationships.

Second, individually optimal index-priced contracts can be welfare-improving upon the current fixed-rate contracts, but only in a tight market, when demand for transportation service is high. The reason for this difference is that, in a tight market, relationships with fixed-rate contracts face a more serious moral hazard problem. Eliminating moral hazard by individually optimal index-priced contracts thus brings about larger gains in a tight market than it does in a soft market. Specifically, index-priced contracts lead to an increase in match-specific gains relative to the current institution by 11% and 28% in the soft and the tight market respectively. Furthermore, index-priced contracts result in a higher search cost in the spot market, by 20 cents/mile in a soft market and 26 cents/mile in a tight

[43]The details of these steps are delegated to Appendix C.2.

[44]That is, we normalize shippers' gains from spot transactions to zero. This is why the average welfare is calculated to be negative.

Figure 12: Welfare comparison across three market institutions

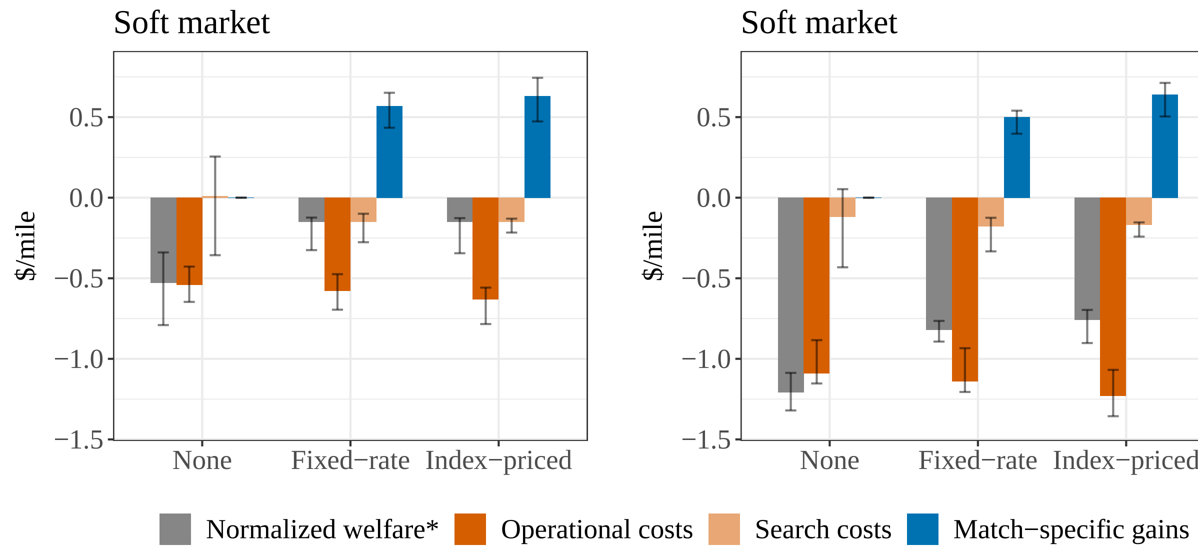


Table 6: Summary of welfare channels

	Soft market			Tight market		
Relationship type	None	Fixed-rate	Index-priced	None	Fixed-rate	Index-priced
Spot rate	1.56	1.63	1.47	2.07	2.06	1.92
Spot share	100%	44%	29%	100%	61%	34%
Δ Search cost	−0.36		0.20	−0.21		0.26
Match-specific gains		0.57	0.63		0.50	0.64
Operational costs	−0.54	−0.58	−0.63	−1.09	−1.14	−1.23
Search costs	0.01	−0.15	−0.15	−0.12	−0.18	−0.17
Normalized welfare*	−0.53	−0.15	−0.15	−1.21	−0.82	−0.76

Notes: *Our welfare measure normalizes the shippers' benefits from spot transactions to zero. That is, it includes carriers' costs and match-specific gains and shippers' match-specific gains. All numbers are in $/mile.

market, but more loads would be accepted within long-term relationships and they entail no search costs. In aggregate, these two forces balance out, resulting in no difference in average search costs between the two types of contracts. There is, however, a more subtle effect of an increase in search costs on aggregate welfare. Higher search costs create a larger wedge in realized operational costs between carriers servicing in long-term relationships and those servicing in the spot market. This tends to hurt allocative cost efficiency. Across both market phases, operational costs increase by about 9% under index-priced contracts relative to fixed-rate contracts. Another factor contributing to this increase is that carriers in long-term relationships with index-priced contracts fully internalize match-specific gains, thus accepting even when operational costs are high.

8.2.2 Distributional effects

Both counterfactual institutions have large distributional consequences in comparison to the current institution. Figure 13 and Table 7 present the average per-period payoff ($/mile) of each individual in the market. There are four groups: shippers and carriers who form relationships, and shippers and carriers who only transact in the spot market. We will refer to the latter two groups as spot shippers and spot carriers.

Figure 13: Distributional effects

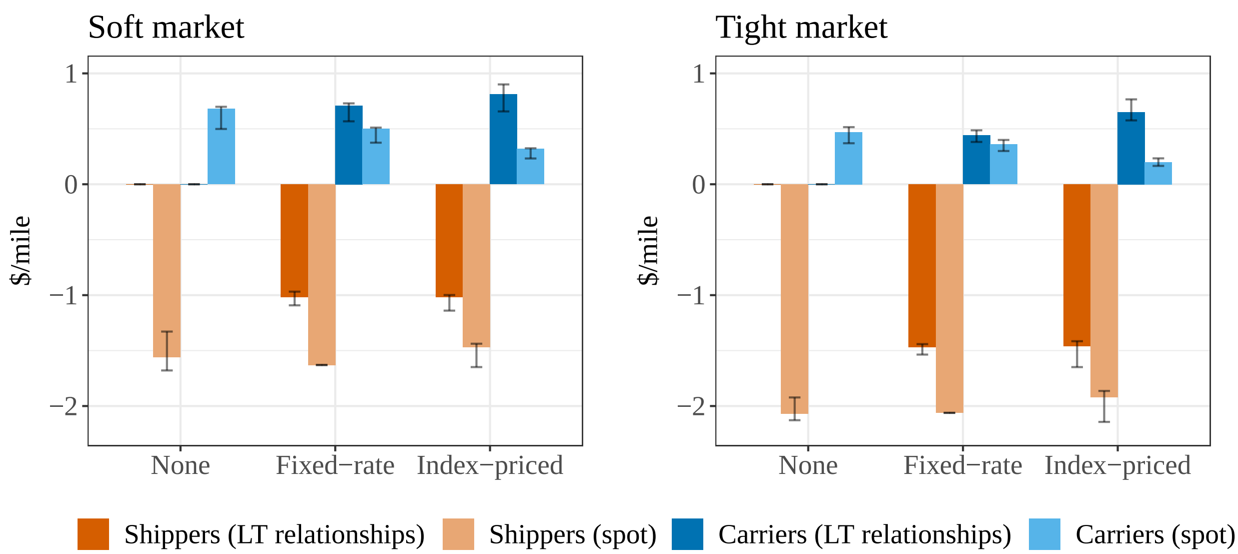


Table 7: Distributional effects

	Soft market			Tight market		
Relationship type	None	Fixed-rate	Index-priced	None	Fixed-rate	Index-priced
LT relationships						
Shippers		−1.02	−1.02		−1.47	−1.46
Carriers		0.71	0.81		0.44	0.65
Spot market						
Shippers	−1.56	−1.63	−1.47	−2.07	−2.06	−1.92
Carriers	0.68	0.50	0.32	0.47	0.36	0.20

Notes: Shippers' benefits from having their loads shipped are normalized to zero. All numbers are in $/mile.

Overall, institutions with long-term relationships tend to benefit those who manage to form relationships and hurt those who transact only in the spot market. Specifically, better performance of long-term relationships affects the spot market by (i) reducing demand for spot loads and (ii) increasing search costs for spot loads. Both of these channels unambiguously hurt spot carriers. Specifically, relative to a centralized spot platform, the average welfare of a spot carrier reduces by approximately 25% under fixed-rate contracts and 55% under index-priced contracts in both soft and tight market. Spot shippers are affected via equilibrium spot rates, with channel (i) pushing towards lower spot rate and channel (ii) pushing towards higher spot rate. Across the three institutions, the institution with index-priced contracts in long-term relationships has the lowest equilibrium spot rate ($1.47/mile), which benefits spot shippers.

Within long-term relationships, index-priced contracts improve the welfare of carriers and not shippers upon fixed-rate contracts. This difference is because the split of surplus from long-term relationships is more favorable to shippers under fixed-rate contracts and more favorable to carriers under index-priced contracts. As the potential surplus from long-term relationships is fully extracted under index-priced contracts, carriers get more information rents in the auctions. Specifically, relative to fixed-rate contracts, index-priced contracts increase the average welfare of carriers in long-term relationships by 14% in a soft market and 48% in a tight market.

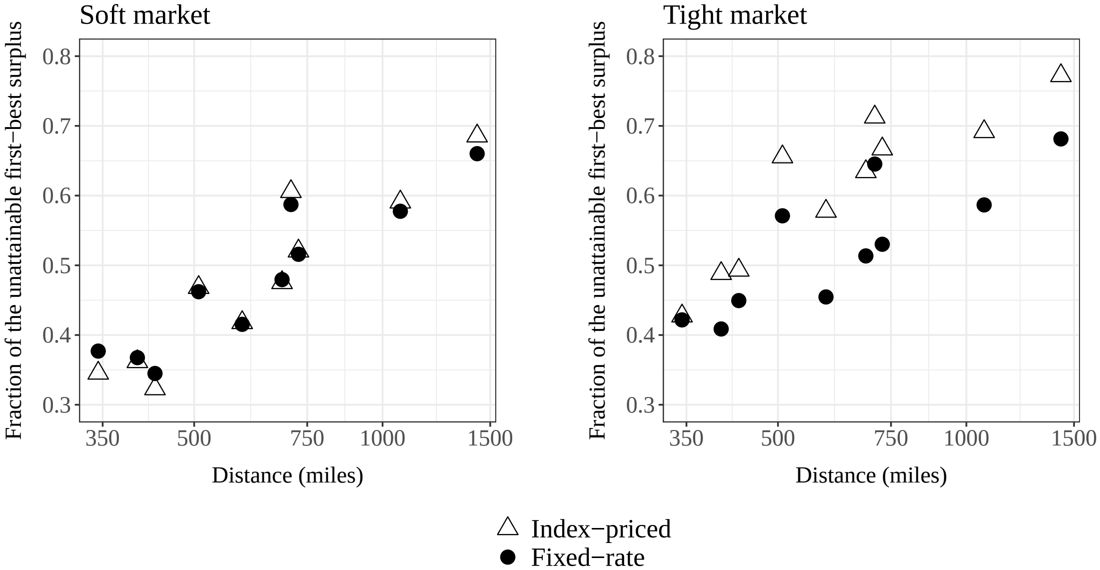
8.2.3 Lower-bound comparison to the market-level first-best welfare

In this section, we provide an upper bound on the market-level first-best welfare to benchmark the performance of fixed-rate and index-priced contracts. Since a centralized spot market is unambiguously the worst performing institution, we use it as a baseline for normalization; market-level welfare gain from this baseline will be referred to as market-level surplus.

An upper bound on the market-level first-best welfare. Since search costs are determined endogenously by spot market thickness, it is difficult to calculate market-level first-best welfare. However, an intuitive (strict) upper bound of this welfare measure can be constructed within our framework. First, we fix search costs to the level obtained in a centralized spot market, which is the lowest feasible level of search costs across all market institutions. Then, we exploit index-priced contracts to achieve allocative efficiency, internalizing all match-specific gains from transactions within long-term relationships and the fixed level of search costs for spot transactions.[45] In other words, our upper bound on the

[45]Index-priced contracts achieve allocative efficiency conditional on a fixed level of search costs by (i) building an aggregate cost curve that internalizes all match-specific gains and search costs and (ii) using a price mechanism to clear the market. Specifically, this aggregate cost curve is made up of the following components. First, each carrier j in a long-term relationship under index-priced contract with shipper i on lane ℓ provides service in period t if its internalized cost is less than the equilibrium spot rate, $c_{j\ell t} - (\psi_{ij\ell} + \eta_{ij\ell}) \leq \tilde{p}_{\ell t}$. Second, each spot carrier j' provides service in period t if the sum of its operational and search costs is less than the equilibrium spot rate, $c_{j'\ell t} + \kappa_\ell \leq \tilde{p}_{\ell t}$. However, the aggregate cost curve is determined endogenously, shifting upwards due to higher search costs as more match-specific gains in long-

Figure 14: Comparison to an upper bound on the market-level first-best welfare



market-level first-best welfare shuts down the negative externalities of transactions in long-term relationships on the spot market. We refer to this upper bound as the unattainable first-best and use it to evaluate the performance of fixed-rate and index-priced contracts. That is, the relative performance of fixed-rate and index-priced contracts to the unattainable first-best provides a lower bound comparison of the welfare of these institutions to the market-level first-best welfare.

Figure 14 plots the ratio of the market-level surplus of fixed-rate and index-priced contracts to the market-level surplus of the unattainable first-best, across our ten clusters of lanes and two market phases. Here, the market-level surplus of an institution is defined as the improvement in market-level welfare from a centralized spot market. Moreover, we order the clusters of lanes by their average distance, since search costs vary less with spot market thickness on longer lanes.

A noteworthy finding is that fixed-rate contracts perform quite well at the market level, capturing around 40% to 70% of the market-level surplus of the unattainable first-best. Index-priced contracts perform similarly to fixed-rate contracts in a soft market and outperform fixed-rate contracts in a tight market. Furthermore, the performance of both fixed-rate and index-priced contracts improves with the distance of lanes, reflecting the lower effect of spot market thickness on search costs on longer lanes. However, the magnitude of this improvement is inflated by the fact that when search costs vary less with spot market thickness,

term relationships are realized. This is why individually optimal index-priced contracts are not necessarily socially optimal.

our upper bound on the market-level first-best welfare is also tighter.

8.3 Discussion

While our market-level welfare analysis shows that a fully centralized spot market is unambiguously worse than both institutions with long-term relationships, it does not suggest that the role of the spot market should be downgraded. In contrast, the spot market provides an important clearing mechanism for hybrid institutions where long-term relationships and the spot market coexist. In particular, index-priced contracts require reliable measures of spot rates, which will not be available if the spot market is too thin. Moreover, technological advances in the near future can improve the spot market in more ways than by reducing search costs. For example, a digital spot platform can suggest carriers to different lanes to exploit network externalities, thus mitigating the aggregate empty mile problem.

Our comparison of fixed-rate and index-priced contracts demonstrates that the contract design of individual relationships can have market-level consequences. For example, we find that during a soft market and on lanes of shorter distance, fixed-rate contracts, which are suboptimal at the relationship level, generate higher market-level welfare than index-priced contracts. It would be interesting to examine the market-level welfare effects of contracts between these two extremes. In fact, there may be barriers to the implementation of individually optimal index-priced contracts. For example, price uncertainties and the fixed fees that carriers need to pay for rejections would pose concerns for shippers and carriers with budget constraints. Acocella, Caplice, and Sheffi (2022a) study the relationship-level effects of index-based contracts that are tuned to practitioners' concerns. Our framework can be used to evaluate the market-level performance of such contracts.

9 Conclusion

This paper studies the interactions and welfare effects of long-term relationships and the spot market. Using detailed data on the US for-hire truckload freight industry, we argue that the two-way crowding-out effects between long-term relationships and the spot market, as hypothesized by Kranton (1996), are present and play a crucial role in this setting. On the one hand, the spot market crowds out long-term relationships by creating a moral hazard problem within relationships. On the other hand, long-term relationships crowd out the spot market by reducing spot market thickness, thereby increasing search costs for spot loads.

Our methodological contribution is a model that captures both the formation of and interactions within long-term relationships. We model relationship formation as an auction and interactions within the winning relationship as a repeated game. We recover a rich set of model primitives by building on tools from the empirical auction and dynamic discrete choice literature. An empirical challenge of the dynamic discrete choice problem in our setting is that payoff-relevant actions are only partially observed. We tackle this challenge with a novel support-based argument. Specifically, we exploit the sensitivity of the observed decision margin (between "accept" and "reject") to a running variable (the current spot rate) that affects the unobserved decision margin (between "spot" and "idle") to pin down the latter decision margin. This identification approach could be generalized to other settings with moral hazard, where actions are naturally not fully observed. Moreover, we develop an identification argument for auctions with two-sided match-specificity. Our argument relies on the observation that the shipper's and the carrier's expected payoffs in a relationship depend on their match-specific gains only through their per-transaction rents. Under empirically plausible conditions, we derive equilibrium conditions that permit a monotone mapping between the carrier's rent and the shipper's rent and pin down this mapping by an approach similar to Guerre, Perrigne, and Vuong (2000). More generally, we demonstrate how in settings where parties share rents, classic insights from the auction literature apply even through there are two latent variables with unrestricted correlation.

Estimating the model, we find that long-term relationships generate large match-specific gains, but realizing more of these gains would come at the cost of a thinner spot market with significantly higher search costs. This market-level tradeoff is a key consideration in evaluating different market institutions. Future technological innovations could either threaten to replace relationships with a more efficient spot market or enhance relationships with more sophisticated contract design. In either case, the welfare effects would be determined by the two-way crowding-out effects that we estimate.

Our counterfactual analysis suggests that the benefits of long-term relationships outweigh their negative externalities. However, this does not mean that the market unambiguously benefits from optimizing the performance of relationships. On the one hand, removing relationships would result in substantial welfare loss, despite achieving the maximal thickness of the spot market. This finding suggests that the dominance of long-term relationships in the current institution is not driven by a coordination failure to form a thick spot market but rather by the large match-specific gains from long-term relationships. On the other hand, optimizing the performance of individual relationships with index-priced contracts leads to only small improvements in market-level welfare and only in periods with high demand. The reason is that such contracts worsen the negative externalities of relationships on the spot market.

While this paper answers key questions about the interactions between long-term relationships and the spot market, there are several possible directions for future research in this setting. First, one could examine the role of brokers in the current market institution, especially in how shippers and carriers search for and haggle on loads. Second, one could extend the model to allow for interactions across lanes, for example, by letting carriers jointly manage multiple relationships and have access to the spot market on different lanes.

More broadly, our paper is a first step towards understanding the role of technological innovations in shifting the boundary between formal and informal interactions. On the one hand, technological innovations can improve the performance of informal interactions by allowing them to incorporate market information. On the other hand, technological innovations can enable new forms of formal interactions that compete with informal interactions. As technology-driven changes take hold in various industries, understanding the implications of both possibilities is key to anticipating, understanding, and responding to technology's effects.

References

Acocella, A., Caplice, C., & Sheffi, Y. (2020). Elephants or goldfish?: An empirical analysis of carrier reciprocity in dynamic freight markets. *Transportation Research Part E: Logistics and Transportation Review*, *142*, 102073.

Acocella, A., Caplice, C., & Sheffi, Y. (2022a). *The end of set-it-and-forget-it pricing? Index-based freight contract design under uncertainty* (MIT Center for Transportation & Logistics Working Paper Series).

Acocella, A., Caplice, C., & Sheffi, Y. (2022b). *When suppliers ignore their contracts: Quantifying the value of a customer in the buyer-supplier relationship* (MIT Center for Transportation & Logistics Working Paper Series).

Adhvaryu, A., Bassi, V., Nyshadham, A., & Tamayo, J. A. (2020). *No line left behind: Assortative matching inside the firm* (Working Paper). National Bureau of Economic Research.

Admati, A. R., & Pfleiderer, P. (1988). A theory of intraday patterns: Volume and price variability. *The Review of Financial Studies*, *1*(1), 3–40.

Allen, J., & Wittwer, M. (2021). *Centralizing over-the-counter markets?* (Working Paper). Bank of Canada Staff Working Paper.

Baker, G. P., Gibbons, R., & Murphy, K. J. (1994). Subjective performance measures in optimal incentive contracts. *The Quarterly Journal of Economics*, *109*(4), 1125–1156.

Baker, G. P., & Hubbard, T. N. (2003). Make versus buy in trucking: Asset ownership, job design, and information. *American Economic Review*, *93*(3), 551–572.

Baker, G. P., & Hubbard, T. N. (2004). Contractibility and asset ownership: On-board computers and governance in us trucking. *The Quarterly Journal of Economics*, *119*(4), 1443–1479.

Barron, D., Gibbons, R., Gil, R., & Murphy, K. J. (2020). Relational adaptation under reel authority. *Management Science*, *66*(5), 1868–1889.

Bester, C. A., & Hansen, C. B. (2016). Grouped effects estimators in fixed effects models. *Journal of Econometrics*, *190*(1), 197–208.

Biais, B., & Green, R. (2019). The microstructure of the bond market in the 20th century. *Review of Economic Dynamics*, *33*, 250–271.

Bonhomme, S., & Manresa, E. (2015). Grouped patterns of heterogeneity in panel data. *Econometrica*, *83*(3), 1147–1184.

Brancaccio, G., Kalouptsidi, M., Papageorgiou, T., & Rosaia, N. (2020). *Search frictions and efficiency in decentralized transportation markets* (Working Paper). National Bureau of Economic Research.

Brugues, F. (2020). *Take the goods and run: Contracting frictions and market power in supply chains* (Working Paper). Brown University.

Buchholz, N. (2022). Spatial equilibrium, search frictions, and dynamic efficiency in the taxi industry. *The Review of Economic Studies*, *89*(2), 556–591.

Cajal-Grossi, J., Macchiavello, R., & Noguera, G. (2022). *Buyers' sourcing strategies and suppliers' markups in Bangladeshi garments* (Working Paper).

Caliendo, L., Parro, F., Rossi-Hansberg, E., & Sarte, P.-D. (2018). The impact of regional and sectoral productivity changes on the us economy. *The Review of Economic Studies*, *85*(4), 2042–2096.

Caplice, C. (2007). Electronic markets for truckload transportation. *Production and Operations Management*, *16*(4), 423–436.

Collin-Dufresne, P., Hoffmann, P., & Vogel, S. (2019). *Informed traders and dealers in the fx forward market* (Working Paper).

Dugast, J., Üslü, S., & Weill, P.-O. (2019). *A theory of participation in otc and centralized markets* (NBER Working Paper).

Elliott, M. (2014). *Heterogeneities and the fragility of labor markets* (Working Paper).

Frechette, G. R., Lizzeri, A., & Salz, T. (2019). Frictions in a competitive, regulated market: Evidence from taxis. *American Economic Review*, *109*(8), 2954–92.

Galenianos, M., & Gavazza, A. (2017). A structural model of the retail market for illicit drugs. *American Economic Review*, *107*(3), 858–896.

Gavazza, A. (2010). Asset liquidity and financial contracts: Evidence from aircraft leases. *Journal of Financial Economics*, *95*(1), 62–84.

Gavazza, A. (2011). Leasing and secondary markets: Theory and evidence from commercial aircraft. *Journal of Political Economy*, *119*(2), 325–377.

Gayle, G.-L., & Miller, R. A. (2015). Identifying and testing models of managerial compensation. *The Review of Economic Studies*, *82*(3), 1074–1118.

Gil, R., Kim, M., & Zanarone, G. (2021). Relationships under stress: Relational outsourcing in the US airline industry after the 2008 financial crisis. *Management Science*.

Guerre, E., Perrigne, I., & Vuong, Q. (2000). Optimal nonparametric estimation of first-price auctions. *Econometrica*, *68*(3), 525–574.

Harris, A., & Nguyen, T. M. A. (2021). *Long-term relationships in the US truckload freight industry* (Working Paper).

Hu, Y., & Xin, Y. (2021). *Identification and estimation of dynamic structural models with unobserved choices* (Working Paper).

Hubbard, T. N. (2001). Contractual form and market thickness in trucking. *RAND Journal of Economics*, 369–386.

Kasahara, H., & Shimotsu, K. (2009). Nonparametric identification of finite mixture models of dynamic discrete choices. *Econometrica*, *77*(1), 135–175.

Kranton, R. E. (1996). Reciprocal exchange: A Self-Sustaining system. *American Economic Review*, *86*(4), 830–851.

Lagos, R. (2000). An alternative approach to search frictions. *Journal of Political Economy*, *108*(5), 851–873.

Lagos, R. (2003). An analysis of the market for taxicab rides in new york city. *International Economic Review*, *44*(2), 423–434.

Lee, T., & Wang, C. (2018). *Why trade over-the-counter? when investors want price discrimination* (Jacobs Levy Equity Management Center for Quantitative Financial Research Paper).

Macaulay, S. (1963). Non-contractual relations in business: A preliminary study. *The Sociology of Economic Life* (pp. 361–377). Springer.

Macchiavello, R. (2022). Relational contracts and development. *Annual Review of Economics*, *14*, 337–362.

Macchiavello, R., & Morjaria, A. (2015). The value of relationships: Evidence from a supply shock to kenyan rose exports. *American Economic Review*, *105*(9), 2911–45.

Macchiavello, R., & Morjaria, A. (2021). Competition and relational contracts in the Rwanda coffee chain. *The Quarterly Journal of Economics*, *136*(2), 1089–1143.

Marcus, A. A. (1987). From market dominance to credible commitments: Shippers' strategies in a deregulated trucking environment. *Transportation journal*, 4–15.

Masten, S. E. (2009). Long-term contracts and short-term commitment: Price determination for heterogeneous freight transactions. *American Law and Economics Review*, *11*(1), 79–111.

Ngai, L. R., & Tenreyro, S. (2014). Hot and cold seasons in the housing market. *American Economic Review*, *104*(12), 3991–4026.

Nickerson, J. A., & Silverman, B. S. (2003). Why aren't all truck drivers owner-operators? Asset ownership and the employment relation in interstate for-hire trucking. *Journal of Economics & Management Strategy*, *12*(1), 91–118.

Ostria, S. J. (2003). *Evaluation of US commercial motor carrier industry challenges and opportunities* (Technical Report).

Pagano, M. (1989). Trading volume and asset liquidity. *The Quarterly Journal of Economics*, *104*(2), 255–274.

Perrigne, I., & Vuong, Q. (2011). Nonparametric identification of a contract model with adverse selection and moral hazard. *Econometrica*, *79*(5), 1499–1539.

Rosaia, N. (2020). *Competing platforms and transport equilibrium: Evidence from New York City* (Working Paper). Harvard University.

Rose, N. L. (1985). The incidence of regulatory rents in the motor carrier industry. *The RAND Journal of Economics*, 299–318.

Rose, N. L. (1987). Labor rent sharing and regulation: Evidence from the trucking industry. *Journal of Political Economy*, *95*(6), 1146–1178.

Rust, J. (1994). Structural estimation of Markov decision processes. *Handbook of econometrics*, *4*, 3081–3143.

Startz, M. (2021). *The value of face-to-face: Search and contracting problems in nigerian trade* (PEDL Research Papers).

Train, K. E. (2008). EM algorithms for nonparametric estimation of mixing distributions. *Journal of Choice Modelling*, *1*(1), 40–69.

Tunca, T. I., & Zenios, S. A. (2006). Supply auctions and relational contracts for procurement. *Manufacturing & Service Operations Management*, *8*(1), 43–67.

Williams, N., & Murray, D. (2020). *An analysis of the operational costs of trucking: 2020 update* (Technical Report). American Transportation Research Institute.

Yang, R. (2021). *(Don't) take me home: Home bias and the effect of self-driving trucks on interstate trade* (Working Paper).

Ying, J. S. (1990). The inefficiency of regulating a competitive industry: Productivity gains in trucking following reform. *The Review of Economics and Statistics*, 191–201.

Zahur, N. B. (2022). *Long-term contracts and efficiency in the liquefied natural gas industry* (Working Paper).

A Data Construction

This section describes the construction of key variables of long-term relationships: per-mile contract rates, primary status, demotion events, and auction events. In addition to the observed routing guide for each load offer, we exploit a complementary data set that records the timestamps of shippers' input into the TMS. These timestamps provide the candidates for demotion and auction events. We refer to the period between two consecutive timestamps of a shipper on a lane as a "date-range".

Contract rates. Shippers seeking long-term relationships define lanes at geographical levels finer than KMA to KMA, sometimes as fine as warehouse-to-warehouse. Shippers can also bundle origin-destination pairs with close proximity as a lane, using the same contract. This means that if a contract specifies a linehaul rate (total payment for a trip) on such a lane, the carrier's per-mile payment would vary with the distance of specific trips. On the other hand, if a contract specifies a per-mile rate, the carrier's total payment would vary with trips' distance. To match the unit of spot rates, we construct per-mile contract rates for specific trips and take the median of these rates within a date-range as the fixed contract rate.

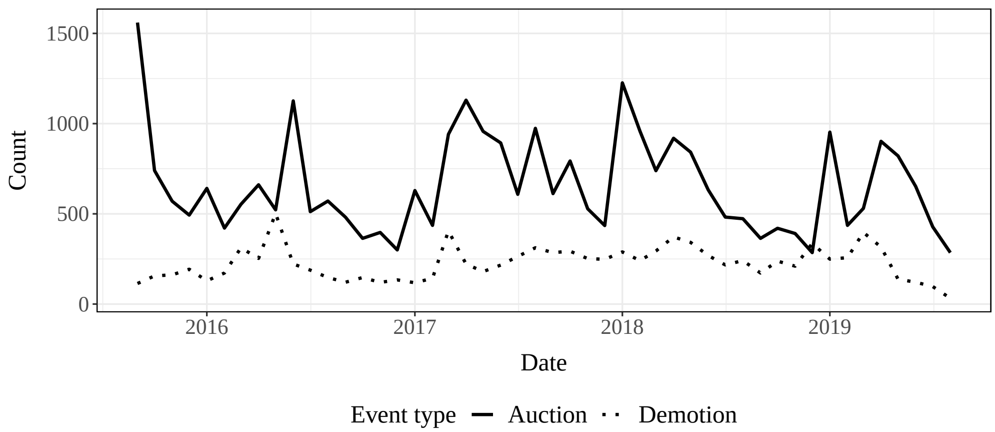
Primary status. We infer the status of carriers from the fact that primary carriers are generally the first to receive shippers' offers. Exceptions are typically due to prespecified capacity constraints that both the shipper and the carrier agreed on, or multiple primary carriers sharing the same lane. In such instances, we assign primary status to the carrier with the most offers within a date-range.

Auction events. Our data do not include records of auctions. However, we can observe when contract rates change. If we observe at least three changes in contract rates within a date-range from the previous date-range, we assign an auction event to the beginning of the current date-range. The secondary indicator of auction events is when a completely new carrier replaces the previous primary carrier. There is a tradeoff in using this indicator. On the one hand, not using this indicator risks missing some auction events because carriers sometimes reuse their bids. On the other hand, using this indicator risks assigning an auction event to what is actually a demotion event, since some backup carriers may not appear in the routing guide. We perceive the second risk to be smaller and use both indicators to detect auction events.

Demotion events. A demotion event is an instance where the current primary carrier is replaced by a different carrier within the same contract period (that is, between two auction events). Measurement errors in our constructed indicator of demotion events can come from measurement errors in our constructed primary status or indicators of auction events.

Figure 15 plots the number of identified auction and demotion events in each month-year in our sample period. Auctions appear to occur at random over time. Additionally, there are some spikes in the number of identified auction events, reflecting the fact that shippers tend to hold auctions on multiple lanes simultaneously. Identified demotion events are relatively evenly distributed over time and do not show a correlation with identified auction events.

Figure 15: Number of identified auction and demotion events

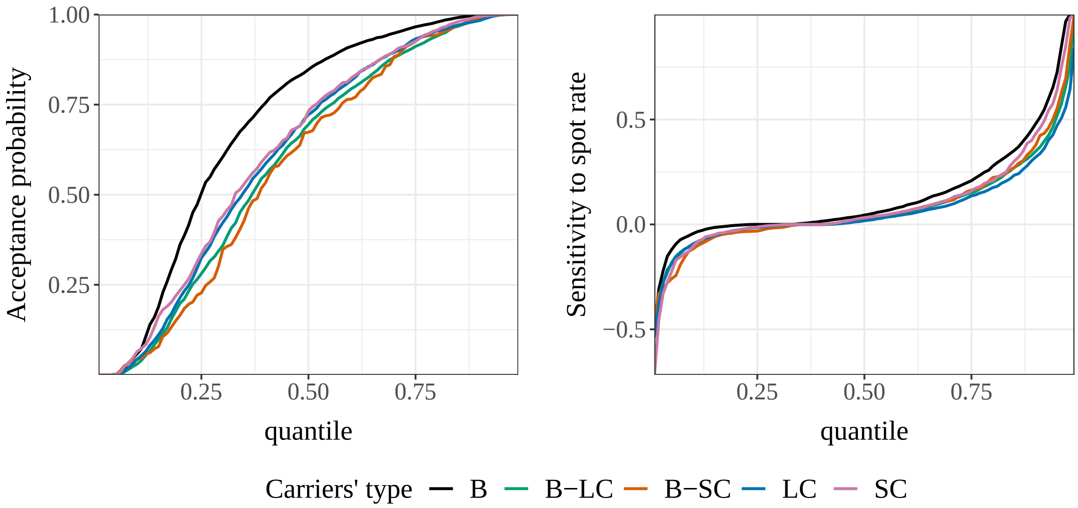


This suggests that our data construction does a reasonable job at separating the two types of event.

Carrier's types. There are two public identifying code systems for carriers: the Standard Carrier Alpha Code (SCAC), maintained by the National Motor Freight Traffic Association (NMFTA) and US DOT for carrier registration at the Department of Transportation. We map the SCAC variable in our data set to US DOT codes using a conversion table from the NMFTA. We then map US DOT codes to carriers' registration at the Department of Transportation for the year 2020. This method matches 90% of carriers in our data set to five types: brokers (B), small asset-onwers (SC), large asset-owners (LC), brokers/small asset-owners (B-SC), brokers/large asset-owners (B-LC). The latter two groups are for carriers with multiple divisions.

Figure 16 plots the acceptance probabilities and sensitivities to spot rates by carrier type. It shows that brokers (solid black) are more likely to accept loads, but also more sensitive to spot rates. The first pattern is likely due to brokers' having more flexibility, with a large pool of carriers to draw from in the spot market. The second pattern is likely due to brokers' costs and thus profit margins on contracted loads being directly tied to spot rates. Moreover, we observe that carriers that are both brokers and asset owners behave similarly to asset owners. In this paper, we drop carriers that are identified as brokers (B) from our data set.

Figure 16: Acceptance tendency across carriers' types



Notes: We run Logit regression $\Pr(d_{ij\ell t} = \text{accepted}) = \beta_{ij\ell,0} + \beta_{ij\ell,1}(\tilde{p}_{\ell t} - p_{ij\ell})$ where i denotes the shipper, j denotes the primary carrier, ℓ denote the lane and t denotes the period in their relationship. The left panel plots the quantile distribution of predicted acceptance probabilities across all relationships by carrier type at $\tilde{p}_{\ell t} = p_{ij\ell}$. The right panel plots the quantile distribution of the decrease in predicted accepted probabilities across all relationships by carrier type when $\tilde{p}_{\ell t}$ increases by one standard deviation from $p_{ij\ell}$.

B Omitted proofs

B.1 Properties of full compensation and acceptance schedules

Our identification argument relies on the observation that under Assumption 2 and Assumption 3, acceptance schedules have well-defined, distinct "jump" points. This section proves the key properties of the full compensation that give rise to these "jump" points.

When proving the properties of a single relaitonship, we drop the dependence of notation on the carrier's rent and subscript i, j, ℓ for shipper, carrier and lane for ease of notation. Recall that the full compensation includes the carrier's rent inclusive of savings on search cost and a dynamic incentive component

$$\bar{p}(R_{t-1}, \tilde{p}_t) = \underbrace{\eta + p + \kappa}_{\text{carrier's transformed rent}} + \underbrace{\frac{\delta}{1-\delta}[V(\alpha R_{t-1}), \tilde{p}_t) - V(\alpha R_{t-1} + (1-\alpha), \tilde{p}_t)]}_{\text{dynamic incentive}},$$

and transformed cost $\tilde{c}_t = c_t + \kappa$ has distribution Normal$(\tilde{\mu}^c, \sigma^c)$, denoted by $\tilde{F}$. The

Bellman's equation of the carrier gives continuation value

$$\begin{aligned}V(R_{t-1},\tilde{p}_{t-1}) =&\mathbf{E}_{\tilde{p}_t}[(1-\sigma_0(R_{t-1},\tilde{p}_t))\underline{V}(\tilde{p}_t)|\tilde{p}_{t-1}] \\ &+\mathbf{E}_{\tilde{p}_t}[\sigma_0(R_{t-1},\tilde{p}_t)\left\{(1-\delta)h(\bar{p}(R_{t-1},\tilde{p}_t),\tilde{p}_t)+\delta V(\alpha R_{t-1}+(1-\alpha),\tilde{p}_t)\right\}|\tilde{p}_{t-1}],\end{aligned} \tag{15}$$

where

$$\begin{aligned}h(\bar{p}(R_{t-1},\tilde{p}_t),\tilde{p}_t) =&\mathbf{1}\{\tilde{p}_t\leq\bar{p}(R_{t-1},\tilde{p}_t)\}\tilde{F}(\bar{p}(R_{t-1},\tilde{p}_t))(\bar{p}(R_{t-1},\tilde{p}_t)-\mathbf{E}[\tilde{c}_t|\tilde{c}_t\leq\bar{p}(R_{t-1},\tilde{p}_t)]) \\ &+\mathbf{1}\{\tilde{p}_t>\bar{p}(R_{t-1},\tilde{p}_t)\}\tilde{F}(\tilde{p}_t)(\tilde{p}_t-\mathbf{E}[\tilde{c}_t|\tilde{c}_t\leq\tilde{p}_t])\end{aligned} \tag{16}$$

is an expression capturing the carrier's current payoff and the gain in continuation value from an acceptance in the current period.

We use a series of lemmas to show that the full acceptance schedules of relationships with different carriers' rents have well-defined, distinct "jump" points.

Lemma B.1. *Suppose that Assumption 3 holds. For every level of carrier's transformed rent $\eta+p+\kappa\geq 0$ and rejection state R_{t-1}, there exists $\hat{p}_{low},\hat{p}_{high}\in\mathbb{R}$ such that $\bar{p}(R_{t-1},\hat{p}_{low})\geq\hat{p}_{low}$ and $\bar{p}(R_{t-1},\tilde{p}_t)<\tilde{p}_t$ for all $\tilde{p}_t>\hat{p}_{high}$.*

Proof. Under Assumption 3 (iii), $\bar{p}(R_{t-1},\hat{p}_{\text{low}})\geq\hat{p}_{\text{low}}$ for all $\hat{p}_{\text{low}}\leq\eta+p+\kappa$. To show the existence of $\hat{p}_{\text{high}}$, it suffices to show that $\bar{p}(R_{t-1},\tilde{p}_t)$ is bounded above. Note that being in a relationship gives the carrier an additional option to accept a load and get a payoff of $\eta+p+\kappa-\tilde{c}_t$, while not being in a relationship only gives the carrier the option to accept a spot load, which gives a payoff of $\tilde{p}_t-\tilde{c}_t$, or to remain idle and get zero. Thus, the continuation value of the carrier at any state is bounded above by the continuation value were the carrier to never be demoted, and bounded below by the continuation value were the carrier to be demoted immediately. It follows that

$$\begin{aligned}&V(\alpha R_{t-1},A),\tilde{p}_t)-V(\alpha R_{t-1}+(1-\alpha),\tilde{p}_t) \\ \leq&(1-\delta)\sum_{\tau=1}^{\infty}\delta^\tau\left(\mathbf{E}[\max\{\eta+p+\kappa-\tilde{c}_{t+\tau},\tilde{p}_{t+\tau}-\tilde{c}_{t+\tau},0\}|\tilde{p}_t]-\mathbf{E}[\max\{\tilde{p}_{t+\tau}-\tilde{c}_{t+\tau},0\}|\tilde{p}_t]\right) \\ \leq&(1-\delta)\sum_{\tau=1}^{\infty}\delta^\tau\left(\mathbf{E}[\max\{\eta+p+\kappa,\tilde{p}_{t+\tau},\tilde{c}_{t+\tau}\}-\max\{\tilde{p}_{t+\tau},\tilde{c}_{t+\tau}\}|\tilde{p}_t]\right)\leq\eta+p+\kappa.\end{aligned}$$

Thus, the full compensation schedule at $(R_{t-1},\tilde{p}_t)$ is bounded above, $\bar{p}(R_{t-1},\tilde{p}_t)\leq\frac{\eta+p+\kappa}{1-\delta}$, completing the proof of the lemma. □

Definition 4. (Jump points) Fix search cost κ. For each level of carrier's rent $\eta+p$ and rejection state R_{t-1}, define the jump point as the lowest spot rate above which the full compensation schedule is always lower than spot rate,

$$p^*(R_{t-1}|\eta + p) = \inf\{\hat{p} : \bar{p}(R_{t-1}, \tilde{p}_t) < \tilde{p}_t, \forall \tilde{p}_t > \hat{p}\}.$$

By Lemma B.1, p^* is well defined. Moreover, under Assumption 2 on the cost distribution and Assumption 3 on the continuity of the full acceptance schedule in spot rates, acceptance probability is positive in the left neighborhood of $p^*(R_{t-1}|\eta+p)$ and zero to the right of this point. This is why we refer to this point as a jump point. Note that we do not rule out the possibility that the full acceptance schedule equals spot rate at multiple points. We focus on the highest such point, since it has the special property that acceptance probability remains zero for any higher level of spot rate. The next lemma shows that these jump points are ordered by the level of carrier rent.

Lemma B.2. *(Order of jump points) Suppose that Assumption 2 and Assumption 3 hold. Fix search cost κ and cost distribution F, any level of rejection state R_{t-1}, and carriers' rents $\eta + p > \eta' + p' \geq 0$. Then,*

$$p^*(R_{t-1}|\eta + p) > p^*(R_{t-1}|\eta' + p').$$

Proof. Suppose that $p^*(R_{t-1}|\eta + p) \leq p^*(R_{t-1}|\eta' + p')$, then at $\tilde{p}_t = p^*(R_{t-1}|\eta' + p')$,

$$\bar{p}(R_{t-1}, \tilde{p}_t|\eta' + p') = \tilde{p}_t \geq \bar{p}(R_{t-1}, \tilde{p}_t|\eta + p),$$

where the last equality follows from the definition of p^* and that $p^*(R_{t-1}|\eta + p) \leq \tilde{p}_t$. This yields a contradiction because under monotonicity of the full compensation schedule (Assumption 3), we have $\bar{p}(R_{t-1}, \tilde{p}_t|\eta + p) > \bar{p}(R_{t-1}, \tilde{p}_t|\eta' + p')$. □

B.2 Identification of the distribution of carriers' rents and costs

This section proves the identification of the distribution $\tilde{F}$ of transformed costs and the distribution $[G^{\eta+p}]^{1:N}$ of winning carriers' rents on a given lane. For ease of notation, we drop the dependence of notation on ℓ.

Lemma B.3. *(Identification of the distribution of transformed costs) Suppose that Assumption 2 holds. If there exist two distinct jump points $p^*(R_{t-1}|\eta + p)$ and $p^*(R'_{t-1}|\eta' + p')$ observed in either two different relationships or two different rejection states of a carrier, then $(\tilde{\mu}^c, \sigma^c)$ are identified.*

Proof. Note that each jump point p^* gives us a point $(p^*, \tilde{F}(p^*))$ on the distribution $\tilde{F}$, where $\tilde{F}(p^*)$ is the observed acceptance probability at this jump point. Thus, two distinct jump points give a system of linear equations

$$\frac{p^*(R_{t-1}|\eta+p)-\tilde{\mu}^c}{\sigma^c} = \Phi^{-1}\left(\Pr(d_t = \text{accept}|R_{t-1}, \tilde{p}_t = p^*(R_{t-1}|\eta+p); \eta+p)\right)$$
$$\frac{p^*(R'_{t-1}|\eta'+p')-\tilde{\mu}^c}{\sigma^c} = \Phi^{-1}\left(\Pr(d_t = \text{accept}|R'_{t-1}, \tilde{p}_t = p^*(R'_{t-1}|\eta'+p'); \eta'+p')\right),$$

the right hand side of which are observed. This system pins down $(\tilde{\mu}^c, \sigma^c)$. □

Lemma B.4. *(Identification of carriers' rents in long relationships) Suppose that Assumption 2 and conditions (ii) and (iii) of Assumption 3 hold. In addition, suppose that search cost κ and the distribution F of operational costs are identified. If a relationship has duration $T \to \infty$, the rent level $\eta + p$ of the carrier in this relationship is identified.*

Proof. Under Assumption 2 and that $T \to \infty$, the acceptance schedule is fully observed. Note that the identification of the carrier rent immediately follows from the monotonicity of the full compensation schedule in carrier rent (Assumption 3 (ii)). Here we present a direct proof that does not rely on monotonicty.

That the acceptance schedule is fully observed means that at any state $(R_{t-1}, \tilde{p}_t)$ in which $\bar{p}(R_{t-1}, \tilde{p}_t) \geq \tilde{p}_t$, the value of the full compensation at that state is identified by

$$\bar{p}(R_{t-1}, \tilde{p}_t) = \tilde{F}^{-1}(\Pr(d_t = \text{accept}|R_{t-1}, \tilde{p}_t; \eta+p)).$$

It follows that at any state $(R_{t-1}, \tilde{p}_t)$, $h(\bar{p}(R_{t-1}, \tilde{p}_t), \tilde{p}_t)$ in Equation (16) is identified. Thus, we can define a mapping $\Gamma : \mathcal{V} \to \mathcal{V}$, where each element $(V(R_{t-1}, \tilde{p}_t))_{R_{t-1}, \tilde{p}_t}$ of $\mathcal{V}$ satisfies that at each state $(R_{t-1}, \tilde{p}_t)$,

$$\mathbf{E}[\underline{V}(\tilde{p}_t)|\tilde{p}_{t-1}] \leq V(R_{t-1}, \tilde{p}_{t-1}) \leq \frac{1}{1-\delta} p^*(R_{t-1}).$$

Under Assumption 3, this means that $\mathcal{V}$ is bounded and that it contains the solution to the Bellman equation. It is straightforward to show that Γ is a contraction mapping by verifying that it satisfies Blackwell's sufficient conditions. Thus, it has a unique fixed point, which is also the collection of continuation values in this relationship. We can find this fixed point by

$$(V(R_{t-1}, \tilde{p}_{t-1}))_{R_{t-1}, \tilde{p}_{t-1}} = \lim_{k\to\infty} \Gamma^k\left((\mathbf{E}[\underline{V}(\tilde{p}_t)|\tilde{p}_{t-1}])_{R_{t-1}, \tilde{p}_{t-1}}\right).$$

That is, Equation (15), the observed acceptance schedule and F pin down V. Finally, it holds at the jump point that

$$p^*(R_{t-1}) = \eta + p + \kappa + \frac{\delta}{1-\delta}[V(\alpha R_{t-1}), p^*(R_{t-1})) - V(\alpha R_{t-1} + (1-\alpha), p^*(R_{t-1}))].$$

This pins down carrier rent $\eta + p$. □

While the identification proof for long relationships helps demonstrate the source of identification power in our setting, we need to develop an argument for the identification of the mixture of carrier rent that includes relationships of all lengths. This argument relies on jump points being strictly monotone in carrier rent. We reproduce the statement and provide the proof below.

Lemma 4. *(Identification of the distribution of carrier rent) Suppose that Assumption 2 and Assumption 3 hold. In addition, suppose that the shipper's incentive scheme σ_s, search cost κ, and the distribution F of operational costs are identified. If the distribution $[G^{\eta+p}]^{1:N}$ of the rents of winning carriers permits an absolutely continuous density, then it is nonparametrically identified.*

Proof of Lemma 4. Fix a rejection state R_{t-1}. We exploit the following equality

$$\Pr(d_t = \text{accept}, R_{t-1}, \tilde{p}_t) = \int \Pr(d_t = \text{accept}, R_{t-1}, \tilde{p}_t | \eta + p = r) d[G^{\eta+p}]^{1:N}(r),$$

where the joint distribution of carriers' acceptance, rejection states and spot rates, both unconditional and conditional on carriers' rents, are either directly observed or identified. Our task is to identify the mixture $[G^{\eta+p}]^{1:N}$.

Another key property is that beyond the "jump" points, acceptance probability equals zero,

$$\Pr(d_t = \text{accept}, R_{t-1}, \tilde{p}_t | \eta + p) = 0, \quad \text{for all } \tilde{p}_t > p^*(R_{t-1} | \eta + p).$$

Take any two distributions $[G^{\eta+p}]^{1:N}$ and $[\hat{G}^{\eta+p}]^{1:N}$ with absolutely continuous densities $[g^{\eta+p}]^{1:N}$ and $[\hat{g}^{\eta+p}]^{1:N}$ that are not everywhere the same. Let $\bar{r} = \inf\{r' : [g^{\eta+p}]^{1:N}(r) = [g^{\eta+p}]^{1:N}(r), \forall r > r'\}$ and suppose, without loss of generality, that $[g^{\eta+p}]^{1:N}(\bar{r}^-) > [\hat{g}^{\eta+p}]^{1:N}(\bar{r}^-)$. The continuity of $[g^{\eta+p}]^{1:N}$ and $[\hat{g}^{\eta+p}]^{1:N}$ further implies that for some $\epsilon > 0$, $[g^{\eta+p}]^{1:N}(\bar{r}) > [\hat{g}^{\eta+p}]^{1:N}(\bar{r})$ for all $r \in [\bar{r} - \epsilon, \bar{r}]$. Then, it follows from Lemma B.2 that

$$\int_{p^*(R_{t-1}|\eta+p=\bar{r}-\epsilon)}^{\infty} \Pr(d_t = \text{accept}, R_{t-1}, \tilde{p}_t > p^*(R_{t-1} | \eta + p = \bar{r} - \epsilon) | \eta + p = r) d[G^{\eta+p}]^{1:N}(r)$$
$$> \int_{p^*(R_{t-1}|\eta+p=\bar{r}-\epsilon)}^{\infty} \Pr(d_t = \text{accept}, R_{t-1}, \tilde{p}_t > p^*(R_{t-1} | \eta + p = \bar{r} - \epsilon) | \eta + p = r) d[\hat{G}^{\eta+p}]^{1:N}(r).$$

That is, two distributions that differ generate different acceptance probability on some range of spot rates. This completes the proof that $[G^{\eta+p}]^{1:N}$ is nonparametrically identified. $\square$

B.3 Existence of a symmetric monotone equilibrium

This section constructs a symmetric monotone equilibrium in two steps. First, we construct a monotone equilibrium in a pseudo-game in which only match quality matters. Second, we derive a symmetric monotone equilibrium in the original game from the monotone equilibrium of the pseudo-game.

Assumption 4. *There exists $\underline{b} \in \mathbb{R}$ such that for all $b \geq \underline{b}$, $U(R_0, \tilde{p}_0|r, b) \geq \underline{U}(R_0, \tilde{p}_0)$ for all $r \geq 0$ and $U(R_0, \tilde{p}_0|r, b)$ is increasing in $r \geq 0$ and $b \geq \underline{b}$.*

The intuition for this assumption is that if the shipper's rent is sufficiently high, then fixing her rent, the shipper benefits from the carrier having higher rent and thus accepting more frequently. While intuitive, this statement relies on the specifics of how the carrier's rent affects its' path of play and how such path of play is correlated with the realized path of spot rates. The right panel of Figure 22 demonstrates that this assumption is satisfied under our estimated spot process and incentive scheme.

Proposition B.1. *Under Assumptions 2, 3 and 4, there exists a symmetric monotone equilibrium.*

Proof. We construct a symmetric monotone equilibrium in two steps.

Step 1: A monotone equilibrium of a pseudo-game.

Consider a bidding game where each carrier j has private information about their match-quality with the shipper, θ_{ij}. Each carrier submits a bid b_{ij} and the shipper chooses the carrier with the highest bid subject to reserve price $\underline{b}$. Here, $\underline{b}$ is the lowest level of shipper's rent that satisfies Assumption 4. The carrier that wins this auction gets expected payoff $V(R_0, \tilde{p}_0|\theta_{ij} - b_{ij})$.

In this game, there exists a strictly increasing bidding function $\mathbf{b} : \theta_{ij} \mapsto b_{ij}$ such that

$$\mathbf{b}(\theta_{ij}) = \arg\max_{b} [G^{\theta}(\mathbf{b}^{-1}(b))]^{N-1}(V(R_0, \tilde{p}_0|\theta_{ij} - b) - \mathbf{E}[\underline{V}(\tilde{p}_1)|\tilde{p}_0]).$$

Note that a relationship strictly benefits the carrier if and only if the carrier's rent is strictly positive. Thus, in this equilibrium, the lowest match quality of a winning carrier gives zero rent to that carrier, $\mathbf{b}(\underline{\theta}) = \underline{\theta}$. That is, individual rationality binds for the carrier with the lowest match quality. Moreover, a carrier with match quality $\theta > \underline{\theta}$ has a strictly positive rent, since it would otherwise strictly benefit from deviating to a lower bid. Denote by $\mathbf{r} : \theta_{ij} \mapsto \theta_{ij} - \mathbf{b}(\theta_{ij})$ the function that maps the carrier's match quality to its rent. We have

$r(\underline{\theta}) = 0$, and for all $\theta \geq \underline{\theta}$, $r(\theta) > 0$ and $\mathbf{b}'(\theta) + \mathbf{r}'(\theta) = 1$. We want to show that $\mathbf{r}$ is strictly increasing.

The first-order condition of the carrier's bidding satisfies that for all $\theta > \underline{\theta}$,

$$(N-1)\frac{g^\theta(\theta)}{G^\theta(\theta)} = \frac{\frac{\partial}{\partial r}V(R_0, \tilde{p}_0 | r = \mathbf{r}(\theta))}{V(R_0, \tilde{p}_0 | r = \mathbf{r}(\theta)) - \mathbf{E}[\underline{V}(\tilde{p}_1)|\tilde{p}_0]}\mathbf{b}'(\theta)$$

Suppose that for some $\theta \geq \underline{\theta}$, $\mathbf{r}'(\theta) \leq 0$ and consider two cases: (i) there exists a strict interval on which $\mathbf{r}'(\theta) = 0$, and (ii) there is no such interval. In case (i), there exist $\theta_1 < \theta_2$ such that $\mathbf{r}(\theta_1) = \mathbf{r}(\theta_2)$ and $\mathbf{r}'(\theta_1) = \mathbf{r}'(\theta_2)$. In case (ii), there exist $\theta_1 < \theta_2$ such that $\mathbf{r}(\theta_1) = \mathbf{r}(\theta_2)$ and $\mathbf{r}'(\theta_1) > 0 > \mathbf{r}'(\theta_2)$. In either case, we have $0 < \mathbf{b}'(\theta_1) \leq \mathbf{b}'(\theta_2)$. Then under the assumption that G^θ has strictly decreasing hazard rate, we have

$$\begin{aligned}\frac{\frac{\partial}{\partial r}V(R_0, \tilde{p}_0 | r = \mathbf{r}(\theta_1))}{V(R_0, \tilde{p}_0 | r = \mathbf{r}(\theta_1)) - \mathbf{E}[\underline{V}(\tilde{p}_1)|\tilde{p}_0]} &= \frac{g^\theta(\theta_1)}{G^\theta(\theta_1)}[\mathbf{b}'(\theta_1)]^{-1} \\ &> \frac{g^\theta(\theta_2)}{G^\theta(\theta_2)}[\mathbf{b}'(\theta_2)]^{-1} = \frac{\frac{\partial}{\partial r}V(R_0, \tilde{p}_0 | r = \mathbf{r}(\theta_2))}{V(R_0, \tilde{p}_0 | r = \mathbf{r}(\theta_2)) - \mathbf{E}[\underline{V}(\tilde{p}_1)|\tilde{p}_0]}.\end{aligned}$$

This is a contradiction, completing the proof that $\mathbf{r}(\theta)$ is strictly increasing in θ.

Step 2: Symmetric monotone equilibrium.

We now map the monotone equilibrium of the pseudo-game to a symmetric monotone equilibrium of the original bidding game. Note that for a carrier j, if the shipper chooses the carrier with the highest effective bid (or proposed shipper's rent) and other carries bid according to $\mathbf{b}$, then carrier j has no incentive to deviate from bidding according to $\mathbf{b}$. It remains to show the shipper's selection rule in the pseudo-game is optimal in the original bidding game.

Under Assumption 4 and by the choice of $\underline{b}$ in the pseudo-game, we have for all $\theta \geq \underline{\theta}$,

$$U(R_0, \tilde{p}_0 | \mathbf{b}(\theta), \mathbf{r}(\theta)) \geq \mathbf{E}[\underline{U}(\tilde{p}_1)|\tilde{p}_0]$$

and $U(R_0, \tilde{p}_0 | \mathbf{b}(\theta), \mathbf{r}(\theta))$ is increasing in θ. This means that by choosing the carrier j with the highest bid such that $b_{ij} \geq \mathbf{b}(\underline{\theta})$, the shipper maximizes her expected payoff from the relationship and never receives an expected payoff lower than her outside option of always going to the spot market. □

C Estimation details

C.1 Estimation of model primitives

Figure 17 presents a roadmap of our estimation procedure.

C.1.1 Estimate the number of bidders

Assume that the number of bidders in an auction is stochastic, $N_a \sim \text{Binomial}(N, q)$. Then the number of effective bidders, who pass the shipper's individual rationality constraint and become either primary or backup carriers, is $n_a \sim \text{Binomial}(N, \tilde{q})$, where $\tilde{q} = q(1 - G^{\eta+p}(\underline{r}))$.

The empirical challenge in estimating $(N, \tilde{q})$ is that we only observe the number of carriers that receive at least an offer within the auction period. This number, denoted by $\hat{n}_a$, could be smaller than the number of effective bidders n_a, since low-rank carriers may never receive an offer. We tackle this issue in two steps. First, we take the maximum of $\hat{n}_a$ in all auctions as an estimate of N. Second, we estimate $\tilde{q}$ through a calibration exercise that captures the bias in the number of observed carriers. This exercise simulates a distribution of $\hat{n}_a$ from $(N, \tilde{q})$, the total number of offers within each auction, and the estimated probability that a load is rejected conditional on previous rejections and the current spot rate. Matching the mean and variance of this simulated distribution to its empirical counterpart pins down $\tilde{q}$.

C.1.2 The likelihood contribution of each relationship

For each relationship, we observe the duration of the relationship $T_{ij\ell}$, and for each period, whether the carrier accepts, the rejection index at the beginning of the period $R_{j\ell t-1}$, and the mean spot rate in that period $\tilde{p}_{\ell t}$. We also observe the standard deviation of spot rates, σ_ℓ^ζ. The likelihood contribution of this relationship depends on the parameters of the carrier's transformed costs $(\tilde{\mu}_{ia\ell}^c, \sigma^c)$ and transformed rent $\eta_{ij\ell} + p_{ij\ell} + \kappa_\ell$ as follows

$$\ln \mathcal{L}\left((\mathbf{1}\{d_{j\ell t} = \text{accept}\}, R_{j\ell t-1}, \tilde{p}_{\ell t})_{t=1}^{T_{ij\ell}}; \sigma_\ell^\zeta, \tilde{\mu}_{ia\ell}^c, \sigma^c, \eta_{ij\ell} + p_{ij\ell} + \kappa_\ell\right)$$
$$\propto \prod_{t=1}^{T_{ij\ell}} \prod_{D \in \{\{\text{accept}\},\{\text{idle,spot}\}\}} \Pr(d_{j\ell t} \in D | R_{j\ell t-1}, \tilde{p}_{\ell t}; \sigma_\ell^\zeta, \tilde{\mu}_{ia\ell}^c, \sigma^c, \eta_{ij\ell} + p_{ij\ell} + \kappa_\ell)^{\mathbf{1}\{d_{j\ell t} \in D\}},$$

Figure 17: Estimation roadmap

Step 1. Estimate instrumental objects

− Calibrate the discount factor δ.

− Calibrate the number of effective bidders to be distributed as Binomial$(N, \tilde{q})$.

− Estimate the spot process by OLS and shippers' strategies by MLE.

Step 2. Estimate cost parameters

− Estimate $\{\sigma^c, (\tilde{\mu}^c_{ia\ell})_{ia\ell}, (\eta_{ij\ell} + p_{ij\ell} + \kappa_\ell)_{ij\ell}\}_{T_{ij\ell} \geq 50}$ by MLE.

Outer: $\sigma^c \in \{0.1, 0.2, ..., 2.0\}$.

Inner: For each shipper-lane-auction, estimate the transformed cost shifter $\tilde{\mu}^c_{ia\ell}$ and the transformed rents of all primary carries, $(\eta_{ij\ell} + p_{ij\ell} + \kappa_\ell)_j$, including both the auction winner and promoted carriers.

− Estimate the scale efficiency parameter γ_1 by 2SLS in

$$\tilde{\mu}^c_{ia\ell} = \frac{\gamma_1 \ln(\text{Volume}^{\text{spot}}_\ell)}{\text{Distance}_\ell} + \gamma_2 \mathbf{1}^a_{\text{tight}} + h_3(\text{Distance}_\ell) + \nu^c_\ell + \epsilon^c_{ia\ell}, \tag{17}$$

using the predicted trade flows (Caliendo, Parro, Rossi-Hansberg, & Sarte, 2018) to instrument for $\text{Volume}^{\text{spot}}_\ell$. Here h_3 is a polynomial of degree 3.

− Extrapolate transformed costs $\tilde{\mu}^c_{ia\ell}$ from relationships with $T_{ij\ell} \geq 50$ to all relationships,

$$\widehat{\tilde{\mu}}^c_{ia\ell} = h'_2(\text{Rate}^{\text{spot}}_\ell, \text{Volume}^{\text{spot}}_\ell, \text{Distance}_\ell) + \gamma'_2 \mathbf{1}^a_{\text{tight}} + h'_3(\text{Distance}_\ell).$$

Here h'_2 and h'_3 and polynomials of degree 2 and degree 3 respectively.
− Decompose each $\widehat{\tilde{\mu}}^c_{ia\ell}$ into a cost shifter $\widehat{\mu}^c_{ia\ell}$ and a search cost $\widehat{\kappa}_\ell$ by normalizing the median operational cost on a lane to \$1.22/mile (\$1.55/mile net of \$0.33/mile fuel surcharge).

Step 3. Estimate the distribution of rents and match-specific gains

− Cluster all relationships:

Outer: 10 K-means clusters based on lane characteristics $(\text{Distance}_\ell, \text{Rate}^{\text{spot}}_\ell, \text{Volume}^{\text{spot}}_\ell)$.

Inner: 2 market phases (soft, tight) based on the start date of each relationship.

− For each sub-cluster:

(i) Estimate the distribution of winning carriers' rents $[G^{\eta+p}]^{1:N}$ by an EM-algorithm.

(ii) Fix the set of median characteristics. Estimate the distribution of shippers' rents from the first-order conditions of carriers' optimal bidding at the percentiles of $[\hat{G}^{\eta+p}]^{1:N}$. Shippers' IR constraint is evaluated at the fifth percentile of $[\hat{G}^{\eta+p}]^{1:N}$.

(iii) Estimate the parameters of the fundamental distribution of match-specific gains, $G^{\psi,\eta}$, by matching the moments of the distribution of carriers' rents conditional on contract rates.

where for each t,

$$\Pr(d_{j\ell t} = \text{accept}|R_{j\ell t-1}, \tilde{p}_{\ell t}; \sigma^{\zeta}_{\ell}, \tilde{\mu}^{c}_{ia\ell}, \sigma^{c}, \eta_{ij\ell} + p_{ij\ell} + \kappa_{\ell}) \tag{18}$$

$$= \Phi\left(\frac{\bar{p}(R_{j\ell t-1}, \tilde{p}_t|\eta_{ij\ell} + p_{ij\ell} + \kappa_{\ell}; \tilde{\mu}^{c}_{ia\ell}, \sigma^{c}) - \tilde{p}_{\ell t}}{\sigma^{\zeta}_{\ell}}\right) \Phi\left(\frac{\bar{p}(R_{j\ell t-1}, \tilde{p}_t|\eta_{ij\ell} + p_{ij\ell} + \kappa_{\ell}; \tilde{\mu}^{c}_{ia\ell}, \sigma^{c}) - \tilde{\mu}_{ia\ell}}{\sigma^{c}}\right). \tag{19}$$

Given a set of parameter values $(\tilde{\mu}^{c}_{ia\ell}, \sigma^{c}, \eta_{ij\ell} + p_{ij\ell} + \kappa_{\ell})$, the optimal strategy and value function of carrier j are obtained through an iterative procedure, using three conditions from the carrier's dynamic programming problem. For ease of notation, we will drop the dependence of these conditions on i, j, ℓ and the parameter values.

1) Optimality condition

$$\bar{p}(R_{t-1}, \tilde{p}_t) = \eta + p + \kappa + \frac{\delta}{1-\delta}(V(\alpha R_{t-1}, A), \tilde{p}_t) - V(\alpha R_{t-1} + (1-\alpha), \tilde{p}_t)).$$

2) Value function

$$\begin{aligned} V(R_{t-1}, \tilde{p}_{t-1}) =& \mathbf{E}_{\tilde{p}_t}[(1 - \sigma_0(R_{t-1}, \tilde{p}_t))\underline{V}(\tilde{p}_t)|\tilde{p}_{t-1}] \\ &+ \mathbf{E}_{\tilde{p}_t, \zeta_t}[\sigma_0(R_{t-1}, \tilde{p}_t)\{(1-\delta)h(\bar{p}(R_{t-1}, \tilde{p}_t), \tilde{p}_t + \zeta_t) + \delta V(\alpha R_{t-1} + (1-\alpha), \tilde{p}_t)\}|\tilde{p}_{t-1}]. \end{aligned}$$

3) Carrier's expected payoff from the outside option

$$\underline{V}(\tilde{p}_t) = (1-\delta)\mathbf{E}_{\zeta_t}[\tilde{F}(\tilde{p}_t + \zeta_t)(\tilde{p}_t + \zeta_t - \mathbf{E}[\tilde{c}_t|\tilde{c}_t \leq \tilde{p}_t + \zeta_t])] + \delta \mathbf{E}_{\tilde{p}_{t+1}}[\underline{V}(\tilde{p}_{t+1})|\tilde{p}_t].$$

Our procedure initializes the carrier value function by its expected payoff from the spot market. In each iteration, we update the full compensation $\bar{p}$ by condition 1) and the value function V by condition 2). The procedure ends when V converges.

Selection on survivals Focusing on long-lasting relationships has two potential selection issues: relationships that last long tend to have either (i) high match-specific gains or (ii) low cost draws. Since the carrier's match-specific gain is a free parameter in an individual relationship when we estimate cost parameters, selection due to (i) is not a concern. Selection due to (ii) can be corrected by conditioning on "surviving" in our construction of the likelihood function. Specifically, acceptance probability at each state will be replaced by

$$\Pr(d_{j\ell t} = \text{accept}|R_{j\ell t-1}, \tilde{p}_{\ell t}, \text{surving})$$
$$= \frac{\sigma_0(\alpha R_{j\ell t-1}, \tilde{p}_t) \Pr(d_{j\ell t} = \text{accept}|R_{j\ell t-1}, \tilde{p}_{\ell t})}{\sigma_0(\alpha R_{j\ell t-1}, \tilde{p}_t) \Pr(d_{j\ell t} = \text{accept}|R_{j\ell t-1}, \tilde{p}_{\ell t}) + \sigma_0(\alpha R_{j\ell t-1} + (1-\alpha), \tilde{p}_t) \Pr(d_{j\ell t} = \text{reject}|R_{j\ell t-1}, \tilde{p}_{\ell t})}$$

if $t < T_{ij\ell}$ and

$$\Pr(d_{j\ell t} = \text{accept}|R_{j\ell t-1}, \tilde{p}_{\ell t}, \text{non-surviving})$$
$$= \frac{[1 - \sigma_0(\alpha R_{j\ell t-1}, \tilde{p}_t)] \Pr(d_{j\ell t} = \text{accept}|R_{j\ell t-1}, \tilde{p}_{\ell t})}{\begin{pmatrix} [1 - \sigma_0(\alpha R_{j\ell t-1}, \tilde{p}_t)] \Pr(d_{j\ell t} = \text{accept}|R_{j\ell t-1}, \tilde{p}_{\ell t}) \\ +[1 - \sigma_0(\alpha R_{j\ell t-1} + (1-\alpha), \tilde{p}_t)] \Pr(d_{j\ell t} = \text{reject}|R_{j\ell t-1}, \tilde{p}_{\ell t}) \end{pmatrix}}$$

if $t = T_{ij\ell}$ and the relationship is ended because of a demotion. We will incorporate these corrections into the next iteration of our estimation. However, since the contribution of an acceptance or rejection to the probability of demotion is small, we expect negligible changes in the results due to these corrections.

C.1.3 The EM-algorithm

We approximate the (continuous) distribution of carriers' rents by a mixture of $K = 5$ Normal distributions. Thus, the parameters to estimate are the mean and variance of each distribution, $(\mu_k^c, \sigma_k^c)_{k=1}^K$, and their shares, $(\pi_k)_{k=1}^K$. For estimation, we adapt an EM algorithm by Train (2008): the M-step integrates the likelihood function over these Normal distributions, and the E-step updates their means, variances, and shares. To speed up the integration step, we discretize carrier rent into a grid, $\{0.0, 0.1, ..., 5.0\}$, and perform linear interpolation on these grid points. In other words, the likelihood contribution of each relationship at each grid point is calculated only once, and when the distribution of carrier rent is updated, we only need to update the weights being put on these grid points.

Treatment of heterogeneity within sub-clusters. In estimating the distribution of carrier rent in each sub-cluster, we keep observable characteristics relationship-specific rather than using a representative set of characteristics for all relationships in the sub-cluster. The reason for our decision is to avoid inflating the heterogeneity of the estimated distribution of carrier rent. Via an auction approach, such inflated heterogeneity would result in an upward bias of the estimated distribution of shipper rent. Representative (median) characteristics are only used in subsequent steps, where we need to estimate the expected payoffs of shippers and carriers conditional on their rents at the auction stage.

C.1.4 Carriers' bidding function

Our empirical model allows for two sources of randomness in the number of effective bidders: (i) the number of carriers who submit a bid, $N_a \sim \text{Binomial}(N, q)$, and (ii) the number of carriers who pass the shippers' individual rationality constraint, $n_a \sim \text{Binomial}(N, \tilde{q})$. We will show that a carrier's bidding problem can be rewritten as follows,

$$\max_r [\tilde{G}^{\eta+p}(r)]^{N-1}(V(R_0, \tilde{p}_0|\theta - \mathbf{b}_r(r)) - \mathbf{E}[\underline{V}(\tilde{p}_1)|\tilde{p}_0], \tag{20}$$

where carriers' probability of winning, $[\tilde{G}^{\eta+p}(r)]^{N-1}$, can be estimated from the distribution of winning carriers' rents.

Recall that $[G^{\eta+p}(r)]^n$ is the distribution of winning carriers' rents in auctions with n effective bidders. The distribution of winning carriers' rents estimated in C.1.3 aggregates auctions with different numbers of bidders, conditional on there being at least one that passes shippers' individual rationality constraint:

$$\frac{\prod_{n=1}^{N} \binom{N}{n} \tilde{q}^n (1-\tilde{q})^{N-n} \left[\frac{G^{\eta+p}(r) - G^{\eta+p}(\underline{r})}{1 - G^{\eta+p}(\underline{r})}\right]^n}{\prod_{n=1}^{N} \binom{N}{n} \tilde{q}^n (1-\tilde{q})^{N-n}} = \frac{\left(1 - \tilde{q} + \tilde{q}\left[\frac{G^{\eta+p}(r) - G^{\eta+p}(\underline{r})}{1 - G^{\eta+p}(\underline{r})}\right]\right)^N - (1-\tilde{q})^N}{1 - (1-\tilde{q})^N}.$$

Given $\tilde{q}$ from Appendix C.1.1, the above equation pins down $\tilde{G}^{\eta+p}(r) \equiv 1 - \tilde{q} + \tilde{q}\left[\frac{G^{\eta+p}(r) - G^{\eta+p}(\underline{r})}{1 - G^{\eta+p}(\underline{r})}\right]$. Moreover, this gives carriers' probability of winning conditional on their rent, since

$$\prod_{n=1}^{N} \binom{N}{n-1} \tilde{q}^{n-1} (1-\tilde{q})^{N-n} \left[\frac{G^{\eta+p}(r) - G^{\eta+p}(\underline{r})}{1 - G^{\eta+p}(\underline{r})}\right]^{n-1} = [\tilde{G}^{\eta+p}(r)]^{N-1}.$$

To estimate carriers' bidding function $\mathbf{b}_r$, we estimate $\mathbf{b}_r'$ from the first-order-condition of (20) and the initial condition $\mathbf{b}_r(\underline{r})$ from the binding individual rationality constraint of shippers. These first-order-conditions are evaluated at the percentiles of the distribution of winning carriers' rents estimated in C.1.3, and the lowest rent-type $\underline{r}$ is the fifth percentile of this distribution.

C.1.5 Estimate the fundamental distribution of match-specific gains

We parameterize the underlying distribution of match-specific gains $G^{\psi,\eta}$ by

$$\begin{pmatrix}\psi\\ \eta\end{pmatrix} \sim \text{Normal}\left(\begin{pmatrix}\mu_\psi\\ \mu_\eta\end{pmatrix}, \begin{pmatrix}\sigma_\psi^2 & \sigma_{\psi\eta}\\ \sigma_{\psi\eta} & \sigma_\eta^2\end{pmatrix}\right).$$

Given the lowest type $\underline{\theta}$ and bidding function $\mathbf{b}_r$ estimated in the previous steps, we can simulate from $G^{\psi,\eta}$ the joint distribution of match-specific gains, contract rates, rents and bids of observed relationships. We estimate the parameters of $G^{\psi,\eta}$ by matching simulated moments to the empirical moments of observed relationships.

Specifically, consider two bins of contract rates divided by the median contract rate p_{med}. We use the first and second moments of the distribution of carrier rent in these two bins, $[G^{\eta+p|p\leq p_{\text{med}}}]^{1:N}$ and $[G^{\eta+p|p>p_{\text{med}}}]^{1:N}$, and the first and second moments of the distribution of contract rates, $[G^p]^{1:N}$. For the empirical moments, we use the empirical distribution of contract rates and use an EM-algorithm to obtain the distributions of carrier rent within each bin of contract rates.

C.2 Details of counterfactual analysis

C.2.1 Estimate cluster-specific market shocks

For each lane-specific cluster and market phase, we combine our sample of long-term relationships and DAT data on the spot market to construct market-level demand and supply factors in a full year. Our sample of long-term relationships gives the total number of loads demanded by the shippers ($\hat{L}_t$), those loads that are accepted by the primary carrier ($\hat{L}_t^{\text{primary}}$), those that are accepted by a backup carrier ($\hat{L}_t^{\text{backup}}$), and those that are rejected by all carriers and fulfilled in the spot market. Our spot market data gives the spot rate $\tilde{p}_t$ and the number of spot loads $\hat{S}_t$. We scale the number of loads in our sample on long-term relationships to reflect the aggregate market share of long-term relationships, obtaining a series $(\hat{L}_t, \hat{L}_t^{\text{primary}}, \hat{L}_t^{\text{backup}}, \hat{S}_t, \tilde{p}_t)_{t=1}^{52}$.

From the series of observed loads and spot rates, we recover a series of demand for long-term relationships, direct spot demand and spot capacity, $(L_t, D_t, C_t)_{t=1}^{52}$ that are consistent with the market equilibrium condition in Equation (8). There are two empirical issues: first, our sample can provide noisy estimates of weekly volumes within long-term relationships and second, our model abstracts from backup carriers. We address these issues by assuming that the demand for long-term relationships is constant within a market phase and counting loads accepted by backup carriers as spot loads. Specifically, we set $L_t = L = \frac{1}{52}\sum_{t=1}^{52}\hat{L}_t$ and assume that all relationships start at $t = 0$. This allows us to capture some correlation in the evolution of the relationship status and the market condition. We then use our estimates of the incentive scheme, distribution of match-specific gains, search and operational costs to estimate the number of loads accepted by the primary carriers (L_t^{primary}) and the number of carriers that reject loads offered within relationships to service the spot market (L_t^{spot}). We also estimate the number of loads accepted by backup carries (L_t^{backup}) and count these loads towards spot loads.

These volume estimates, the estimated model primitives, and the observed spot rate allow us to pin down direct spot demand and spot capacity through the market equilibrium condition

$$L + D_t = \underbrace{L \int_{\tilde{p}_t}^{\infty} F(\bar{p} - \kappa) d\mu(\bar{p}|\tilde{p}_t)}_{L_t^{\text{primary}}} + \underbrace{[C_t F(\tilde{p}_t - \kappa) + \overbrace{L_t \mu(\tilde{p}_t|\tilde{p}_t) F(\tilde{p}_t - \kappa)}^{L_t^{\text{spot}}}]}_{S_t = \hat{S}_t + L_t^{\text{backup}}}.$$

Specifically,

$$D_t = L_t^{\text{primary}} + L_t^{\text{backup}} + \hat{S}_t - L \quad \text{and} \quad C_t = \frac{\hat{S}_t + L_t^{\text{backup}} - L_t^{\text{spot}}}{F(\tilde{p}_t - \kappa)}.$$

In our counterfactual analysis, we keep fixed $(L, D_t, C_t)_{t=1}^{52}$.

C.2.2 Market-level welfare of a centralized spot market

In a centralized spot market, demand for long-term relationships is combined with direct spot demand, and all carriers in the market make up spot capacity. Thus, the equilibrium spot rate in each period is pinned down by

$$L + D_t = \underbrace{(L + C_t) F(\tilde{p}_t - \kappa^1)}_{S_t^{\text{centralized}}},$$

where

$$\kappa^1 = \kappa + \frac{\gamma_1 \ln \left(\frac{\sum_t S_t^{\text{centralized}}}{\sum_t S_t} \right)}{\text{Distance}}$$

is the equilibrium search cost in a centralized spot market. Notice how the new level of search cost depends only on the scale efficiency parameter γ_1, distance and how much spot market volume has scaled up due to the centralization of all transactions into the spot market. Normalizing shippers' gains from having their loads shipped to zero, we obtain the following measure of aggregate welfare of a centralized spot market in period t,

$$W_t^1 = \sum_t (L + D_t)(-\kappa^1 - \mathbf{E}[c_t | c_t \le \tilde{p}_t - \kappa^1]).$$

C.2.3 Market-level welfare of index-priced contracts

Denote by κ^2 the equilibrium search cost under index-priced contracts. In addition, recall that $\theta_{ij} = \psi_{ij} + \eta_{ij}$ denotes the match-quality of carrier i and shipper j, excluding savings on search costs. Under index-priced contracts, any relationship with $\theta_{ij} + \kappa^2 \ge 0$ generates surplus over spot transactions. That is, the lowest match-quality in a relationship is $\underline{\theta} = -\kappa^2$. Moreover, carriers in long-term relationships with individually optimal index-priced contracts might reject when costs are high, but never reject in order to service the spot market. Specifically, carrier j rejects if and only if $\theta_{ij} + \tilde{p}_t \le c_t$. Thus, the equilibrium spot rate in each period is pinned down by

$$L + D_t = L \int_{\theta \ge -\kappa^2} F(\theta + \tilde{p}_t) d[G^\theta]^{1:N}(\theta) + \underbrace{C_t F(\tilde{p}_t - \kappa^2)}_{S_t^{\text{index}}},$$

where

$$\kappa^2 = \kappa + \frac{\gamma_1 \ln \left(\frac{\sum_t S_t^{\text{index}}}{\sum_t S_t} \right)}{\text{Distance}}.$$

The aggregate welfare under index-priced contracts in period t is

$$W_t^2 = L \int_{\theta \ge -\kappa^2} F(\theta + \tilde{p}_t)(\theta - \mathbf{E}[c_t | c_t \le \theta + \tilde{p}_t]) d[G^\theta]^{1:N}(\theta) + C_t F(\tilde{p}_t - \kappa^2)(-\kappa^2 - \mathbf{E}[c_t | c_t \le \tilde{p}_t - \kappa^2]).$$

To estimate the split of relationship surplus between the shipper and the carrier, we rely on two observations. First, under an individually optimal index-priced contract, the full surplus from an individual relationship is realized. Second, the shipper's surplus is precisely the fixed fee b_{ij}^0 that the carrier bids on. Denote by $\text{Surplus}(\tilde{p}_0 | \theta_{ij})$ the total expected surplus of the relationship between shipper i and carrier j with match quality θ_{ij} and initial spot rate $\tilde{p}_0$. There exists a symmetric monotone bidding equation $\mathbf{b}^0 : \text{Surplus}(\tilde{p}_0 | \theta_{ij}) \mapsto b_{ij}^0$ mapping each level of expected surplus to a bid. Specifically, $\mathbf{b}^0$ satisfies that for all $\theta_{ij} \ge -\kappa^2$,

$$\mathbf{b}^0(\theta_{ij}) = \text{Surplus}(\tilde{p}_0 | \theta) - \frac{\int_{-\kappa^2}^{\theta_{ij}} \text{Surplus}(\tilde{p}_0 | \theta) d[G^\theta]^{1:N}(\theta)}{[G^\theta]^{1:N}(\theta_{ij})}.$$

This in turn pins down the expected surplus of the shipper and the carrier in each relationship.

C.2.4 An upper bound on market-level first-best welfare

Our upper bound on the market-level first-best welfare in a period uses the same formula as the welfare under index-priced contracts but replaces the search cost under index-priced contracts with the search cost of a centralized spot market,

$$\overline{W}_t = L \int_{\theta \geq -\kappa^1} F(\theta + \tilde{p}_t)(\theta - \mathbf{E}[c_t | c_t \leq \theta + \tilde{p}_t]) d[G^\theta]^{1:N}(\theta) + C_t F(\tilde{p}_t - \kappa^1)(-\kappa^1 - \mathbf{E}[c_t | c_t \leq \tilde{p}_t - \kappa^1]).$$

In each counterfactual, we report the welfare averaged over time: $W^0 = \frac{1}{52}\sum_t W_t^0$ for fixed-rate contracts (baseline), $W^1 = \frac{1}{52}\sum_t W_t^1$ for a centralized spot market, $W^2 = \frac{1}{52}\sum_t W_t^2$ for index-priced contracts and $\overline{W} = \frac{1}{52}\sum_t \overline{W}_t$ for the unattainable market-level first-best welfare. The performance of fixed-rate and index-priced contracts are measured as the ratio of their surplus relative to a centralized spot market over the surplus of the unattainable first-best relative to a centralized spot market, $\frac{W^0 - W^1}{\overline{W} - W^1}$ and $\frac{W^2 - W^1}{\overline{W} - W^1}$ respectively.

C.3 Lane-specific shares of spot volumes

We construct fitted values from the coefficient estimate $\hat{\beta}_1 = 0.345$ in the (IV) specification in Table 4 and the fact that the aggregate share of spot volume is 20%. Denote by SpotShare_ℓ the share of spot volume on lane ℓ. We have

$$\frac{\text{Volume}_\ell^{\text{spot}}}{\text{Volume}_\ell^{\text{LT}}} = \exp(\tilde{\beta}_0)\left(\text{Volume}_\ell^{\text{spot}} + \text{Volume}_\ell^{\text{LT}}\right)^{0.345} \quad \text{and} \quad \frac{\sum_\ell \text{Volume}_\ell^{\text{spot}}}{\sum_\ell (\text{Volume}_\ell^{\text{LT}} + \text{Volume}_\ell^{\text{spot}})} = 20\%,$$

where $\text{Volume}_\ell^{\text{LT}}$ is the population long-term relationship volume on lane ℓ, and $\tilde{\beta}_0$ is the scaling parameter that we need to calibrate. Rewriting the above equations gives,

$$\frac{\text{SpotShare}_\ell}{1 - \text{SpotShare}_\ell} = \exp(\tilde{\beta}_0)\left(\frac{\text{Volume}_\ell^{\text{spot}}}{\text{SpotShare}_\ell}\right)^{0.345} \quad \text{and} \quad \frac{\sum_\ell \text{Volume}_\ell^{\text{spot}}}{\sum_\ell \text{Volume}_\ell^{\text{spot}}/\text{SpotShare}_\ell} = 20\%,$$

which we use to calibrate $\tilde{\beta}_0$ and $(\text{SpotShare}_\ell)_\ell$ from the observed spot volume $(\text{Volume}_\ell^{\text{spot}})_\ell$.

Table 8: Estimates of the relational incentive scheme

Rejection index I_R			Demotion probability $1-\sigma_0$		
Parameter	Estimate	95% CI	Variable	Estimate	95% CI
α	0.97	$(0.97, 0.98)$	Constant	-2.58	$(-2.60, -2.54)$
R_0	0.97	$(0.96, 0.99)$	R_{t-1}	0.69	(0.59,0.72)
			$\tilde{p}_t$	-0.29	$(-0.37, -0.19)$
			Frequency	-0.29	$(-0.31, -0.24)$
			Inconsistency	0.46	(0.41,0.53)
			$R_{t-1} \times \tilde{p}_t$	0.90	(0.70,1.14)
			$R_{t-1}\times$ Frequency	0.22	(0.11,0.29)
			$R_{t-1}\times$ Inconsistency	0.10	$(-0.09, 0.30)$

Note: α is the daily decay parameter; Frequency is the log of average monthly volume; Inconsistency is the average coefficient of variation of weekly volume within a month. The confidence intervals are bootstrapped at the auction level.

D Other results

D.1 Shipper's incentive scheme and instrumental objects

Shippers' incentive scheme. Table 8 presents our estimates of the shipper's incentive scheme. The coefficient of the rejection index R_{t-1} is positive and highly significant, confirming Assumption 1 in our model that shippers punish carriers' rejections with higher probability of demotion. To interpret the magnitude of this coefficient, we simulate two sets of relationships, one with an initial rejection and one with an initial acceptance. We find that a carrier's initial rejection instead of an acceptance reduces the expected number of offers it receives by 3% (from 91 loads to 88 loads). This suggests that the shipper's incentive scheme is soft but generates dynamic incentives that are economically significant.

Consistent with Harris and Nguyen (2021), we find that when the current spot rate ($\tilde{p}_t$) is high or if the shipper has large volume on a lane, that is, when the relationship is more valuable to the shipper, demotion probability is lower. However, also in such cases the shipper strengthens the incentive scheme, punishing the carrier's rejections more harshly to induce more acceptances.[46] Additionally, we find that the daily discount rate α on the carrier's past rejections is close to one. This suggests that a rejection of the carrier affects the continuation probability of its relationship in many periods.

[46] In Harris and Nguyen (2021), we specified a linear probability model that includes both asset-owners and brokers, and estimate our specification by GMM.

Table 9: The link between spot shares and total market thickness

	Dependent variable: $\ln(\text{Volume}_\ell^{\text{spot}}) - \ln(\text{Volume}_\ell^{\text{LT}})$							
	≥ 5 relationships		≥ 10 relationships		≥ 15 relationships		≥ 20 relationships	
	(OLS2)	(IV)	(OLS2)	(IV)	(OLS2)	(IV)	(OLS2)	(IV)
$\ln(\text{Volume}^{\text{total}})$	0.162	0.257	0.269	0.345	0.272	0.334	0.321	0.408
	(0.032)	(0.039)	(0.037)	(0.045)	(0.044)	(0.052)	(0.048)	(0.059)
ln(distance)	−0.090	0.001	−0.028	0.032	0.002	0.042	0.048	0.097
	(0.070)	(0.074)	(0.075)	(0.077)	(0.083)	(0.085)	(0.089)	(0.091)
Frequency	−0.261	−0.262	−0.156	−0.159	−0.221	−0.228	−0.366	−0.367
	(0.067)	(0.067)	(0.076)	(0.076)	(0.109)	(0.110)	(0.133)	(0.134)
Inconsistency	1.597	1.591	1.682	1.653	1.415	1.387	1.325	1.303
	(0.278)	(0.279)	(0.319)	(0.321)	(0.429)	(0.430)	(0.494)	(0.497)
origin = MidWest	−0.407	−0.381	−0.378	−0.354	−0.304	−0.281	−0.221	−0.186
	(0.082)	(0.082)	(0.085)	(0.086)	(0.095)	(0.096)	(0.098)	(0.100)
destination = MidWest	−0.061	−0.030	−0.090	−0.071	−0.141	−0.122	−0.053	−0.029
	(0.090)	(0.091)	(0.096)	(0.096)	(0.103)	(0.104)	(0.108)	(0.109)
Instrument								
ln(PredictedFlow)	✓	✓	✓	✓	✓	✓	✓	✓
N	887	887	588	588	427	427	321	321

Notes: Standard errors are in parentheses.

Other instrumental objects. We calibrate the daily discount rate to 0.992, under the assumption that (i) shippers and carriers are patient and (ii) auction periods end randomly with an estimated average duration of 320 days. We estimate that the number of effective bidders is distributed as $\text{Binomial}(15, 0.21)$, averaged to 3 effective bidders per auction.

D.2 Robustness of the link between spot market thickness and efficiency

Table 9 presents regression results of Equation (3) as we vary the sets of lane characteristics and the sample restriction to include lanes with at least 5, 10, 15 or 20 relationships (in the microdata). All sample restrictions and specifications show a strong positive link between spot shares and total market thickness.

Table 10 presents our decomposition of the mean transformed costs into the mean operational costs and search costs. Specification (1) is our main specification, where we estimate Equation (17) by two-stage least squares, including as controls a polynomial of degree three of distance and an indicator of market tightness. One potential concern of the main specification is that patterns of trade affect the equilibrium movements of trucks and thus may correlate with unobserved cost shifters. For example, a thick lane may appear desirable not because search costs are lower on this lane but because it is connected to other thick lanes; this connectivity would help carriers reduce empty miles. To control for such network effects, we construct a measure for the imbalance between forehauls and backhauls, calculated as

$$\text{Imbalance}_\ell = \ln(\text{Volume}^{\text{spot}}_{-\ell}) - \ln(\text{Volume}^{\text{spot}}_{\ell}),$$

where $-\ell$ denotes the backhaul going from the destination of lane ℓ to the origin of lane ℓ. This measure captures the likelihood of finding a backhaul, which is the key concern for carriers on long trips (at least 250 miles). Specification (2) includes this measure of volume imbalance on a lane and specification (3) additionally controls for the frequency of interactions and consistency of load timing of the shipper-lane within the contract period. Our estimate of the scale efficiency parameter, the coefficient of $\ln(\text{Volume}^{\text{spot}})/\text{distance}$, is similar across these three specifications.

Specifications (4), (5) and (6) in Table 10 lend support to our functional form assumption on the relationship between per-mile search costs and spot market thickness. They show that per-mile search costs decrease with the thickness of the spot market, but less so on longer lanes.

D.3 Cluster-specific results

We estimate the distribution of rents and match-specific gains, and perform welfare analysis separately on each of our ten lane-specific clusters and in two market phases.

Figure 18 plots our lane-specific clusters in different shades of gray. The left panel plots the clustered relationships against the average spot rate and log of the average spot volume, which are equilibrium objects used in our K-means clustering. The right panel plots these clusters against their median characteristics, with the sum of search and operational costs representing the supply factor and the log of predicted trade flows representing the demand factor; the size of each cluster represents the number of relationships in that cluster. These scatter plots show that our K-means clustering performs well in separating lanes by their underlying demand and supply factors.

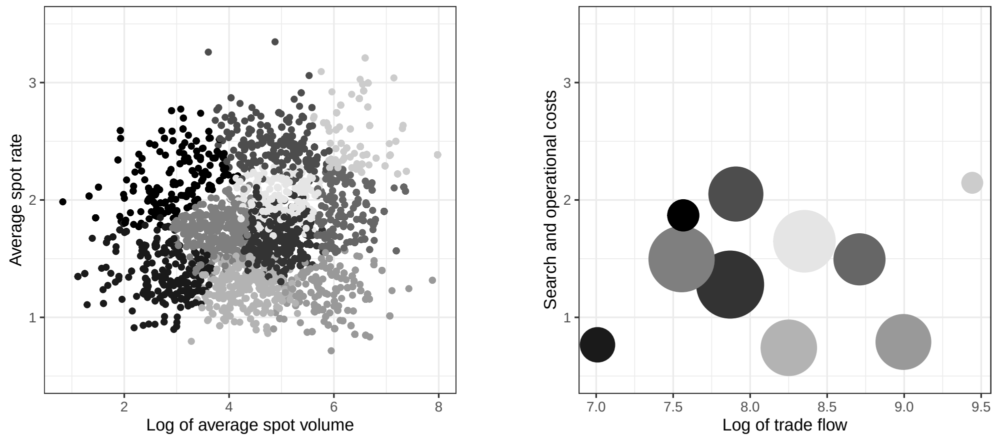
Although the clusters are different in their underlying demand and supply factors, the takeaways from the cluster-specific results are consistent with the aggregated results reported in Section 7. Figure 19 plots the distribution of match quality across all relationships within each lane-specific cluster and each market phase, showing large heterogeneity in match quality within each cluster. Figure 20 plots the median match-specific gains of shippers and carriers in each of our 20 clusters, showing that shippers tend to have larger match-specific

Table 10: Cost decomposition

Mean transformed cost	(1)	(2)	(3)	(4)	(5)	(6)
$\ln(\text{Volume}^{\text{spot}})/\text{distance}$	−255.850	−229.428	−229.866			
	(77.625)	(68.241)	(67.363)			
$\ln(\text{Volume}^{\text{spot}})$				−0.883	−0.719	−0.747
				(0.202)	(0.151)	(0.149)
$\ln(\text{Volume}^{\text{spot}}) \times \ln(\text{distance})$				0.575	0.536	0.518
				(0.441)	(0.392)	(0.397)
Tight market	0.536	0.547	0.505	0.548	0.558	0.505
	(0.093)	(0.086)	(0.083)	(0.106)	(0.092)	(0.091)
Imbalance		−0.359	−0.354		−0.443	−0.446
		(0.082)	(0.080)		(0.100)	(0.098)
Instruments						
ln(PredictedFlow)/distance	✓	✓	✓			
ln(PredictedFlow)				✓	✓	✓
ln(PredictedFlow) × distance				✓	✓	✓
Controls						
h_3(distance)	✓	✓	✓	✓	✓	✓
Frequency			✓			✓
Inconsistency			✓			✓
N	1162	1162	1162	1162	1162	1162

Notes: Standard errors (in parentheses) are constructed by bootstrapping at the auction level.

Figure 18: Ten lane-specific clusters by K-means method



gains than carriers from their relationships. Finally, the left panel of Figure 21 shows that under the current fixed-rate contracts, the median relationship in each cluster achieves 40-50% of the relationship-level first-best surplus. The right panel of Figure 21 shows that in the median relationship of each cluster, the shipper gets 75% of the total surplus of the relationship relative to spot transactions.

E Additional figures

Figure 22 confirms that under our estimates of the spot process and shippers' incentive scheme, substantive assumptions on carriers' full compensation and shippers' expected payoff are satisfied. The left panel shows that the carrier's full compensation is increasing in its rent (Assumption 3). The right panel shows that the shipper's expected payoff is increasing in both her rent and the carrier's rent (Assumption 4).

Figure 23 shows a screenshot of DAT load board when a carrier searches for a load. This carrier conducted multiple searches; in each search, it input an origin and a destination, possibly with a radius around these locations, date availability, and some basic information on equipment. The highlighted search is for a load from Houston, Texas to any location on December, 11. The search results show a list of shippers and their contact information. The carrier would contact these shippers and negotiate rates off the platform.

Figure 19: Match-quality (including savings on search costs)

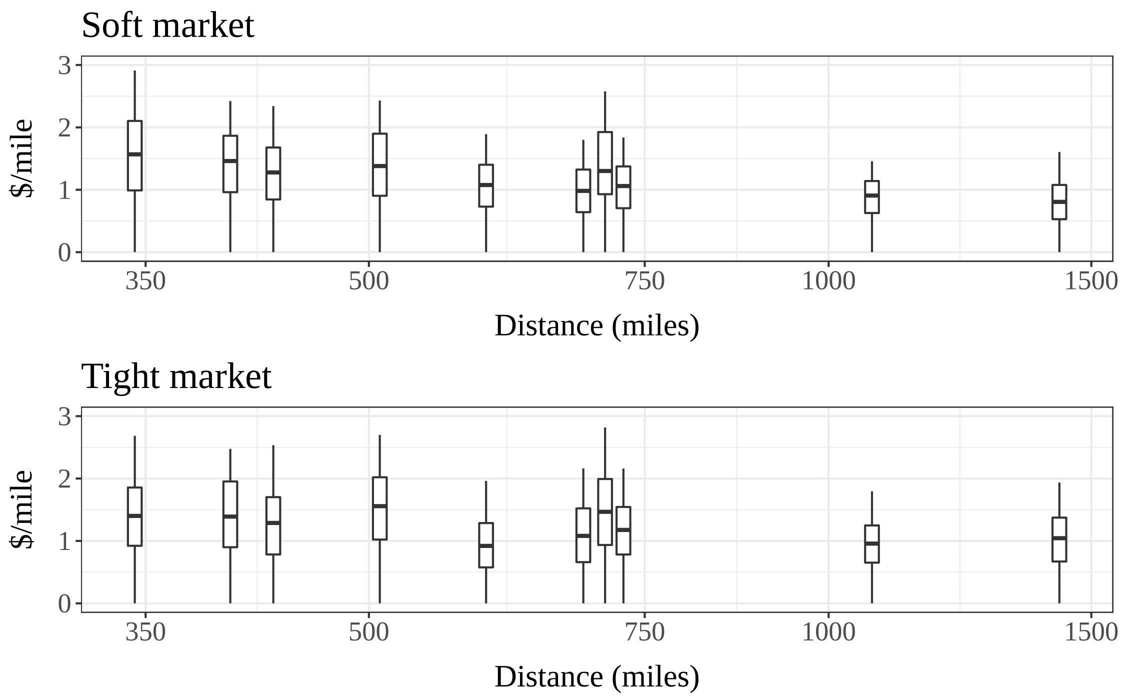


Figure 20: Shippers and carriers' match-specific gains (including savings on search costs)

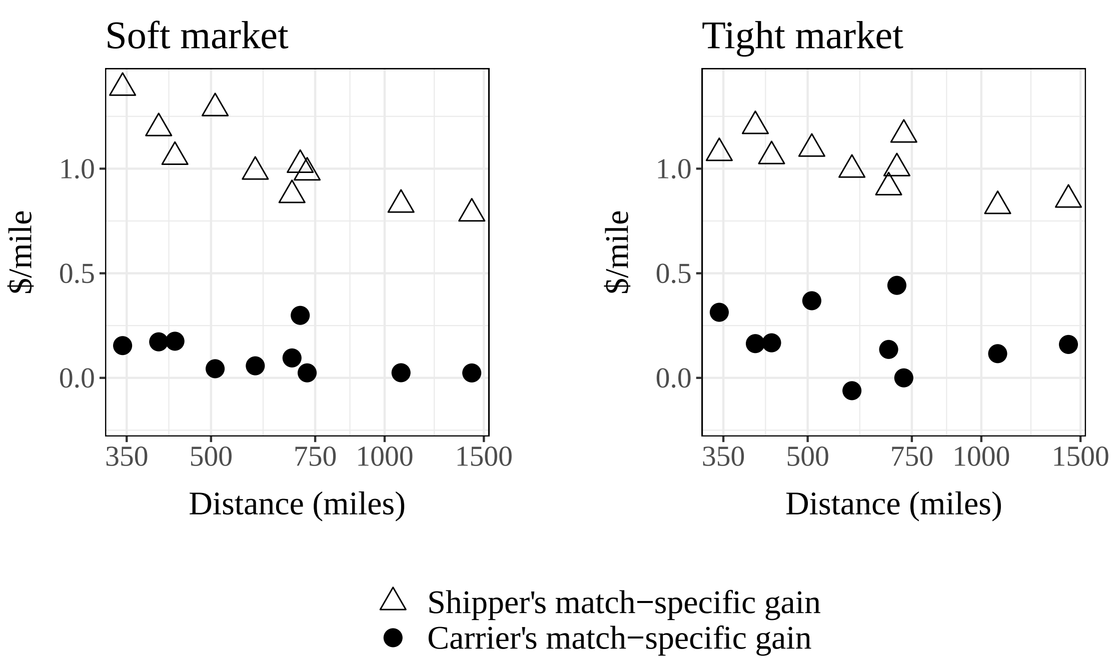


Figure 21: The median share of total surplus to the first-best surplus and its share by shippers

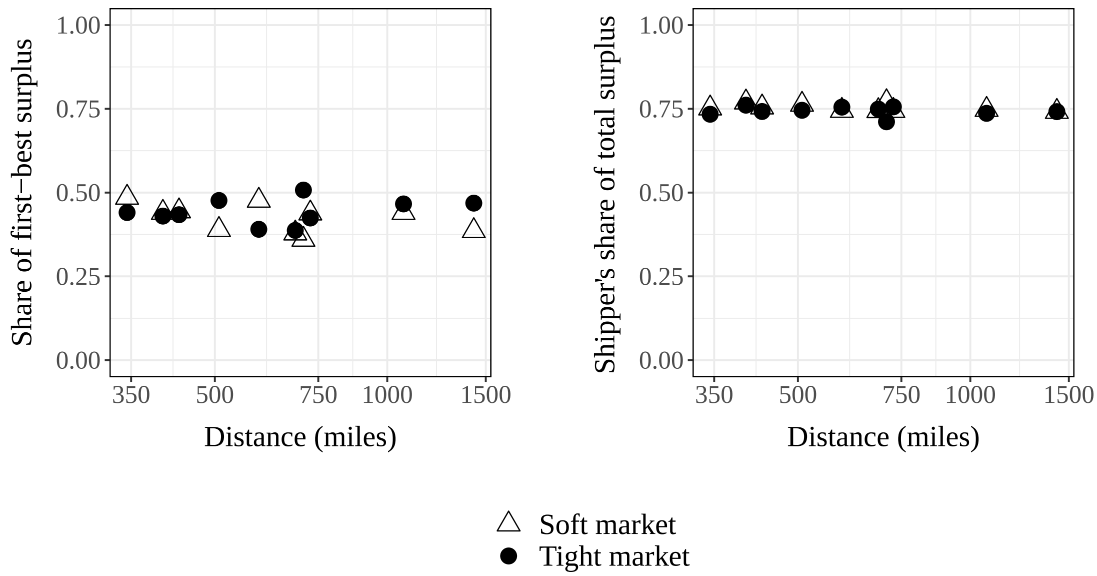


Figure 22: Monotonicity of carrier's full compensation and shipper's expected payoffs in rents

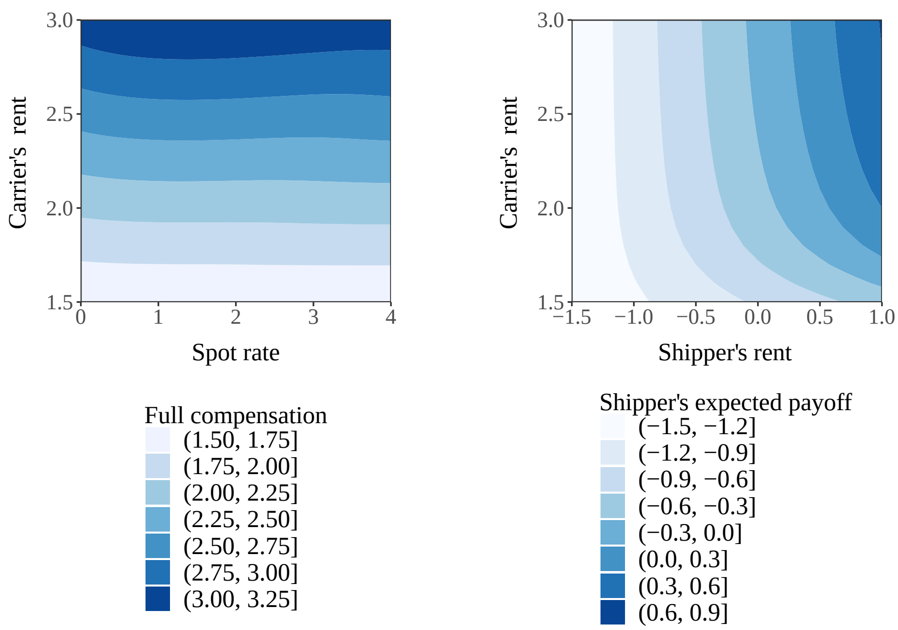
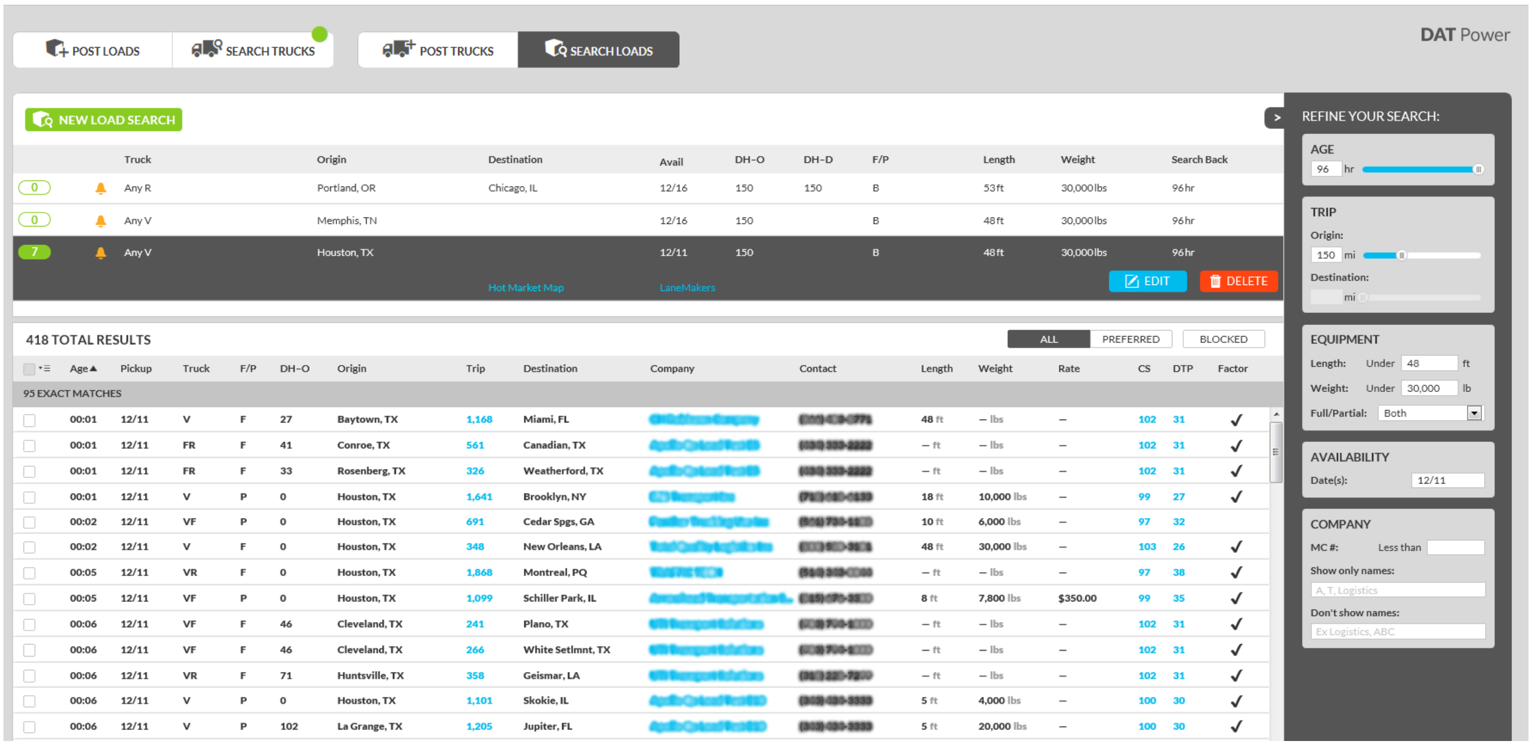


Figure 23: Example of DAT load board



Source: https://forms.dat.com/resources/product-sheets/dat-load-boards#.

Chapter 3

Empiricist Learning Rules on Social Networks: Learning and Quality of Information Aggregation*

Thi Mai Anh Nguyen

1 Introduction

People interact in complicated networks while often having little understanding about the structure of these networks. For example, using network data and surveys from 75 villages in India, Breza et al. (2018) find that 46% of respondents are not certain enough to elicit a guess about whether two given individuals have financial, social or informational links. Furthermore, when the respondents do make guesses, only 37% of them are correct. The reality that knowledge about the network structure is local poses a challenge to information aggregation, since individuals may fail to correctly assess the reliability of information from their neighbors.

Previous papers in the literature that relax common knowledge of the full network have abstracted from learning.[1] An open question is whether individuals with limited knowledge of their complicated environment, through repeated interactions, could learn to aggregate information from their neighbors optimally. Furthermore, how does such learning affect the quality of information aggregation by the network as a whole?

To answer these questions, I propose a novel model of information diffusion and aggregation in settings like Facebook or Twitter. Three key features of these social platforms motivate my modeling choices. First, individuals interact frequently on multiple, relatively

*Department of Economics, Massachusetts Institute of Technology. Email: anhng@mit.edu. I am particularly indebted to Glenn Ellison and Stephen Morris. I am also grateful to Krishna Dasaratha, Drew Fudenberg, Giacomo Lanzani, Claudio Mezzetti, Alex Wolitzky and participants in the MIT Theory Lunch and the 2021 European Summer Meeting of the Econometric Society for helpful comments and suggestions.

[1] Most papers in the social learning literature maintain common knowledge of the network topology and focus on Perfect Bayesian equilibria. DeGroot-style learning papers, for example, relax these common knowledge assumptions but use heuristics of updating rules.

short-lived topics. Second, the diffusion of information depends on an underlying network of individuals, who are quick to respond to the posts of their friends (Facebook) or of those they follow (Twitter). Third, the underlying network is rich, making it impossible for an individual to perfectly know others' links. My model captures these features in a stylized way. It features in each period a new draw of nature and random, asynchronous timing of signal arrivals and decision making. Local knowledge of the network structure and local observability are assumed. Despite individuals' limited knowledge and observability, a class of empiricist learning rules achieve convergence of learning on all networks. The induced long-run play is then shown to achieve strong efficiency on clique trees, tree-like networks where a node can be replaced by a set of fully linked individuals. In general circumstances, however, there are systematic ways in which misalignment of interests arises, challenging efficiency.

My model departs from the literature in two main ways:

First, it has two layers of timing that separate learning about the environment from aggregating information about the current state. The infinite number of periods put a repetitive structure on the interactions between each individual and her environment, facilitating learning about the environment. The multiple rounds within each period allow information about a newly drawn state of the world to spread through network links. Together with the network structure, the random arrivals of newly drawn private signals determine the order of play and the set of actions observed by each individual in each period. This means that each period is a game of information aggregation where individuals know little about how information from different sources have diffused throughout the network before reaching them.

Second, each individual in my model faces vast uncertainty about the underlying environment. This includes objective uncertainty, captured by the network structure, the distribution of private signals and their arrivals, and strategic uncertainty, encoded in how other individuals map their private observations to actions. Specifically, each individual only knows the set of her neighbors, her own private signal as well as her own belief-updating rule and decision rule. Moreover, others' belief-updating rules and decision rules are treated as primitives of the learning environment rather than strategic choices. This paper's objective is to find a class of learning rules whose performance is robust to individuals' vast uncertainty about their environment.

One such class is the class of *empiricist learning rules*. These are learning rules that believe in stationary environments and asymptotically believe in the empirical distribution of observations. Empiricist learning rules rest on two ideas. First, to optimally aggregate information in stationary environments, an individual only need to learn the stationary distributions of her observation conditional on each state of the world. In other words, the collection of these stationary conditional distributions, referred to as a *local model*, is the key object of learning for her. Second, since each individual possesses an *independent* and *informative* signal, there is a one-to-one map between an individual's local model and the unconditional distribution of her observation. This further reduces the object of learning to the unconditional distribution of an individual's observation, which, in stationary environments, can be learned asymptotically from its empirical counterpart. In Proposition 1, I show that if an individual's environment is indeed stationary, empiricist learning rules ensure asymptotic learning of her true local model and thus, asymptotic optimality.

What if all individuals on the network actively learn with empiricist learning rules, rendering the underlying environment nonstationary? In which cases will play converge and information aggregation be socially optimal?

Theorem 1 shows that convergence of beliefs and of play holds in all environments where individuals adopt empiricist learning rules and "smooth" optimal decision rules. In each round, as the play of those neighbors that acted before an individual converges, the local environment of that individual becomes approximately stationary. Empiricist learning rules then ensure convergence of her belief about her local environment, in that round, and smooth decision rules translate convergence of beliefs into convergence of play. Crucially, empiricist learning rules separate learning in different rounds, thus avoiding potential mislearning in later rounds to contaminate learning in earlier rounds.

In fact, the convergence result in my setting extends to richer settings, suggesting that adaptations of my model could provide a framework for studying other games of information aggregation. This framework starts with building a model where the game of interest is played repeatedly. This model is then analyzed in two steps. First is to achieve convergence of learning about local models, or elements of the environment that are relevant to each individual's decision making. Second is to evaluate the per-period game where each individual knows her local model perfectly. Compared to a direct analysis of the game of interest, this framework generates predictions about individuals' play as the long-run outcome of a learning process. In doing so, it makes minimal assumptions on individuals' knowledge of their environment and of others' play.

The remainder of this paper focuses on the long-run quality of information aggregation in my setting when all individuals use empiricist learning rules and smooth decision rules.

Theorem 2 establishes a strong positive result: on *clique trees*, the long-run play induced by empiricist learning rules and smooth decision rules is *strongly efficient*, among all stationary plays that treat the two states symmetrically. Two properties of clique trees guarantee this result. First is *conditional independence of neighbors' actions*. Under symmetric treatment of the two states of the world, this property implies that the overall informativeness of all neighbors' actions is increasing in the informativeness of each neighbor's action. Second, clique trees ensure *local alignment of interests*: an individual maximizes the informativeness of her action to her neighbors exactly by optimally aggregating information from other neighbors of hers. The proof of Theorem 2 uses an inductive argument to show that under conditional independence of neighbors' actions, local alignment of interests leads to global alignment of interests. Since the long-run play induced by empiricist learning rules and smooth decision rules is individually optimal, it follows that this long-run play is strongly efficient; that is, it is optimal for every individual on the network.

What if the underlying network is not a clique tree, or the notion of efficiency is weakened, or the environment is not symmetric? Theorem 3, which is a weak converse to Theorem 2, illustrates the role of clique trees in achieving efficiency: on any network that is not a clique tree, substantive misalignment of interests arises for some diffusion process and signal structure. This theorem generalizes two examples of networks that are not clique trees, where either conditional independence of neighbors' actions or local alignment of interests fails. Example 1 shows that when there is conditional correlation between different information sources, it might be possible to reduce the informativeness of an individual source in a way that breaks the correlation and improves the informativeness of all sources combined. Example 2 illustrates how an individual might play optimally against a distribution that pools diffusions irrelevant to her neighbors, and thus fails to optimize the informativeness of her action to these neighbors. Moreover, forces that challenge strong efficiency are likely to also challenge Pareto efficiency. In two examples that extend Example 1 and Example 2, I show that Pareto efficiency fails because individuals can trade favors across rounds. Lastly, coordinated biases about the two states of the world can improve the informativeness of combined sources. This makes strong efficiency hard to achieve when asymmetric plays are considered for comparison.

In the setting of this paper, that individuals cannot trace the path of their information does not hurt learning but it generates misalignment of interests. In other words, empiricist learning rules solve the problem of local knowledge of the network structure, ensuring asymptotic learning and individual optimality. However, the gap between individual optimality and social optimality remains. This gap depends on all elements of the objective environment: the network, the diffusion process, and the quality of private signals.

My paper does not follow any previous work closely but it shares elements with different branches of literature. The modeling of each period is similar to recent models of social learning by Acemoglu et al. (2011) and Lobel and Sadler (2015) in that individuals play sequentially, after observing a random subset of others' actions. My model additionally generates randomness in the order of play, tying both the realized order of play and observation sets to the underlying network structure. Moreover, their papers focus on late decision makers in large populations, as standard in the social learning literature, while my paper concerns the quality of information aggregation of every individual.

The idea of exploiting the stationarity of the objective environment across periods to learn other individuals' persistent types is shared with Sethi and Yildiz (2016, 2019), who study advice-seeking networks where advice reveals information both about the current state of the world and about the perspective of the advice giver. In their papers, confoundedness of sources is absent and advice-seeking is an active choice. In contrast, individuals in my paper receives information passively on exogenous networks, but their types as perceived by others arise endogenously from how they learn and aggregate information.

My paper draws a connection between the network literature and the literature on learning in games, where learning rules are often motivated by stationary problems. Specifically, the class of empiricist learning rules defined in my paper relates to Fudenberg and Kreps (1993) in the idea that the long-run belief about a stationary distribution should be concentrated on its empirical frequency. Empiricist learning rules take a further step. They use properties that hold in all permissible network environments to back out the conditional distributions of observables from the unconditional distribution.

Lastly, the motivation from the empirical finding that network knowledge is local is shared with Li and Tan (2020). They study a model of misspecified learning, where each individual believes that the underlying network is only a subgraph including her neighbors and herself. In contrast, I take a robustness approach, constructing learning rules that work well when individuals acknowledge their uncertainty about the environment.

The rest of the paper is structured as follows. Section 2 describes the model. Section 3 defines empiricist learning rules and proves their asymptotic individual optimality in stationary environments (Proposition 1). Section 4 shows the convergence result (Theorem 1) and sketches out directions for extensions. Section 5 shows strong efficiency on clique trees (Theorem 2). Section 6 explores the challenges to achieving efficiency of information aggregation in general circumstances, including a formal converse to the positive efficiency result (Theorem 3). Section 7 reviews related literature and Section 8 concludes.

2 Model

My model features asynchronous arrivals of private signals and asynchronous actions within each period, when a new state of the world is drawn. In each period, an individual's observation updates both her belief about the underlying environment and her belief about the current state, the latter of which determines her action. This section focuses on clarifying the elements of an environment.

An environment consists of an objective environment, described in Subsection 2.1, and a profile of belief-updating rules and decision rules, described in Subsection 2.3. That is, belief-updating rules and decision rules are treated as primitives of the learning environment rather than strategic choices. Subsection 2.2 provides details on the timing of actions within each period. Subsection 2.4 defines convergence of beliefs and of play.

2.1 Objective environment

Fix an undirected network G. Without loss of generality, assume that G is connected. Denote by $N = \{1, ..., n\}$ the set of individuals on this network and by N_i the set of neighbors of individual i.

There are infinitely many periods, each of which has R rounds. In period t, a state of the world θ_t is drawn from $\{-1, 1\}$ with $\Pr(\theta_t = -1) = \Pr(\theta_t = 1) = 1/2$. Denote by $s_{1t}, ..., s_{nt}$ the private signals of individuals $1, ..., n$ in period t. Assume that these private signals are independent conditional on the realized state of the world, with the private signal of each individual i having conditional distribution $\Pr(s_{it} = \theta_t | \theta_t) = q_i$ for all $\theta_t \in \{-1, 1\}$. Assume that $q_i > 1/2$ so that private signals are informative. Refer to q_i as the quality of i's private signal. Private signals arrive to individuals at different rounds according to a signal-timing vector $\tau_t = (\tau_{1t}, ..., \tau_{nt})$ drawn from some distribution $H \in \Delta(\{1, ..., R, \infty\}^n)$, independently over time and independently of the state of the world; $\tau_{it} = r \in \{1, ..., R\}$ means that in period t a private signal arrives to i in round r and $\tau_{it} = \infty$ means that no private signal is generated for i that period. Assume that H has full support and the arrivals of private signals are independent across individuals, that is, $H = H_1 \times ... \times H_n$ where $H_i \in \Delta(\{1, ..., R, \infty\})$ for each i.

In each period t, individual i decides on an action $a_{it} \in \{-1, 1\}$.[2] In the context of social platforms, an action could be a post expressing one's view on the topic of that period and individuals want to express the correct view. Assume that an individual takes action at the end of the round in which she first receives some private signal or observes some actions of her neighbors (or both). Moreover, it takes one round for one's action to be observable to her neighbors. Assume further that once an individual chooses an action in a period, she does not pay attention to any private signal or actions chosen by her neighbors in later rounds of that period. As in the social learning literature, this assumption prevents the same piece

[2] For $M \subseteq N$, write $a_{M,t}$ for the profile of actions chosen in period t by individuals in M. Write a_t for the profile of actions of all individuals in period t.

of information spreading from one individual to her neighbors and then spreading back to her.[3]

A network G, a distribution H of the signal-timing vector and a vector $(q_i)_{i \in N}$ of signal qualities constitute an *objective environment*.

2.2 Timeline in each period

For a given network G, the diffusion of information and timing of actions in each period t are determined solely by the realized signal-timing vector τ_t. This vector induces for each i an *action round* $r_i(\tau_t)$ and an *observation set* $M_i(\tau_t)$. At the beginning of round $r_i(\tau_t)$, i first receives a private signal or observes the actions of some neighbors of hers. The observation set $M_i(\tau_t)$ is the set of individuals whom i hears from in round $r_i(\tau_t)$, including herself if she has received a private signal. At the end of round $r_i(\tau_t)$, individual i chooses an action.

Let $l(i, j)$ denote the distance between individuals i and j on the network. Given signal-timing vector τ, individual i's action round is

$$r_i(\tau) = \min\left\{\min_{j \in N}\{\tau_j + l(i, j)\}, R\right\}.$$

This reflects the assumption that each individual reacts to the first piece of information she receives and that it takes one round for information to travel across one link. If an individual does not receive any information by the last round, she will take action in the last round. The observation set of i is defined by

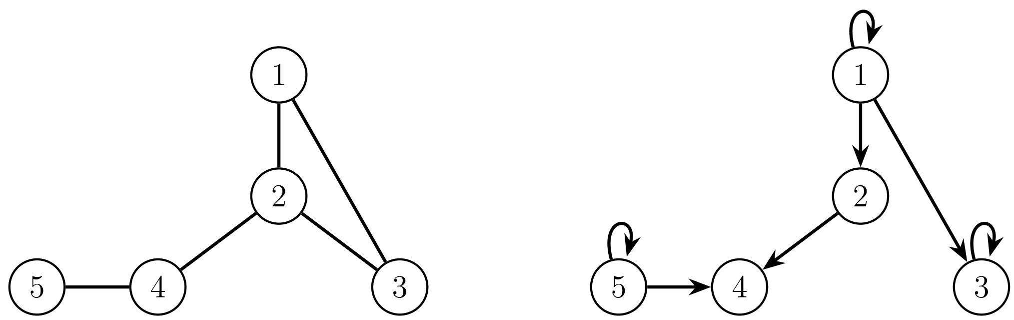
$$M_i(\tau) = (N_i \cap \{j : r_j(\tau) = r_i(\tau) - 1\}) \cup \{i\}^{\mathbf{1}\{r_i(\tau)=\tau_i\}}.$$

Figure 1 illustrates how a signal-timing vector determines the diffusion of information and timing of actions on one particular network. In the left panel is the undirected network and in the right panel is the diffusion of information on this network for $\tau = (1, 3, 2, 5, 2)$. In the latter, an arrow from $j \in N_i$ to i means that i observes j's action and an arrow from i to herself means that she receives her private signal. The action rounds are $r_1(\tau) = 1, r_2(\tau) = r_3(\tau) = r_5(\tau) = 2, r_4(\tau) = 3$ and the observation sets are $M_1(\tau) = \{1\}, M_2(\tau) = \{1\}, M_3(\tau) = \{1, 3\}, M_4(\tau) = \{2, 5\}, M_5(\tau) = \{5\}$. In round 1, individual 1 observes her private signal and takes action. Her action takes one round to reach individuals 2 and 3, the latter of which receives a private signal in round 2. In round 2, individual 2 chooses an

[3] If feedback and updates of actions were instead allowed, then an individual might benefit from distorting her initial decision in ways that improve the quality of information she would receive before making the final decision.

action based only on individual 1's action while individual 3 aggregates both the action of individual 1 and his private signal. Individual 5 observes her private signal in round 2 and chooses an action. In round 3, individual 4 sees the actions of both individuals 2 and 5 and chooses an action before seeing any private signal.

Figure 1: A realized diffusion with diffusion vector $\tau = (1, 3, 2, 5, 2)$



2.3 Belief-updating rules and decision rules

Recall that each individual is assumed to react to the first piece of information she receives in a period and not to pay attention to information arriving after she has chosen an action that period. Thus, an individual's observation at the beginning of her action round in period t includes the actions of some subset of her neighbors and her private signal, and the action round itself. Denote by h_{it} what i observes in period t before choosing an action. That is, $h_{it} = (a_{M_i(\tau_t)\setminus\{i\},t}, s_{it}, r_i(\tau_t))$ if $i \in M_i(\tau_t)$ and $h_{it} = (a_{M_i(\tau_t),t}, r_i(\tau_t))$ if $i \notin M_i(\tau_t)$. Her observation in period t after taking action that period, denoted by $\tilde{h}_{it}$, further includes the action taken. Let $\tilde{h}_i^t = (\tilde{h}_{i1}, ..., \tilde{h}_{it})$.

A *belief-updating rule* of i is a function $\beta_i : (\tilde{h}_i^{t-1}, h_{it}) \mapsto b_i \in \Delta(\{-1, 1\})$ that keeps track of i's belief about the current state of the world at the beginning of her action round each period. A *stationary belief-updating rule* $\bar{\beta}_i : h_{it} \mapsto b_i \in \Delta(\{-1, 1\})$ forms individual i's belief about the current state of the world only from her current observation.

A *decision rule* is a time-dependent function $\sigma_i : (b_i, t) \mapsto d_i \in \Delta(\{-1, 1\})$ that maps i's belief $b_i \in \Delta(\{-1, 1\})$ about the current state of the world to a distribution over feasible actions. A *stationary decision rule* is a time-independent function $\bar{\sigma}_i : b_i \mapsto d_i \in \Delta(\{-1, 1\})$. A natural candidate for consideration is the stationary decision rule that maximizes the subjective probability that i's action matches the current state of the world and breaks ties equally between the two states,

$$\bar{\sigma}_i^*(b_i) = \begin{cases} \delta_1 & \text{if } b_i(1) > 1/2, \\ \delta_{-1} & \text{if } b_i(-1) > 1/2, \\ (1/2, 1/2) & \text{otherwise.} \end{cases}$$

Here δ is the Dirac measure. Notice that this decision rule has a discontinuity at the point where belief splits exactly $(1/2, 1/2)$.

To ensure that convergence of beliefs leads to convergence of play, I consider *smooth decision rules*, which are technical modifications of $\bar{\sigma}_i^*$. A decision rule σ_i^η is a smooth decision rule if

$$\sigma_i^\eta(b_i, t)(\theta) = \frac{\exp\left(b_i(\theta)/\eta_{it}\right)}{\exp\left(b_i(\theta)/\eta_{it}\right) + \exp\left(b_i(-\theta)/\eta_{it}\right)}.$$

for some sequence $\eta = (\eta_{it})_t$, where $\eta_{it} \to 0$ as $t \to \infty$, and for all $\theta \in \{-1, 1\}$. That is, σ_i^η can be derived as the choice probability of a random utility model where the mean utility is the subjective belief and the logistic error decays with time. Since σ_i^η converges to $\bar{\sigma}_i^*$, smooth decision rules asymptotically optimize the probability of matching an individual's action to the true state of the world if the beliefs are correct. While letting the smoothing parameter decay to zero with time is not important for convergence, this approach allows cleaner statements on asymptotic efficiency than the alternative of keeping a constant smoothing parameter.[4]

An *environment* is a tuple $E = (G, H, (q_i)_{i\in N}, (\beta_i)_{i\in N}, (\sigma_i)_{i\in N})$ that combines an objective environment, with a profile of belief updating rules and decision rules.

2.4 Notions of convergence

To separate the stochasticity of play induced by the objective environment from the implications of different belief-updating rules and decision rules, I define the notion of system states. Fix an objective environment $(G, H, (q_i)_{i\in N})$. For each period t, let $\zeta_t = (\zeta_{it})_{i\in N}$ be a vector of independent standard uniform random variables, which are used instrumentally to capture the randomness in actions induced by mixing decisions. Define the *system state* at time t as the tuple $\omega_t = (\tau_t, s_t, \zeta_t)$, which includes a signal-timing vector, a vector of private signals, and a vector of instrumental draws. Note that system states are drawn independently over time from a stationary distribution, so any nonstationarity in the environment must be

[4]Smooth decision rules can also be motivated by the behavior of an individual with smooth ambiguity aversion who faces a decision problem repeatedly and whose uncertainty about the problem gets resolved over time. Such an individual tends to hedge, but her hedging tendency goes away with time. Battigalli et al. (2019) generate decaying hedging tendency from Bayesian learners with smooth ambiguity aversion. For the characterization of smooth ambiguity aversion, see Klibanoff et al. (2005).

induced by the belief-updating rules and decision rules.

For a given objective environment, a profile $(\beta_i)_{i\in N}$ of belief-updating rules and a profile $(\sigma_i)_{i\in N}$ of decision rules define a history-dependent function $\psi : (\omega_t, \tilde{h}^{t-1}, t) \mapsto \{-1,1\}^n$ that, for each given history $\tilde{h}^{t-1}$, maps the current system state to a profile of actions. Refer to such function as a *function of play* and each component function $\psi_i : (\omega_t, \tilde{h}^{t-1}, t) \mapsto \{-1,1\}$ as a *function of i's play.* Take a period t and a history $\tilde{h}^{t-1}$, the action profile $a_t = \psi(\omega_t, \tilde{h}^{t-1}, t)$ is constructed as following:

- Let $r_0 = \min\{\tau_{1t}, .., \tau_{nt}\}$. For all i such that $\tau_{it} = r_0$, $h_{it} = (s_{it}, r_0)$. Set $a_{it} = -1$ if $\zeta_{it} \leq \sigma_i(\beta_i(\tilde{h}_i^{t-1}, h_{it}))(-1)$, otherwise set $a_{it} = 1$.
- Inductively, for $r \in \{r_0+1, ..., R\}$ and i such that $r_i(\tau_t) = r$, $h_{it} = (a_{M_i(\tau_t)\setminus\{i\},t}, s_{it}, r)$ if $\tau_{it} = r$ and $h_{it} = (a_{M_i(\tau_t),t}, r)$ if $\tau_{it} > r$. Set $a_{it} = -1$ if $\zeta_{it} \leq \sigma_i(\beta_i(\tilde{h}_i^{t-1}, h_{it}))(-1)$, otherwise set $a_{it} = 1$.

A *stationary function of play* is a history-independent function $\bar{\psi} : \omega \mapsto a \in \{-1,1\}^n$, mapping each system state to a profile of actions.

Write $\theta^t = (\theta_1, ..., \theta_t)$ for the sequence of states of the world up to t and write $\omega^t = (\omega_1, ..., \omega_t)$ for the sequence of system states up to t. The objective environment of an environment E defines a probability space on the set of all $(\theta^\infty, \omega^\infty)$. Denote by $\Pr^E$ the corresponding probability measure.

An important question when studying a learning rule is whether, or under which conditions, it leads to convergence of beliefs and of play. This paper focuses on almost sure convergence with respect to the true underlying environment.

Definition 1. (Convergence) Fix an environment $E = (G, H, (q_i)_{i\in N}, (\beta_i)_{i\in N}, (\sigma_i)_{i\in N})$. A belief-updating rule β_i converges to a stationary belief-updating rule $\bar{\beta}_i$ if

$$\Pr^E\left(\lim_{t\to\infty} \|\beta_i(\tilde{h}_i^{t-1}, h_{it}) - \bar{\beta}_i(h_{it})\| = 0\right) = 1.$$

The induced function of play ψ converges to a stationary function of play $\bar{\psi}$ if

$$\Pr^E\left(\lim_{t\to\infty} \|\psi(\omega_t, \tilde{h}^{t-1}, t) - \bar{\psi}(\omega_t)\| = 0\right) = 1.$$

3 Empiricist learning rules

This section builds the class of empiricist learning rules and proves their desirable properties in stationary environments. Subsection 3.1 defines learning rules. Subsection 3.2 derives key properties of stationary environments. Subsection 3.3 defines empiricist learning rules and shows that in stationary environments, empiricist learning rules ensure asymptotic learning of relevant elements of the environment and asymptotic individual optimality.

3.1 Learning rules

Over time, individuals update their beliefs about the environment based on what they have observed. To capture the idea that individuals have minimal knowledge of the underlying environment, I assume that each individual is certain only about the set of her neighbors, the quality of her private signal, her own belief-updating rule and decision rule. For each i, denote by $\mathcal{E}_i \subseteq \mathcal{E}$ the set of environments that are consistent with the true tuple $(N_i, q_i, \beta_i, \sigma_i)$ of the underlying environment.

Definition 2. (Learning rules) A *learning rule* $\Gamma_i : (\tilde{h}_i^{t-1}, h_{it}) \mapsto \tilde{p} \in \Delta(\mathcal{E}_i)$ of individual i is a map from i's histories to i's beliefs over environments she deems feasible. Belief-updating rule β_i is founded by learning rule Γ_i if for all history $(\tilde{h}_i^{t-1}, h_{it})$ and all θ,

$$\beta_i(\tilde{h}_i^{t-1}, h_{it})(\theta) = \int_{E \in \mathcal{E}_i} \left(\frac{\Pr^E(h_{it}|\theta_t = \theta, \tilde{h}_i^{t-1})}{\Pr^E(h_{it}|\theta_t = \theta, \tilde{h}_i^{t-1}) + \Pr^E(h_{it}|\theta_t = -\theta, \tilde{h}_i^{t-1})} \right) d\Gamma_i(\tilde{h}_t^{t-1}, h_{it})(E).$$

Note that inside the integral is the likelihood that the current state is θ conditional on i observing $(\tilde{h}_i^{t-1}, h_{it})$ in environment E. Intuitively, with a learning rule, an individual updates her belief about the underlying environment and then updates her belief about the current state using Bayes' rule.

3.2 Local models for stationary environments

In stationary environments, the key object of learning simplifies substantially. It reduces to a collection of conditional distributions of observations at each pair of an observation set and an action round. This collection is referred to as a local model. I then show that there is a one-to-one map between an individual's local model and the unconditional distribution of her observation, further reducing the object of learning to this unconditional distribution.

Definition 3. (Stationary environments) An environment $(G, H, (q_i)_{i\in N}, (\beta_i)_{i\in N}, (\sigma_i)_{i\in N})$ is stationary for individual i if for all $j \neq i$, there exists a stationary belief-updating rule $\bar{\beta}_j$ such that $\beta_j(\tilde{h}_j^{t-1}, h_{jt}) = \bar{\beta}_j(h_{jt})$ for all history $(\tilde{h}_j^{t-1}, h_{jt})$, and a stationary decision rule $\bar{\sigma}_j$ such that $\sigma_j(b_j, t) = \bar{\sigma}_j(b_j)$ for all t and $b_j \in \Delta(\{-1, 1\})$. That is, the belief-updating rules of other individuals depend only on their current observations and their decision rules do not depend on time.

Denote by $\bar{\mathcal{E}}_i \subseteq \mathcal{E}_i$ the set of all environments stationary for i. It holds for all $E \in \bar{\mathcal{E}}_i$, histories $(\tilde{h}_i^{t-1}, h_{it})$ and θ that

$$\Pr^E(h_{it}|\theta_t = \theta, \tilde{h}_i^{t-1}) = \Pr^E(h_{it}|\theta_t = \theta).$$

Thus, the likelihood that the current state is θ conditional on histories $(\tilde{h}_i^{t-1}, h_{it})$ is

$$\begin{aligned}&\frac{\Pr^E(h_{it}|\theta_t = \theta)}{\Pr^E(h_{it}|\theta_t = \theta) + \Pr^E(h_{it}|\theta_t = -\theta)}\\ =&\frac{\Pr^E(h_{it}|\theta_t = \theta, M_i(\tau_t) = M_i, r_i(\tau_t) = r)}{\Pr^E(h_{it}|\theta_t = \theta, M_i(\tau_t) = M_i, r_i(\tau_t) = r) + \Pr^E(h_{it}|\theta_t = -\theta, M_i(\tau_t) = M_i, r_i(\tau_t) = r)},\end{aligned}$$

where $M_i \subseteq N_i \cup \{i\}$ is i's observation set and r is i's action round when she observes h_{it}. The equality follows because the signal-timing vector is independent of the state of the world. The above equation means that for i to form her belief about the current state of the world, she only need to specify her belief about the stationary distribution of her observation conditional on the state of the world at each observation set and action round.

For each stationary environment $E \in \bar{\mathcal{E}}_i$, each nonempty set $M_i \subseteq N_i$, and each $r \in \{2, ..., R\}$, denote by $f_i^E(a_{M_i}, s_i|\theta, M_i \cup \{i\}, r)$ the stationary probability that i observes (a_{M_i}, s_i) given observation set $M_i \cup \{i\}$ and action round r, conditional on the state of the world being θ. Similarly, write $f_i^E(a_{M_i}|\theta, M_i, r)$ for the probability that i observes a_{M_i} given observation set M_i and action round r, conditional on the state of the world being θ. Refer to the tuple $f_i^E = (f_i^E(.|\theta, M_i \cup \{i\}, r), f_i^E(.|\theta, M_i, r))_{\theta\in\{-1,1\}, M_i\subseteq N_i, r\in\{2,...,R\}}$ as i's local model in E. Appendix A.1 shows that f_i^E is well defined.[5]

With some abuse of notation, write

$$f_i^E(a_{M_i}|\theta, M_i \cup \{i\}, r) = f_i^E(a_{M_i}, 1|\theta, M_i \cup \{i\}, r) + f_i^E(a_{M_i}, -1|\theta, M_i \cup \{i\}, r)$$

for the conditional probability that i observes a_{M_i} from her neighbors when she receives some private signal. The following lemma proves two properties that hold in all stationary learning environments.

Lemma 1. *Take any $i \in N$ and any $E \in \bar{\mathcal{E}}_i$. The following properties hold:*

[5] Note also that the conditional distribution of an individual's observation when she receives only her private signal is known by her. For ease of notation, it is thus excluded from her local model.

1) For every nonempty $M_i \subseteq N_i$, $r \in \{2, ..., R\}$ *and* $a_{M_i} \in \{-1, 1\}^{|M_i|}$,

$$\begin{pmatrix} f_i^E(a_{M_i}, -1|M_i \cup \{i\}, r) \\ f_i^E(a_{M_i}, 1|M_i \cup \{i\}, r) \end{pmatrix} = \begin{pmatrix} \Pr^E(s_i = -1, \theta = -1) & \Pr^E(s_i = -1, \theta = 1) \\ \Pr^E(s_i = 1, \theta = -1) & \Pr^E(s_i = 1, \theta = 1) \end{pmatrix} \times \begin{pmatrix} f_i^E(a_{M_i}|\theta = -1, M_i \cup \{i\}, r) \\ f_i^E(a_{M_i}|\theta = 1, M_i \cup \{i\}, r) \end{pmatrix}.$$

Furthermore, matrix $(\Pr^E(s_i, \theta))_{s_i, \theta \in \{-1,1\}}$ *is full rank.*

2) For every nonempty set $M_i \subseteq N_i, r \in \{2, .., R\}$ *and* $(a_{M_i}, s_i) \in \{-1, 1\}^{|M_i|+1}$,

$$\begin{cases} f_i^E(a_{M_i}|\theta, M_i, r) = f_i^E(a_{M_i}|\theta, M_i \cup \{i\}, r) \\ f_i^E(a_{M_i}, s_i|\theta, M_i \cup \{i\}, r) = \Pr^E(s_i|\theta) f_i^E(a_{M_i}|\theta, M_i \cup \{i\}, r). \end{cases}$$

Proof sketch. The second equation in part 2 holds because private signals are independent of the diffusion process and across individuals, conditional on the state of the world. This implies the system of linear equations in part 1. Moreover, matrix $(\Pr^E(s_i, \theta))_{s_i, \theta \in \{-1,1\}}$ is full rank because i's private signal is informative. For the first equation in part 2, notice that the arrivals of private signals are independent across individuals. This implies that conditional on i acting later than her neighbors, the way information has diffused to her neighbors is not affected by whether i receives her private signal. See Appendix A.2 for the details. □

Since i knows the quality of her informative private signal, part 1 of Lemma 1 implies that the conditional distribution of the neighbor actions observed by i when she receives a private signal can be derived from the unconditional distribution of her observation when she receives a private signal. Then part 2 of the lemma completes a one-to-one mapping between i's local model and the tuple of unconditional distributions of i's observation when she receives a private signal.

Corollary 1. *For every* E *and* $E' \in \bar{\mathcal{E}}_i$, *if the respective local models* f_i^E *and* $f_i^{E'}$ *of individual* i *satisfy that* $f_i^E(.|M_i \cup \{i\}, r) = f_i^{E'}(.|M_i \cup \{i\}, r)$ *for all nonempty set* $M_i \subseteq N_i$ *and* $r \in \{2, ..., R\}$, *then* $f_i^E = f_i^{E'}$.

3.3 Empiricist learning rules

Empiricist learning rules build on two ideas. First, if an individual believes that her learning environment is stationary, she does not need to learn the underlying environment but only the local model it induces. Second, this local model can be inferred from the unconditional distribution of her observation, which in turn could be learned from its empirical counterpart.

Learning about a stationary environment is connected to learning about a local model. Denote by $\mathcal{F}_i = \{f_i^E \text{ for some } E \in \bar{\mathcal{E}}_i\}$ the set of all local models induced by feasible environments that are stationary to i. Let Γ_i be i's learning rule. If $\Gamma_i(\tilde{h}_i^{t-1}, h_{it}) \in \Delta(\bar{\mathcal{E}}_i)$ for all histories $(\tilde{h}_i^{t-1}, h_{it})$, then Γ_i induces a *learning rule over local models*, $\gamma_i : h_i^t \mapsto p \in \Delta(\mathcal{F}_i)$, that satisfies for all $F_i \subseteq \mathcal{F}_i$,

$$\gamma_i(h_i^t)(F_i) = \int_{E \in \bar{\mathcal{E}}_i} \mathbf{1}\{f_i^E \in F_i\} d\Gamma_i(\tilde{h}_i^{t-1}, h_{it})(E).$$

Notice that to update her belief about the true local model, individual i only need to keep track of past observed actions of her neighbors and not her own. The reason is that in stationary environments, an individual's own action has no influence on the actions of other individuals in future periods, and thus does not matter for her learning.

Let β_i be the belief-updating rule founded by such Γ_i. It follows that for all histories $(\tilde{h}_i^{t-1}, h_{it})$ and state of the world θ,

$$\beta_i(\tilde{h}_i^{t-1}, h_{it})(\theta) = \int_{f_i \in \mathcal{F}_i} \left(\frac{f_i(a_{M_i}, s_i|\theta, M_i \cup \{i\}, r)}{f_i(a_{M_i}, s_i|\theta, M_i \cup \{i\}, r) + f_i(a_{M_i}, s_i| -\theta, M_i \cup \{i\}, r)} \right) d\gamma_i(h_i^t)(f_i)$$

if $h_{it} = (a_{M_i}, s_i, r)$, and

$$\beta_i(\tilde{h}_i^{t-1}, h_{it})(\theta) = \int_{f_i \in \mathcal{F}_i} \left(\frac{f_i(a_{M_i}|\theta, M_i, r)}{f_i(a_{M_i}|\theta, M_i, r) + f_i(a_{M_i}| -\theta, M_i, r)} \right) d\gamma_i(h_i^t)(f_i)$$

if $h_{it} = (a_{M_i}, r)$.

Next is to connect an individual's learning of her local model to the empirical distribution of her observation. Denote by $\hat{f}_i(.|M_i \cup \{i\}, r)(h_i^t)$ the empirical distribution of i's observation conditional on observation set $M_i \cup \{i\}$ and action round r given history h_i^t. Formally, for each nonempty set $M_i \subseteq N_i, r \in \{2, ..., R\}$ and $(a_{M_i}, s_i) \in \{-1, 1\}^{|M_i|+1}$, let

$$\hat{f}_i(a_{M_i}, s_i|M_i \cup \{i\}, r)(h_i^t) = \frac{\sum_{t'=1}^t \mathbf{1}\{h_{it'} = (a_{M_i}, s_i, r)\}}{\sum_{t'=1}^t \sum_{(a'_{M_i}, s'_i) \in \{-1,1\}^{|M_i|+1}} \mathbf{1}\{h_{it'} = (a'_{M_i}, s'_i, r)\}}$$

if the denominator is positive, otherwise set $\hat{f}_i(a_{M_i}, s_i|M_i \cup \{i\}, r)(h_i^t) = 1/2^{|M_i|+1}$. For every $\epsilon > 0$ and history h_i^t, define $F_i^\epsilon(h_i^t) \subseteq \mathcal{F}_i$ as the set of all local models f_i such that for all nonempty set $M_i \subseteq N_i$ and $r \in \{2, ..., R\}$,

$$\|f_i(.|M_i \cup \{i\}, r) - \hat{f}_i(.|M_i \cup \{i\}, r)(h_i^t)\| < \epsilon.$$

That is, $F_i^\epsilon(h_i^t)$ is the set of local models that induce unconditional distributions within ϵ-distance from their empirical counterparts.

Along every history, empiricist learning rules put probability one on stationary environments, so learning about the underlying environment reduces to learning about the local model. Furthermore, empiricist learning rules asymptotically put probability one on local models consistent with the empirical distribution of individuals' observations.

Definition 4. (Empiricist learning rules) Learning rule $\Gamma_i : (\tilde{h}_i^{t-1}, h_{it}) \mapsto \tilde{p} \in \Delta(\mathcal{E}_i)$ is an empiricist learning rule if

1) for all history $(\tilde{h}_i^{t-1}, h_{it})$, $\Gamma_i(\tilde{h}_i^{t-1}, h_{it}) \in \Delta(\bar{\mathcal{E}}_i)$,

2) and the map $\gamma_i : h_i^t \mapsto p \in \Delta(\mathcal{F}_i)$ induced by Γ_i satisfies that for all $\epsilon > 0$ and h_i^t,

$$\lim_{t \to \infty} \gamma_i(h_i^t)(F_i^\epsilon(h_i^t)) = 1.$$

A belief-updating rule is empiricist if it is founded by an empiricist learning rule.

Empiricist learning rules have a high-level connection with the literature on fictitious play, in particular, the concept of asymptotically empirical assessment rules by Fudenberg and Kreps (1993). In a fictitious play, each player in a repeated game believes that their opponents employ stationary mixed strategies. Asymptotically empirical assessment rules require that a player's assessment of others' mixed strategies asymptotically agrees with their empirical distribution. Analogously, in my model, individuals that use empiricist learning rules believe that their environments are stationary and that they should eventually take the empirical distribution of observed play as the true stationary play. The difference is that in my model, the distribution of play is not the final object of learning, but how it correlates with the unobserved state of the world. To learn the latter, empiricist learning rules exploit Lemma 1, which connects the distribution of observed neighbors' actions unconditional on the state of the world to that conditional on the state of the world.

Another intuition that carries analogously from fictitious play to my model is that fictitious play would be asymptotically optimal if the opponents indeed used stationary strategies. The reason is that by the Strong Law of Large Numbers, the empirical distribution of a stationary random variable converges to its true stationary distribution almost surely. An individual that asymptotically believed in the empirical distribution would, therefore, asymptotically learn the true stationary distribution of opponents' play and respond optimally. In my model, when the underlying learning environment is stationary, empiricist learning rules ensure asymptotic learning of the true local model and thus, asymptotic optimality. More specifically, part 2 of the following proposition says that conditional on sufficiently long histories, a pair of an empiricist learning rule and a smooth decision rule approximately optimizes the probability that an individual's action matches the true state of the world in the current period and in all future periods. Note that in an environment stationary to i, the distribution over histories h_i^∞ depends only on the objective environment and the profile of belief-updating rules and decision rules of individuals other than i. Since asymptotic learning holds at almost every history h_i^∞, so does asymptotic optimality.

Proposition 1. *Take any $i \in N, E \in \bar{\mathcal{E}}_i$ and let f_i^E be i's local model induced by E. Suppose that i adopts an empiricist learning rule Γ_i, which induces learning rule γ_i of local models and empiricist belief-updating rule β_i. The following hold:*

1) (Asymptotic Learning) For all $\epsilon > 0$,

$$\Pr^E\left(\lim_{t\to\infty} \gamma_i(h_i^t)\left(B^\epsilon\left(f_i^E\right)\right) = 1\right) = 1,$$

where $B^\epsilon\left(f_i^E\right) = \{f_i \in \mathcal{F}_i : \|f_i - f_i^E\| < \epsilon\}$ is the ϵ-ball around f_i^E.

2) (Asymptotic Optimality) For all belief-updating rule β_i', decision rule σ_i' and $\epsilon > 0$,

$$\Pr^E\Big(\exists T : \mathbb{E}^E(\sigma_i^\eta(\beta_i(\tilde{h}_i^{t-1}, h_{it}))(\theta_t)|h_i^t) \geq \mathbb{E}^E(\sigma_i'(\beta_i'(\tilde{h}_i^{t-1}, h_{it}))(\theta_t)|h_i^t) - \epsilon, \forall t > T\Big) = 1,$$

where $\mathbb{E}^E$ is the expectation with respect to probability measure $\Pr^E$.

Proof sketch. In every stationary environment, it holds by the Strong Law of Large Numbers that the empirical distribution of an individual's observation converges to the true unconditional distribution almost surely. Empiricist learning rules, which asymptotically put probability one on local models consistent with the empirical distribution of observations, thus asymptotically put probability one on local models consistent with the true unconditional distribution. By Corollary 1, it follows that beliefs under empiricist learning rules asymptotically concentrate on the true local model. Part 2 follows from part 1 and the asymptotic optimality of smooth decision rules. See Appendix A.3 for the details. □

4 Convergence of learning

Section 3 showed that empiricist learning rules are individually asymptotically optimal in stationary environments. But how do they perform in nonstationary environments, where individuals actively learn? This section shows the first main theorem of my paper: convergence of beliefs and of play is achieved in every environment where all individuals adopt empiricist learning rules and smooth decision rules. Subsection 4.1 sketches the proof of this theorem. Subsection 4.2 discusses how the convergence result can be extended to more general settings.

4.1 Convergence on all networks

When all individuals adopt empiricist learning rules and smooth decision rules, individuals' play converges to a stationary play, where only the current observation affects an individual's belief about the current state of the world. Moreover, each individual asymptotically learns the true local model induced by the long-run play.

Theorem 1. *Fix any environment $E = (G, H, (q_i)_{i \in N}, (\beta_i)_{i \in N}, (\sigma_i^\eta)_{i \in N})$ where for each i, β_i is an empiricist belief-updating rule and σ_i^η is a smooth decision rule. Let γ_i be individual i's learning rule over local models and ψ be the function of play induced by E. The following hold.*

1) *For each $i \in N$, β_i converges to some stationary belief-updating rule $\bar{\beta}_i^*$.*

2) *ψ converges to some stationary function of play $\bar{\psi}^*$.*

3) *Let f_i denote the local model induced by $\bar{\psi}^*$. Then for each $i \in N$, γ_i asymptotically learns f_i. That is, for every ϵ-ball $B^\epsilon(f_i)$ around f_i,*

$$\Pr^E\left(\lim_{t \to \infty} \gamma_i(h_i^t)\left(B^\epsilon(f_i)\right) = 1\right) = 1.$$

Moreover, $\bar{\psi}^*$ is the symmetric and individually optimal stationary function of play that is uniquely defined by the objective environment $(G, H, (q_i)_{i \in N})$. For each i, $\bar{\beta}_i^*$ gives the correct likelihoods of the state of the world at each current observation, according to local model f_i.

Proof sketch. The proof of parts 1 and 2 proceeds inductively on action rounds, showing for each $r \in \{1, ..., R\}$ that for every individual i,

$$\Pr^E\left(\lim_{t \to \infty} \mathbf{1}\{r_i(\tau_t) = r\} \|\beta_i(\tilde{h}_i^{t-1}, h_{it}) - \bar{\beta}_i^*(h_{it})\| = 0\right) = 1$$

and

$$\Pr^E\left(\lim_{t\to\infty} \mathbf{1}\{r_i(\tau_t)=r\}\|\psi_i(\omega_t,\tilde{h}_{t-1},t)-\bar{\psi}_i^*(\omega_t)\|=0\right)=1.$$

Consider the base case with $r=1$ and take an arbitrary i. For every τ_t such that $r_i(\tau_t)=1$, it must hold that $\tau_{it}=1$ and $M_i(\tau_t)=\{i\}$. That is, i must receive her private signal in round 1 and choose an action using only her private signal. For all $\omega=(\tau,s,\zeta)$ such that $r_i(\tau)=1$, construct

$$\bar{\beta}_i^*(s_i,1)(\theta)=q_i^{\mathbf{1}\{s_i=\theta\}}(1-q_i)^{\mathbf{1}\{s_i\neq\theta\}} \text{ for all } \theta\in\{-1,1\}, \quad \text{and} \quad \bar{\psi}_i^*(\omega)=s_i.$$

Because individual i knows the quality of her private signal, $\beta_i(\tilde{h}_i^{t-1},h_{it})=\bar{\beta}_i^*(h_{it})$ for all observation $h_{it}=(s_{it},1)$. Moreover, since $q_i>1/2$, her optimal action at such h_{it} is to choose $a_{it}=s_{it}$. It then follows from the definition of smooth decision rules that ψ_i asymptotically puts probability one on the action that agrees with her private signal when she receives only her private signal. This completes the proof of the base case.

The inductive step from r to $r+1$ relies on showing that as an individual's environment becomes approximately stationary, her belief converges and thus so does her play. Empiricist learning rules ensure convergence of learning of approximately stationary environments, and smooth decision rules ensure that convergence of beliefs leads to convergence of play. This argument is formalized by Lemma A2 in Appendix A.4. Finally, the proof of part 3 extends the proof of part 1 of Proposition 1 to approximately stationary environments. □

Theorem 1 demonstrates a form of learning externalities: if an individual learns the informativeness of her neighbors' actions well, then the informativeness of her action to other neighbors will be learned well by these other neighbors. As a general pattern of learning, convergence spreads from early receivers of news to late receivers of news. This pattern is inherent in the inductive construction of the profile of stationary belief-updating rules $(\bar{\beta}_i^*)_{i\in N}$ and the stationary function of play $\bar{\psi}^*$ (see Appendix A.4).

While the way information diffuses across rounds suggests a natural order for an inductive argument, the inductive step relies crucially on the fact that empiricist learning rules do not allow (mis)learning in later rounds to contaminate learning in earlier rounds. The latter holds because under empiricist learning rules, learning in a round depends only on the empirical frequency of neighbors' actions observed in that round, which in turn depends only on others' learning in earlier rounds. Note that neighbors' play across periods is in fact linked through the fundamentals of the environment, such as the network structure and the informativeness of private signals. By forgoing information from such linkages, empiricist learning rules may slow down learning but they avoid potential failures due to misspecification.

That convergence of play under empiricist learning rules holds in all networks has important implications. First, it lends support to the assumption of empiricist learning rules that individuals believe in stationary learning environments. Second and most importantly, it allows the long-run performance of empiricist learning rules to be evaluated at the stationary play to which play under empiricist learning rules and smooth decision rules converges. This stationary play is the focus of Sections 5 and 6.

Another implication of the general convergence result is that my model can provide a learning foundation for the common knowledge assumptions in a direct analysis of the one-period game. Consider the one-period game of my model where the objective environment is common knowledge and each individual aims to maximize the probability that her action matches the true state of the world. In this game, the unique Perfect Bayesian equilibrium that breaks ties equally between the two states of the world induces the same stationary play as the long-run play in my model when individuals use empiricist learning rules and smooth decision rules.

4.2 Convergence in more general settings

In fact, the convergence result can be generalized to richer settings.

First, the objective environment can allow any finite number of states of the world and general diffusion processes that are not necessarily tied to a permanent underlying network. The crucial assumption to maintain is the stationarity of the objective environment. This assumption ensures that at each inductive step, the local environment becomes approximately stationary as others' play converges.

Second, individuals can have multiple signals with arbitrary correlation, as long as each individual has a private signal that satisfies *independence* and *informativeness*. Here independence means that the arrival and realization of this private signal is independent of other components of the individual's observation set. This allows for a mapping between the unconditional distribution of an individual's observation when she receives this private signal and the corresponding conditional distributions, analogous to property 1 of Lemma 1. The independence of this private signal also allows a direct connection between the conditional distribution of an individual's observation when she receives and when she does not receive this signal, analogous to property 2 of Lemma 1. If the private signal is informative, these two mappings result in a one-to-one mapping between the unconditional distribution of one's observation and her local model. Here, informativeness means that the matrix $(\Pr^E(s_i, \theta))_{s_i,\theta}$ has full column rank.

Finally, as in the main setting, individuals that use empiricist learning rules believe in stationary environments and asymptotically believe in the empirical distribution of their observation. Furthermore, they use asymptotically optimal decision rules that are smoothed to ensure that convergence of beliefs leads to convergence of play. As in the main setting, one way of smoothing is to derive these rules from the choice probabilities of a random utility model that takes subjective beliefs as mean utilities and has decaying logistic errors.

5 Strong efficiency on clique trees

This section builds on the convergence result of the previous section to prove a positive efficiency result: on clique trees, the long-run play induced by empiricist learning rules and smooth decision rules achieves strong efficiency within the class of symmetric feasible stationary plays. That is, no other symmetric feasible stationary play can make some individual strictly better off, even at the expense of some other individuals.

Definition 5. (Strong efficiency) Fix an objective environment, which induces probability measure Pr on (θ, ω). A stationary function of play $\bar{\psi}$ is said to dominate another stationary function of play $\bar{\psi}'$ if for every i,

$$\Pr(\bar{\psi}_i(\omega) = \theta) \geq \Pr(\bar{\psi}'_i(\omega) = \theta).$$

A stationary function of play is **strongly efficient** within a class of stationary functions of play if it belongs to that class and it dominates all stationary functions of play in that class.

Subsection 5.1 defines the class of symmetric feasible stationary functions of play. Subsection 5.2 derives properties of clique trees that are key to the efficiency result, which is shown in Subsection 5.3.

5.1 Symmetric feasible stationary functions of play

Recall from Theorem 1 that if all individuals adopt empiricist learning rules and smooth decision rules, the induced play converges to a stationary function of play $\bar{\psi}^*$ that is well-defined by the underlying objective environment. More specifically, $\bar{\psi}^*$ is associated with the profile $(\bar{\beta}^*_i)_{i \in N}$ of stationary belief-updating rules that correctly learn the local environments and the profile $(\bar{\sigma}^*_i)_{i \in N}$ of optimal stationary decision rules that treat the two states symmetrically.

For efficiency comparison, it is only meaningful to compare $\bar{\psi}^*$ to stationary functions of play that respect local observability. That is, an individual's play can only condition on her observation.

Definition 6. (Feasibility) A stationary function of play $\bar{\psi}$ is feasible if and only if it can be induced by a profile of stationary belief-updating rules $(\bar{\beta}_i)_{i\in N}$ and a profile of stationary decision rules $(\bar{\sigma}_i)_{i\in N}$.

Moreover, to study environments without biases towards either state of the world, one can focus on the class of symmetric stationary functions of play. Since the objective environment is symmetric with respect to the state of the world, play should also be symmetric with respect to the state of the world if individuals care equally about false positives and false negatives.

A stationary belief-updating rule $\bar{\beta}_i$ is *symmetric* if for every observation (a_{M_i}, s_i, r),

$$\bar{\beta}_i(a_{M_i}, s_i, r)(\theta) = \bar{\beta}_i(-a_{M_i}, -s_i, r)(-\theta),$$

and for every observation (a_{M_i}, r) of i,

$$\bar{\beta}_i(a_{M_i}, r)(\theta) = \bar{\beta}_i(-a_{M_i}, r)(-\theta).$$

A stationary decision rule $\bar{\sigma}_i$ is symmetric if for all $b \in \Delta(\{-1, 1\})$, $\bar{\sigma}_i(\mathbf{1} - b) = \mathbf{1} - \bar{\sigma}_i(b)$. That is, a symmetric belief-updating rule flips one's belief between the two states of the world if the observed actions and private signal flip signs; a symmetric decision rule flips the probability of each action when one's belief flips between the two states of the world.

Definition 7. (Symmetry) A feasible stationary function of play $\bar{\psi}$ is symmetric if

$$\bar{\psi}(\tau, s, \zeta) = -\bar{\psi}(\tau, -s, \mathbf{1} - \zeta)$$

for all τ, s and ζ. Equivalently, $\bar{\psi}$ is symmetric if and only if it can be induced by some profile $(\bar{\beta}_i)_{i\in N}$ of symmetric belief-updating rules and some profile $(\bar{\sigma}_i)_{i\in N}$ of symmetric stationary decision rules.[6]

5.2 Clique trees

The main efficiency theorem of this paper establishes that individuals on clique trees have their interests perfectly aligned. This subsection describes clique trees and the two properties of clique trees that are sufficient for strong efficiency.

Definition 8. (Clique trees) A network G is a clique tree if any two individuals i and j that belong to the same cycle are linked.

[6]The equivalence relation is shown in Appendix A.5.

Essentially, clique trees are extensions of trees, where a node can be replaced by a clique, that is, a set of fully linked individuals. Qualitatively, clique trees are stylized examples of networks of different groups with dense within-links and sparse between-links.

Recall that information travels from one individual to another individual only through the shortest paths connecting them. Thus, every feasible stationary function of play $\bar{\psi}$ must satisfy that for every i, $\bar{\psi}_i$ depends on the system state ω only through:

- the signal-timing vector τ,
- the private signal s_j of each $j \in \text{argmin}_{k\in N}\{\tau_k + l(k,i)\}$,
- the instrumental draws ζ_j of each $j \in \text{argmin}_{k\in N}\{r_k(\tau) + l(k,i)\}$ and ζ_i.

This means that the role of the underlying network in information diffusion is exactly summarized by the structure it imposes on the shortest paths connecting individuals. Let $IF_{j\setminus i} = \{k \in N : l(k,j) + l(j,i) = l(k,i)\}$ denote the set of individuals whose shortest connecting paths to i go through j. This notion is useful for establishing two technical properties of clique trees. Note that on geodetic networks, any two individuals are connected through a unique shortest path, and that clique trees are geodetic.

Lemma 2. *Take any connected network G.*

1) If G is geodetic then for all $i \in N$, $\{IF_{j\setminus i}\}_{j\in N_i}$ is a partition of $N\setminus\{i\}$.

2) If G is a clique tree then for all $i \in N$, $j \in N_i$ and $k \in N_j\setminus(N_i \cup \{i\})$, $IF_{k\setminus j} \subseteq IF_{j\setminus i}$.

Proof. See Appendix A.6. □

The above lemma establishes two technical properties of clique trees. First, an individual's neighbors act as separate channels through which signals and actions of other individuals can influence her. Second, sources that are not jointly shared between two neighbors i and j and influence j must also influence i exactly through j. In Lemma 3 and Lemma 4, these two technical properties of clique trees are translated into two properties crucial for establishing strong efficiency. First, the separation of sources that can influence an individual is exploited to show that for every individual on a geodetic network, the observed actions of her neighbors are always independent conditional on the state of the world. As a result, the overall informativeness of these neighbors' actions is increasing in the informativeness of each of their actions, captured by a sufficient statistic under symmetry of play. Then, the relation between the set of those individuals that can influence an individual and the set of those influencing her neighbors is shown to align neighbors' interests. That is, an individual maximizes the informativeness of her action to her neighbors exactly by optimally aggregating information at each of her conditioning sets.

Lemma 3 establishes the conditional independence of observed neighbors' actions. The proof involves two steps, each of which exploits part 1 of Lemma 2. The first step further uses the independence of the realizations of private signals to show that the actions of i's neighbors are independent conditional on the state of the world and the signal-timing vector. The second step further uses the independence of signal arrivals to show that the realized path of information that reaches an individual through each neighbor of hers is independent across these neighbors.

Lemma 3. *(Conditional independence of observed neighbors' actions) Fix an objective environment $(G, H, (q_i)_{i\in N})$ where G is geodetic and a feasible stationary function of play $\bar{\psi}$. For every i, nonempty set $M_i \subseteq N_i \cup \{i\}$ and $r \in \{2, ..., R\}$,*

$$\Pr(\bar{\psi}_{M_i\setminus\{i\}}(\omega)|\theta, M_i(\tau) = M_i, r_i(\tau) = r) = \prod_{j\in M_i\setminus\{i\}} \Pr(\bar{\psi}_j(\omega)|\theta, M_i(\tau) = M_i, r_i(\tau) = r).$$

Proof. See Appendix A.7. □

An implication of Lemma 3 together with symmetry is that for each conditioning set of an individual, what matters for her prediction is a set of sufficient statistics, each capturing the quality of each source that she hears from. Specifically, fix an objective environment where the network is a clique tree and consider a symmetric feasible stationary function of play $\bar{\psi}$. Take an individual i with observation set $M_i \subseteq N_i$ and action round $r \in \{2, ..., R\}$. For each $j \in M_i$, define the quality of j's signal given i's conditioning set being (M_i, r) by

$$\begin{aligned}\tilde{q}_j^{M_i,r} &= \Pr(\bar{\psi}_j(\omega) = -1|\theta = -1, M_i(\tau) = M_i, r_i(\tau) = r)\\ &= \Pr(\bar{\psi}_j(\omega) = 1|\theta = 1, M_i(\tau) = M_i, r_i(\tau) = r).\end{aligned}$$

The second equality follows from the symmetry of $\bar{\psi}$ and the assumption that $\Pr(s|\theta) = \Pr(-s| - \theta)$ for all $s \in \{-1, 1\}^N$. Then the likelihood ratio of the state of the world θ conditional on i observing a_{M_i} in round r is

$$\frac{\mathcal{L}(\theta = 1|a_{M_i}, r)}{\mathcal{L}(\theta = -1|a_{M_i}, r)} = \prod_{j\in M_i}\left(\frac{\tilde{q}_j^{M_i,r}}{1-\tilde{q}_j^{M_i,r}}\right)^{\mathbf{1}\{a_j=1\}-\mathbf{1}\{a_j=-1\}}.$$

When individual i also observes her private signal, so i's observation set is $M_i \cup \{i\}$, the likelihood ratio becomes

$$\frac{\mathcal{L}(\theta = 1|a_{M_i}, s_i, , r)}{\mathcal{L}(\theta = -1|a_{M_i}, s_i, r)} = \left(\frac{q_i}{1-q_i}\right)^{\mathbf{1}\{s_i=1\}-\mathbf{1}\{s_i=-1\}} \prod_{j\in M_i}\left(\frac{\tilde{q}_j^{M_i,r}}{1-\tilde{q}_j^{M_i,r}}\right)^{\mathbf{1}\{a_j=1\}-\mathbf{1}\{a_j=-1\}}.$$

The optimal predictor for individual i is to choose $a_i = 1$ when the likelihood ratio is larger than 1 and to choose $a_i = -1$ otherwise. Given that the different sources that i hears from are conditionally independent, the probability that her optimal predictor matches the state of the world is weakly increasing in the probability that each source matches the state of the world. The reason is that were she to have sources of higher quality, she could always distort their quality downward by adding symmetric independent noise and use the optimal predictor of the lower quality sources. Note that under $\bar{\psi}^*$ induced in the long run by all individuals using empiricist learning rules and smooth decision rules, each individual uses the optimal predictor at each of her conditioning sets. The next lemma shows that on clique trees, when a neighbor j of i optimizes the probability that her action matches the state of the world, she simultaneously maximizes the informativeness of her action to i. This is an important property for selfish learning to be socially efficient.

Lemma 4. *(Local alignment of interests) Fix an objective environment $(G, H, (q_i)_{i \in N})$ where G is a clique tree. Let $\bar{\psi}$ be a symmetric feasible stationary function of play of this objective environment. Take an individual i, a nonempty set $M_i \subseteq N_i$ and an action round $r \in \{2, ..., R\}$. For every $j \in M_i$,*

$$\tilde{q}_j^{M_i,r} = \sum_{M_j \subseteq (N_j \setminus (N_i \cup \{i\})) \cup \{j\}} w_{M_j} \times \Pr(\bar{\psi}_j(\omega) = \theta | \theta, M_j(\tau) = M_j, r_j(\tau) = r - 1),$$

where each $w_{M_j} \geq 0$ depends only on G and H, and $\sum_{M_j \subseteq (N_j \setminus (N_i \cup \{i\})) \cup \{j\}} w_{M_j} = 1$.

Proof. See Appendix A.8. □

The proof of this lemma relies on part 2 of Lemma 2, which ensures that any neighbor j of i, given her observation set and action round, can distinguish between diffusions that affect her but not i and diffusions that reach i through her. Thus, the informativeness of j's action to i can be written as a weighted average of the probabilities that j's action matches the state of the world at different conditioning sets of j. This establishes local alignment of interests.

5.3 Strong efficiency

Theorem 2 shows that under conditional independence of observed neighbors' actions, local alignment of interests implies global alignment of interests. Formally, let $\bar{\psi}^*$ be the long-run play when all individuals in a network adopt empiricist learning rules and smooth decision rules. Strong efficiency on clique trees is established by inductively applying the following implication of Lemma 3 and Lemma 4: for an individual to aggregate information better under an alternative symmetric feasible stationary function of play $\bar{\psi}$ than she does under $\bar{\psi}^*$, some of her neighbors must aggregate information better under $\bar{\psi}$ than they do under $\bar{\psi}^*$.

Theorem 2. *Fix an objective environment $(G, H, (q_i)_{i \in N})$. Let $\bar{\psi}^*$ be the stationary function of play induced in the long run by all individuals adopting empiricist learning rules and smooth decision rules. If G is a clique tree then $\bar{\psi}^*$ is strongly efficient within the class of symmetric feasible stationary functions of play.*

Proof. Suppose to the contrary that there exists some symmetric feasible stationary function of play $\bar{\psi}$ such that for some i,

$$\Pr(\bar{\psi}_i(\omega) = \theta) > \Pr(\bar{\psi}_i^*(\omega) = \theta).$$

Then there must exist some $M_i \subseteq N_i \cup \{i\}$ and $r \in \{1, ..., R\}$ such that

$$\Pr(\bar{\psi}_i(\omega) = \theta | \theta, M_i(\tau) = M_i, r_i(\tau) = r) > \Pr(\bar{\psi}_i^*(\omega) = \theta | \theta, M_i(\tau) = M_i, r_i(\tau) = r).$$

Take r to be the smallest round such that the above inequality holds for some i and some $M_i \subseteq N_i \cup \{i\}$. If $M_i = \{i\}$ then

$$q_i = \Pr(\bar{\psi}_i^*(\omega) = \theta | \theta, M_i(\tau) = \{i\}, r_i(\tau) = r) \geq \Pr(\bar{\psi}_i(\omega) = \theta | \theta, M_i(\tau) = \{i\}, r_i(\tau) = r),$$

that is, the above strict inequality cannot hold. Thus M_i includes some of i's neighbors, which also implies that $r \geq 2$.

Since individuals use the optimal predictor under $\bar{\psi}^*$, by Lemma 3 and symmetry of play, $\bar{\psi}_i$ is strictly more likely to match the state of the world at conditioning set (M_i, r) than does $\bar{\psi}_i^*$ only if the action of some neighbor of i is more informative to i under $\bar{\psi}$ than under $\bar{\psi}^*$. That is, there exists some $j \in M_i \backslash \{i\}$ such that

$$\Pr(\bar{\psi}_j(\omega) = \theta | \theta, M_i(\tau) = M_i, r_i(\tau) = r) > \Pr(\bar{\psi}_j^*(\omega) = \theta | \theta, M_i(\tau) = M_i, r_i(\tau) = r).$$

By Lemma 4, this means that there must exist some $M_j \subseteq (N_j \backslash (N_i \cup \{i\})) \cup \{j\}$ such that

$$\begin{aligned} &\Pr(\bar{\psi}_j(\omega) = \theta | \theta, M_j(\tau) = M_j, r_j(\tau) = r - 1) \\ > &\Pr(\bar{\psi}_j^*(\omega) = \theta | \theta, M_j(\tau) = M_j, r_j(\tau) = r - 1). \end{aligned}$$

This contradicts that r is the smallest round in which someone can be made strictly better off at some conditioning set. □

Theorem 2 identifies a condition on the underlying network that ensures strong efficiency regardless of the timing of signal arrivals and the quality of private signals. On a tree, the informativeness of a neighbor's action is summarized by a sufficient statistic and the overall informativeness of the actions of one's neighbors is increasing in each of these sufficient statistics. Inductively, better information aggregation starting from the leaves of a tree helps information aggregation towards the root of that tree. Since cliques, or complete subgraphs, do not add confoundedness of sources, the above intuition generalizes to clique trees.

6 Challenges to efficiency

This section explains the challenges to generalizing Theorem 2. It includes a formal converse, Theorem 3, showing that on any network that is not a clique tree, strong efficiency fails for some diffusion process and signal structure. Recall from Lemma 3 and Lemma 4 that clique trees possess two key properties. First, observed neighbors' actions are always conditionally independent. Second, an individual optimizes the informativeness of her action to her neighbors exactly by optimizing her information aggregation at each of her conditioning sets. Theorem 3 generalizes the intuition of two examples: one fails the first property of clique trees and not the second; the other fails the second property and not the first. These examples are presented in Subsections 6.1 and 6.2. The converse is presented in Subsection 6.3. Subsection 6.4 explores whether it is the strong notion of efficiency that demands a restrictive class of networks. The short answer is no. The general forces that lead to failure of strong efficiency as illustrated in Subsections 6.1 and 6.2, when coupled with opportunities for favor trading across rounds, can result in failure of Pareto efficiency. Finally, Subsection 6.5 illustrates the implications of asymmetric environments, where asymmetry comes from either the underlying environment or individuals' play.

In each example of this section, $\bar{\psi}^*$ denotes the stationary function of play induced in the long run by all individuals using empiricist learning rules and smooth decision rules.

6.1 Failure of conditional independence of neighbors' actions

The following example shows that when neighbors' actions correlate, it is possible to distort downward the individual informativeness of some neighbor's action to break the correlation in a way that improves the overall informativeness of all neighbors' actions.

Example 1. An objective environment has $G = \{\{1,2\},\{1,3\},\{2,4\},\{3,4\}\}$, $R = 3$, $\Pr(\tau_1 = 1) \approx 1, \Pr(\tau_2 = 2) = \Pr(\tau_2 = \infty) = \Pr(\tau_3 = 2) = \Pr(\tau_3 = \infty) \approx 0.5, \Pr(\tau_4 = 3) \approx 1$, and $(q_1, q_2, q_3, q_4) = (0.5 + \epsilon, 0.75, 0.9, 0.75)$ for $\epsilon > 0$ small.

There exists a symmetric feasible stationary function of play where individual 2 plays a suboptimal strategy that makes individual 4 strictly better off compared to $\bar{\psi}^*$.

Proof. The first panel of Figure 2 plots the network, which is not a clique tree. The second panel illustrates the flow of information when $\tau = (1, \infty, \infty, 3)$, which occurs with probability close to 0.25. The third panel illustrates the flow of information when $\tau \in \{(1, 2, 2, 3), (1, 2, \infty, 3), (1, \infty, 2, 3)\}$, which occurs with probability close to 0.75. In both cases, information flows from individual 1 to individuals 2 and 3 and then to individual 4, who chooses an action based on what she observes from 2 and 3 and her private signal. The difference is that in the second panel neither 2 nor 3 receives their private signal while in the third panel at least one of them does.

Consider $\bar{\psi}^*$ first. Since $q_2 > q_1$, individual 2 passes onto individual 4 her private signal if she receives it, $a_2 = s_2$, and otherwise passes on the action of individual 1, $a_2 = a_1 = s_1$. Similarly individual 3 chooses $a_3 = s_3$ if he receives his private signal and otherwise chooses $a_3 = a_1 = s_1$. Thus, a_2 and a_3 are perfectly correlated when neither individual 2 nor 3 receives their private signal, and a_2 and a_3 are conditionally independent otherwise.

Here optimally, individual 4 chooses $a_4 = s_4$ when $a_2 \neq a_3$ or $a_2 = a_3 = s_4$. The key question is whether individual 4 would follow the consensus actions of individuals 2 and 3 when $a_2 = a_3 \neq s_4$. The informativeness of $a_2 = a_3$ is summarized by the following ratio

$$\begin{aligned}
&\frac{\Pr(a_2 = a_3 = \theta | \theta, M_4(\tau) = \{2, 3, 4\}, r_4(\tau) = 3)}{\Pr(a_2 = a_3 = -\theta | \theta, M_4(\tau) = \{2, 3, 4\}, r_4(\tau) = 3)} \\
\approx &\frac{0.25(0.5 + \epsilon) + 0.25(0.5 + \epsilon)(0.75) + 0.25(0.5 + \epsilon)(0.9) + 0.25(0.75)(0.9)}{0.25(0.5 - \epsilon) + 0.25(0.5 - \epsilon)(0.25) + 0.25(0.5 - \epsilon)(0.1) + 0.25(0.25)(0.1)} \\
\approx &2.86 < 3 = \frac{\Pr(s_4 = \theta | \theta)}{\Pr(s_4 = -\theta | \theta)}.
\end{aligned}$$

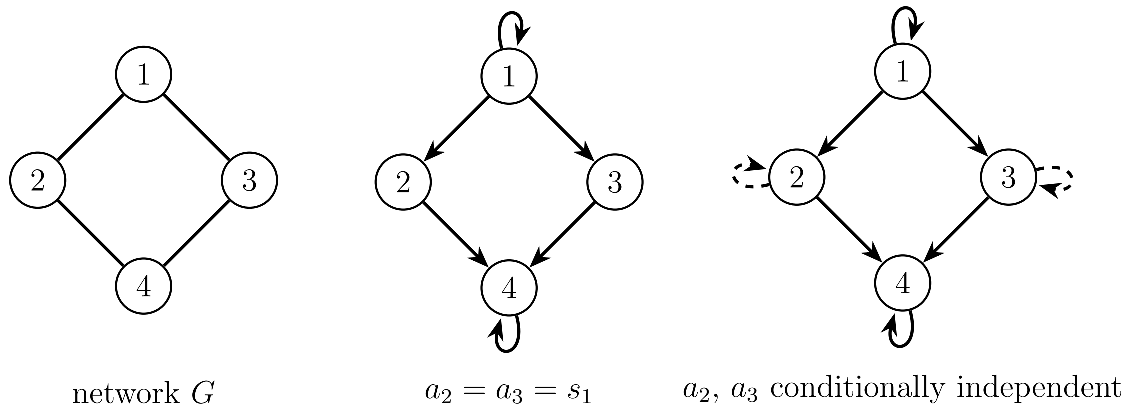
This means that under $\bar{\psi}^*$, individual 4 always chooses $a_4 = s_4$.

Now consider an alternative stationary function of play $\bar{\psi}$ where individual 2 randomizes with equal probabilities between 1 and -1 when she does not receive her private signal. This change breaks the correlation between a_2 and a_3. Then from $\Pr(a_2 = \theta | \theta, M_4(\tau) = \{2, 3, 4\}, r_4(\tau) = 3) \approx 0.625$ and $\Pr(a_3 = \theta | \theta, M_4(\tau) = \{2, 3, 4\}, r_4(\tau) = 3) \approx 0.7 + 0.5\epsilon$,

$$\begin{aligned}
&\frac{\Pr(a_2 = a_3 = \theta | \theta, M_4(\tau) = \{2, 3, 4\}, r_4(\tau) = 3)}{\Pr(a_2 = a_3 = -\theta | \theta, M_4(\tau) = \{2, 3, 4\}, r_4(\tau) = 3)} \\
\approx &\frac{(0.625)(0.7 + 0.5\epsilon)}{(0.375)(0.3 - 0.5\epsilon)} \approx 3.89 > \frac{\Pr(s_4 = \theta | \theta)}{\Pr(s_4 = -\theta | \theta)}.
\end{aligned}$$

This means that when $a_2 = a_3 \neq s_4$, individual 4 is now strictly better-off choosing $a_4 =$

Figure 2: The network and key diffusions in Example 1



$a_2 = a_3$ than choosing $a_4 = s_4$. □

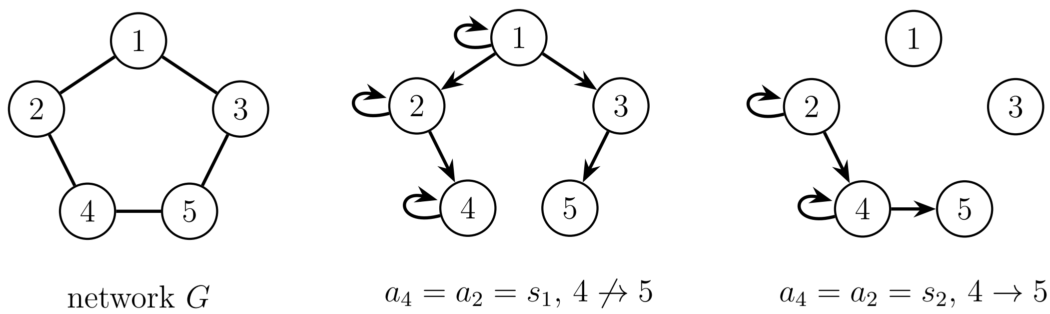
The key intuition of this example is that the correlation between the actions of individuals 2 and 3 when they jointly receive low-quality information from individual 1 downgrades the average informativeness of their consensus actions to individual 4. Note that correlation between different information sources does not hurt learning. Individual 4 correctly learns the correlation structure of the actions of individuals 2 and 3, but it is because she cannot distinguish whether the second panel or the third panel of Figure 2 takes place that she disregards her neighbors' actions altogether. Furthermore, since the private signal of individual 1 is of low quality, the modification made in the alternative function of play hurts individual 2 little but lends noticeable help to individual 4.

Finally, notice that by setting $R = 3$, I ensure that this example does not fail the second property of clique trees. In this example, individual informativeness of each neighbor's action is indeed optimized by the neighbors' selfish learning. Intuitively, failure of the second property requires that an individual hear from some neighbor who himself faces confoundedness of sources, that is, he heard from another neighbor who had heard from some other neighbor. This chain is not possible when $R = 3$. The purpose of this technical trick is to separate the importance of each property, showing that the failure of one property does not imply the failure of the other.

6.2 Failure of local alignment of interests

By local observabiliity, the local model of each individual is averaged over different diffusion realizations. This means that an individual may play optimally to an average distribution that pools some diffusions irrelevant to her neighbors. In other words, the informative-

Figure 3: The network and key diffusions in Example 2



ness of an individual's action to her neighbors is not necessarily maximized by her playing optimally at each of her conditioning sets, as is the case in the following example.

Example 2. An objective environment has $G = \{\{1,2\},\{1,3\},\{2,4\},\{3,5\},\{4,5\}\}$, $R = 4$, $\Pr(\tau_1 = 1) = \Pr(\tau_1 = \infty) \approx 0.5, \Pr(\tau_2 = 2) = \Pr(\tau_4 = 3) \approx 1, \Pr(\tau_3 = \infty) = \Pr(\tau_5 = \infty) \approx 1$, and $(q_1, q_2, q_3, q_4, q_5) = (0.9, 0.5+\epsilon, 0.5+\epsilon, 0.7, 0.7)$.

There exists a symmetric feasible stationary function of play where individual 4 plays a suboptimal strategy that makes individual 5 strictly better off compared to $\bar{\psi}^*$.

Proof. The first panel of Figure 3 plots the network, which is geodetic but not a clique tree. The second and the third panels of this figure illustrate the two main realized diffusions of this objective environment, each occurring with probability approximately 0.5. On the second panel where individual 1 receives his private signal in round 1, individual 5 takes action in round 3 after observing the action of individual 3. On the third panel where individual 1 does not receive his private signal, individual 5 takes action in round 4 after observing the action of individual 4.

Under $\bar{\psi}^*$, individual 2 passes onto individual 4 the private signal of individual 1 when she hears from him and passes on her own private signal otherwise. Thus, individual 4 learns that

$$\Pr(a_2 = \theta | \theta, M_4(\tau) = \{2,4\}, r_4(\tau) = 3) \approx 0.5(0.9) + 0.5(0.5+\epsilon) = 0.7 + 0.5\epsilon > q_4.$$

This means that under $\bar{\psi}^*$, individual 4 chooses $a_4 = a_2$ when he hears from individual 2 and receives a private signal himself in round 3. Note that while individual 4 plays optimally against the distribution of individual 2's action averaged over both cases illustrated in Figure 3, only in the case in the third panel of this figure does individual 4's action matter to individual 5. Thus,

$$\Pr(a_4 = \theta|\theta, M_5 = \{4\}, r_5(\tau) = 4) = \Pr(s_2 = \theta|\theta) = 0.5 + \epsilon.$$

Consider an alternative symmetric feasible stationary function of play $\bar{\psi}$ where individual 4 chooses $a_4 = s_4$ upon hearing from individual 2 in round 3. Then

$$\Pr(a_4 = \theta|\theta, M_5 = \{4\}, r_5(\tau) = 4) = \Pr(s_4 = \theta|\theta) = 0.7.$$

Thus $\bar{\psi}$ makes individual 5 strictly better off, at a (small) expense of individual 4. □

6.3 A converse of Theorem 2

The following theorem generalizes Example 1 and Example 2 to show that on any network that is not a clique tree, strong efficiency fails for some diffusion process and some set of private signals.

Theorem 3. *Suppose that G is not a clique tree. Then there exist a distribution H of the signal-timing vector and a vector of signal qualities $(q_i)_{i\in N}$ such that the stationary function of play $\bar{\psi}^*$ induced in the long-run by all individuals adopting empiricist learning rules and smooth decision rules is not strongly efficient, even within the class of symmetric stationary functions of play.*

Proof sketch. Note that a cycle is incomplete if some individuals in the cycle are not linked. A cycle of length at least four is chordless if any two non-consecutive individuals in the cycle are not linked. There are three cases based on the smallest incomplete cycle G' of G.

Case 1: G' is a chordless cycle of size $2m$ for $m \geq 2$.

Case 2: G' is a chordles cycle of size $2m+1$ for $m \geq 2$.

Case 3: G' is a chordal graph, that is, it does not have any chordless cycle.

The construction of H and $(q_i)_{i\in N}$ for Case 1 is a generalization of Example 1, where the correlation in the actions of two neighbors influenced by a common source makes the individual disregards their consensus actions and breaking such correlation could benefit the individual. The construction for Case 2 is a generalization of Example 2 where an individual follows a neighbor's action that is less informative than her own private signal exactly when her action will be observed by another neighbor. If she used her private signal, the latter neighbor would strictly benefit. In Case 3, it can be shown that G' must be a cycle of size four with exactly one missing link. Then the construction for Case 1 applies. See Appendix A.9 for the details. □

6.4 Pareto improvement from favor trading

This section explores whether it is the strong notion of efficiency that demands a restrictive class of networks. First, I show that a weak form of Pareto efficiency holds on all networks: compared to the long-run play induced by empiricist learning rules and smooth decision rules, there is no symmetric play that strictly improves the probability of correct prediction by an individual without hurting the probability of correct prediction by some other individual in some round. This weak efficiency result does not extend to the standard notion of Pareto efficiency where an individual's welfare is measured by the probability of correct prediction pooling over all action rounds. The reason is that compensations can be made across rounds. An individual may forgo some benefits for her neighbor's sake when she receives information early if her neighbor forgoes some of his benefits for her sake when she receives information late. Examples 3 and 4 build on Examples 1 and 2 and the idea of compensations across rounds to illustrate how Pareto efficiency may not extend much beyond clique trees.

Proposition 2. *Fix an objective environment. Let $\bar{\psi}^*$ be the stationary function of play induced as the long-run play when all individuals on the network adopt empiricist learning rules and smooth decision rules. There does not exist any symmetric feasible stationary function of play $\bar{\psi}$ such that for each $i \in N$ and each $r \in \{1, ..., R\}$,*

$$\Pr(\bar{\psi}_i(\omega) = \theta | r_i(\tau) = r) \geq \Pr(\bar{\psi}^*_i(\omega) = \theta | r_i(\tau) = r),$$

with the inequality holding strictly for some i and some r.

Proof. See Appendix A.10. □

This proposition shows that in environments without biases, failure of Pareto efficiency, as defined below, must be due to some form of favor-trading across rounds.

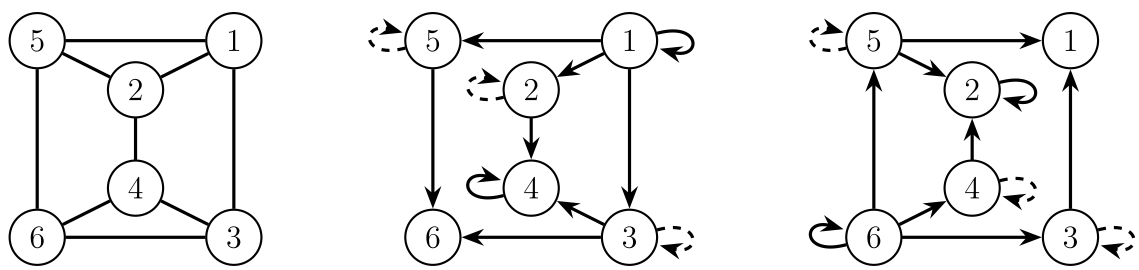
Definition 9. (Pareto efficiency) Fix an objective environment. A feasible stationary function of play $\bar{\psi}$ is **Pareto efficient** within some class of stationary functions of play if it belongs to that class and there does not exist any stationary function of play $\bar{\psi}'$ in that class such that for all i,

$$\Pr(\bar{\psi}'_i(\omega) = \theta) \geq \Pr(\bar{\psi}_i(\omega) = \theta),$$

with a strict inequality holding for some i.

Following I present two examples showing that forces that present challenges to strong efficiency, correlation of neighbors' actions and pooling over diffusions irrelevant to one's

Figure 4: The network and key diffusions in Example 3



neighbors, similarly present challenges to Pareto efficiency when individuals can trade favors across rounds.

Example 3. An objective environment has $G = \{\{1,2\},\{1,3\},\{1,5\},\{2,4\},\{2,5\},$ $\{3,4\},\{3,6\},\{4,6\},\{5,6\}\}$, $R = 3$, $\Pr(\tau_1 = 1) = \Pr(\tau_1 = \infty) = \Pr(\tau_6 = 1) = \Pr(\tau_6 = \infty) \approx 0.5, \Pr(\tau_2 = 2) = \Pr(\tau_2 = 3) = \Pr(\tau_4 = 2) = \Pr(\tau_4 = 3) \approx 0.5, \Pr(\tau_3 = 2) = \Pr(\tau_3 = \infty) = \Pr(\tau_5 = 2) = \Pr(\tau_5 = \infty) \approx 0.5$, and $(q_1, q_2, q_3, q_4, q_5, q_6) = (0.5+\epsilon, 0.75, 0.9, 0.75, 0.9, 0.5+\epsilon)$.

Compared to $\bar{\psi}^*$, there exists an alternative symmetric stationary function of play where individuals 2 and 4 play suboptimally in round 2 to make the other strictly better off in round 3 in a way that both of them are overall better off.

Proof. The first panel of Figure 4 presents the network and the next two panels illustrate the two key diffusions where an opportunity for mutually beneficial favor exchange exists for individuals 2 and 4. The second panel illustrates the case when $\tau_1 = 1, \tau_6 = \infty$ and $\tau_4 = 3$, which occurs with probability close to 0.125. It replicates the flow of information in Example 1, where under $\bar{\psi}^*$ the induced correlation between the actions of individuals 2 and 3 causes individual 4 to downgrade their consensus actions and disregard their actions altogether. The third panel is a mirrored case of the second panel, where individual 6 plays the role of individual 1, individual 5 plays the role of individual 3 and individuals 2 and 4 swap roles. This occurs with probability close to 0.125, when $\tau_6 = 1, \tau_1 = \infty, \tau_2 = 3$.

Similar to Example 1, individual 2 can make individual 4 strictly better off in round 3 by randomizing between 1 and -1 with equal probabilities when she only hears from individual 1 in round 2. Similarly, individual 4 can make individual 2 strictly better off in round 3 by randomizing between 1 and -1 with equal probabilities when he only hears from individual 6 in round 2. The network structure and the diffusion process are chosen so that such changes do not affect any other individuals. Moreover, as ϵ gets closer to zero, the losses in round

2 incurred by individuals 2 and 4 for randomizing rather than respectively passing on the action of individual 1 and the action of individual 6 shrink to zero. However, their gains from breaking the correlations remain. This means that for sufficiently small ϵ, such changes in the play of individuals 2 and 4 lead to a Pareto improvement over $\bar{\psi}^*$. □

Example 4. An objective environment has $G = \{\{1,2\},\{1,3\},\{2,4\},\{3,5\},\{4,5\}\}$, $R = 4$, $\Pr(\tau_1 = 1) \approx 1/3, \Pr(\tau_1 = \infty) \approx 2/3, \Pr(\tau_2 = 2) = \Pr(\tau_2 = \infty) = \Pr(\tau_3 = 2) = \Pr(\tau_3 = \infty) \approx 0.5, \Pr(\tau_4 = 3) = \Pr(\tau_4 = \infty) = \Pr(\tau_5 = 3) = \Pr(\tau_5 = \infty) \approx 0.5$ and $(q_1, q_2, q_3, q_4, q_5) = (0.9, 0.5 + \epsilon, 0.5 + \epsilon, 0.7, 0.7)$.

Compared to $\bar{\psi}^*$, there exists an alternative symmetric stationary function of play where individuals 4 and 5 play suboptimally in round 3 to make the other strictly better off in round 4 in a way that both of them are overall better off.

Proof. The network is a cycle of length five, as in Example 2. Figure 5 illustrates the key diffusions. The first panel occurs with probability close to 1/3, when $\tau_1 = 1$. In this case both individuals 4 and 5 take action in round 3 without observing the action of the other. In the second panel, information flows from individual 2 to individual 4 and then to individual 5 as in Example 2. In the third panel, information flows from individual 3 to individual 5 and then to individual 4.

The distribution H is chosen so that under $\bar{\psi}^*$,

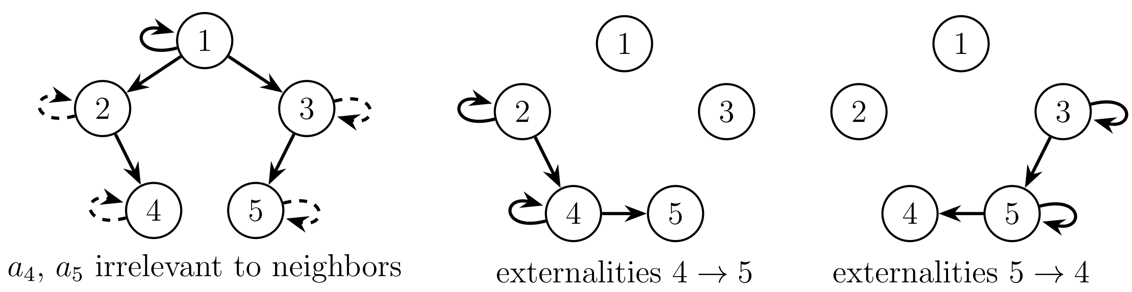
$$\Pr(a_2 = \theta | \theta, M_4(\tau) = \{2,4\}, r_4(\tau) = 3) \approx 0.5q_1 + 0.5q_2 = 0.7 + 0.5\epsilon > q_4,$$

which means that $a_4 = a_2$ when $M_4(\tau) = \{2,4\}$ and $r_4(\tau) = 3$. If instead individual 4 chooses $a_4 = s_4$ at this conditioning set, individual 5 will do strictly better in the second panel of Figure 5. Symmetrically, under $\bar{\psi}^*$, $a_5 = a_3$ when $M_5(\tau) = \{3,5\}$ and $r_5(\tau) = 3$. If instead individual 5 chooses $a_5 = s_5$ at this conditioning set, then individual 4 will do strictly better in the third panel of Figure 5. The losses to individuals 4 and 5 in round 3 vanish as ϵ gets close to zero while their gains in round 4 do not. Moreover, such changes have no effects on other individuals. This completes the proof that a symmetric feasible stationary function of play that adopts these changes is a Pareto improvement over $\bar{\psi}^*$. □

6.5 Asymmetric environments

In general, the overall informativeness of multiple sources of information depends on the quality of each source, their correlation, and their biases towards either false positives or false negatives. Intuitively speaking, the restriction of Theorem 2 to clique trees shuts down the

Figure 5: Key diffusions in Example 4



second channel of externalities and the restriction to symmetric environments shuts down the third channel. The following two examples illustrate how perfect alignment of interests is hard to achieve in asymmetric environments, even on special networks.

First is an example that shows how strong efficiency could fail even on the simplest tree if the two states are asymmetric.

Example 5. Consider a setting where $\Pr(\theta = 1) = 3/4$ and $\Pr(\theta = -1) = 1/4$. An objective environment $(G, H, (q_i)_{i \in N})$ has $G = \{\{1, 2\}\}$, $R = 2$, $\Pr(\tau_1 = 1) = \Pr(\tau_2 = 1) = \Pr(\tau_1 = 2) = \Pr(\tau_2 = 2) \approx 0.5$ and $(q_1, q_2) = (2/3, 2/3)$. Then there does not exist a feasible stationary function of play that is strongly efficient within the class of feasible stationary functions of play.

Proof. First, notice that one negative signal is not sufficient to overturn the positive state, since

$$\frac{\Pr(\theta = -1 | s_1 = -1)}{\Pr(\theta = 1 | s_1 = -1)} = \frac{\Pr(\theta = -1 | s_2 = -1)}{\Pr(\theta = 1 | s_2 = -1)} = \frac{(1/4)(2/3)}{(3/4)(1/3)} = \frac{2}{3} < 1,$$

This means that whoever receives only her own private signal optimally chooses the positive action, regardless of the signal realization. This renders her action completely uninformative to the individual who moves next. As a result, individual optimality implies that $a_1 = a_2 = 1$ regardless of the realized diffusion and signals.

However, two negative signals would indicate that the negative state is more likely,

$$\frac{\Pr(\theta = -1 | s_1 = s_2 = -1)}{\Pr(\theta = 1 | s_1 = s_2 = -1)} = \frac{(1/4)(2/3)(2/3)}{(3/4)(1/3)(1/3)} = \frac{4}{3} > 1.$$

Thus, the second mover would strictly benefit from the first mover reporting her private signal rather than her optimal action. This completes the proof that in every feasible stationary play, there exists an individual that can be made strictly better off at the expense of the other individual. □

Second, even in symmetric objective environments, some individual might benefit from asymmetric play of others, as illustrated by the following example.

Example 6. An objective environment has $G = \{\{1,2\},\{1,3\},\{2,4\},\{3,5\}\}$, $\Pr(\tau_1 = 3) \approx 1, \Pr(\tau_2 = 2) = \Pr(\tau_2 = \infty) = \Pr(\tau_3 = 2) = \Pr(\tau_3 = \infty) \approx 0.5, \Pr(\tau_4 = 1) = \Pr(\tau_5 = 1) \approx 1$, and $(q_1, q_2, q_3, q_4, q_5) = (0.75, 1, 1, 0.5 + \epsilon, 0.5 + \epsilon)$.

There exists a stationary function of play, where the actions of individuals 2 and 3 are biased towards opposite states of the world, that makes individual 1 strictly better off compared to $\bar{\psi}^*$.

Proof. Figure 6 illustrates the diffusions: individuals 4 and 5 receive their private signals in round 1 and pass on their private signals to individuals 2 and 3 respectively, who may or may not receive their private signals in round 2; in round 3, individual 1 receives her private signal and observes the actions of individuals 2 and 3. Since G is a tree, individuals 2 and 3's actions, as observed by individual 1, are always conditionally independent. Under $\bar{\psi}^*$, $a_2 = s_4$ when individual 2 does not receive her private signal and $a_2 = s_2$ when she does. Similarly, $a_3 = s_5$ when individual 3 does not receive his private signal and $a_3 = s_3$ when he does. Overall,

$$\begin{aligned}\Pr(a_2 = \theta|\theta, M_1(\tau) = \{1,2,3\}, r_1(\tau) = 3) &= \Pr(a_3 = \theta|\theta, M_1(\tau) = \{1,2,3\}, r_1(\tau) = 3)\\ &= 0.5(0.5 + \epsilon) + 0.5(1) = 0.75 + 0.5\epsilon.\end{aligned}$$

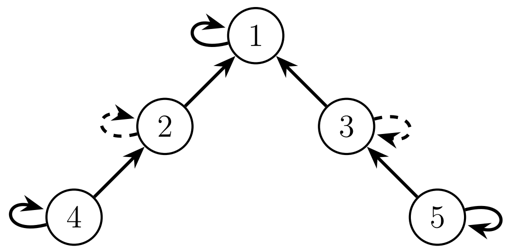
For $\epsilon > 0$ small, the best predictor for individual 1 given (s_1, a_2, a_3) agrees with the majority, thus matching the true state of the world with probability approximately equal to $(3)(0.75)^2(0.25) + (0.75)^3 = \frac{27}{32}$.

Consider an alternative feasible stationary function of play $\bar{\psi}$ where $a_2 = 1$ if individual 2 does not receive her private signal and $a_2 = s_2$ if she does, and $a_3 = -1$ if individual 3 does not receive his private signal and $a_3 = s_3$ if he does. These modifications give

$$\begin{aligned}&\Pr(a_2 = 1|\theta = 1, M_1(\tau) = \{2,3\}, r_1(\tau) = 3) = 1,\\ &\Pr(a_2 = 1|\theta = -1, M_1(\tau) = \{2,3\}, r_1(\tau) = 3) = 0.5,\\ &\Pr(a_3 = 1|\theta = 1, M_1(\tau) = \{2,3\}, r_1(\tau) = 3) = 0.5,\\ &\Pr(a_3 = 1|\theta = -1, M_1(\tau) = \{2,3\}, r_1(\tau) = 3) = 0.\end{aligned}$$

That is, from the perspective of individual 1, a_2 is now conclusive of the negative state and a_3 is now conclusive of the positive state. In this case the best predictor for individual 1 agrees with the conclusive news if there is one and agrees with her private signal if the actions of

Figure 6: Key diffusions in Example 6



individuals 2 and 3 are inconclusive. This best predictor matches the state of the world with probability $0.5 + (0.5)(0.75) = \frac{7}{8} > \frac{27}{32}$. □

This example illustrates that exposure to biases in opposite directions could help with making better predictions. If the biases were instead in the same direction, then an individual who has no intrinsic bias in either state of the world would mechanically appear as biased in that same direction. Note that two key properties of empiricist learning rules, convergence of learning and long-run individual optimality, do not rely on the environments being symmetric. To capture time-invariant biases, one can modify the smooth decision rules accordingly, and given that convergence always holds, one can then focus on studying the stationary function of play induced by everyone correctly learning her local model. While it is potentially interesting to study environments with biases, how they spread or get neutralized, this direction is deferred to future research.

7 Literature review

The majority of papers about learning on networks focus on learning about a single state of the world. My paper belongs to a relatively small group of papers that model varying states across time. Sethi and Yildiz (2016, 2019) study endogenous communication networks with independently drawn states, where plays across periods connect through individuals' learning about the persistent types of other individuals. Alatas et al. (2016) and Dasaratha et al. (2018) model states that evolve according to an AR(1) process. Individuals take action simultaneously in each period after observing their neighbors' past actions, which are informative about past states and thus informative about the current state. The key feature distinguishing my model from these models is the diffusion process within each period, which randomly decides the order of play and the observation sets.

This per-period feature relates to the social learning literature, where each agent takes action exactly once and sequentially. Classical papers of this literature are Banerjee (1992), Bikhchandani et al. (1992) and Smith and Sørensen (2000); recent papers include, for example, Acemoglu et al. (2011) and Lobel and Sadler (2015). The modeling of each period in my model is more related to these recent papers, but further allows randomness in the order of play. Moreover, the efficiency metrics of interest in this literature concerns the n-th individual in large populations, while my efficiency analysis concerns everyone on the network. The crucial difference between my model and the social learning literature goes beyond these per-period features. While most papers in this literature study perfect Bayesian equilibria under common knowledge of the network topology, my model lets individuals learn relevant information about their environments from repeated interactions.[7]

The construction of empiricist learning rules relates to the literature on fictitious play (Brown, 1951; Fudenberg and Kreps, 1993), where individuals asymptotically believe that the empirical frequency of others' play is their true stationary play. My paper brings this idea of the learning literature to a network game of information aggregation.

Local observability, and to a lesser extent, local knowledge of the network structure, are not new assumptions in the network literature. For example, these assumptions are implicit in DeGroot-style learning models, where each individual takes a weighted average of their neighbors' actions. Li and Tan (2020) takes the localness of knowledge to a model of misspecified learning, where each individual assumes her local network is the full network. McBride (2006, 2008) studies network formation models with imperfect monitoring of other' network relationships, allowing incorrect perceptions of others' links to arise as an equilibrium phenomenon. Breza et al. (2018), from their empirical finding that network knowledge is limited and local, suggest using models with incomplete information of the network structure. Rather than turning to heuristics, misspecifications or richer information structure, I build a model with repeated interactions for robust learning.

Since my model is not closely related to any previous work, its concepts of convergence and quality of information aggregation are not direct analogs of similarly named concepts in other papers. Convergence in my model means that asymptotically individuals' actions depend only on their current observations. Each individual asymptotically learns the long-run conditional distribution of her observation, based on which she forms a belief about the current state of the world given her current observation. Knowledge of the distribution of private observation conditional on the state of the world is exactly what is needed for the analysis of social learning models by, for example, Acemoglu et al. (2011) and Lobel and Sadler (2015). In these models, such local knowledge is immediately implied by common

[7] Some exceptions in the social learning literature include, for example, Wiseman (2009), Eyster and Rabin (2010), Guarino and Jehiel (2013), and Bohren (2016). However, individuals' behaviors in these models are also not founded by learning.

knowledge of the network topology and perfect Bayesian equilibria. On the technical side, the per-period play of these models satisfies two key assumptions for the convergence result in my paper. First is the sequential nature of moves, which is crucial to the iterative argument over action rounds. Second is access to private signals, which allows individuals to back out the conditional distribution of their observation from the unconditional distribution. This means that with some small modifications, empiricist learning rules can be used to provide a learning foundation for these social learning models, relaxing the common knowledge assumption of the network topology and of opponents' play.

Using a model very different from mine, with only one state of the world and individuals communicating their beliefs over time as more information about that state arrives, Li and Tan (2020) identify clique trees as the necessary and sufficient condition for strong efficiency.[8] While this result sounds similar to my result on quality of information aggregation, failure of efficiency in their model must come from inferring mistakes induced by misspecified beliefs about the network structure. In contrast, individuals in my model always correctly learn the conditional distribution of their observations. It is local observability and coarse communication that give rise to an endogenous form of misalignment of interest, challenging social optimality.

8 Conclusion

This paper proposes a novel model of information diffusion and aggregation that allows a separation between learning about the environment and aggregating information about the current state. Empiricist learning rules abstract from the details of the environment, focusing directly on the relevant object of learning, that is, the distribution of private observation conditional on the true state of the world. In stationary environments, empiricist learning rules achieve asymptotic learning of the local environment, and thus asymptotic individual optimality. I study convergence of play and efficiency of information aggregation when all individuals use empiricist learning rules and smooth decision rules, which are technical modifications of optimal decision rules.

In this paper, convergence of play means that asymptotically individuals' actions depend only on their current observations. If convergence holds then the long-run quality of information aggregation, including for example, how likely individuals' actions agree with the true state of the world or agree with each other, can be evaluated at a stationary play. I show that if all individuals adopt empiricist learning rules and smooth decision rules, then play converges for all objective environments. The proof uses an inductive argument on action

[8]See Proposition 2 of Li and Tan (2020).

rounds. Convergence of play in previous rounds means that the conditional distribution of an individual's observation in the current round converges to a stationary distribution. Empiricist learning rules then ensures asymptotic learning of this stationary distribution. Finally, convergence of beliefs leads to convergence of play under smooth decision rules. Note that the conditional distributions of an individual's observation at different pairs of an observation set and an action round are all linked together by the fundamentals of the underlying environment. However, empiricist learning rules essentially treat these conditional distributions separately. While forgoing some information, this feature avoids contamination of learning from later rounds to earlier rounds. This is the exact reason why the inductive argument works.

My analysis on quality of information aggregation focuses on the long-run probability that each individual's action matches the true state of the world. Despite individuals' limited knowledge of the network structure, local observability, selfish and myopic motives, the long-run play induced by empiricist learning rules and smooth decision rules achieves strong efficiency on clique trees. Specifically, it maximizes each individual's probability of matching the state of the world within the class of stationary plays that treat the two states symmetrically. Two key properties of clique trees ensure this result. First, neighbors' actions are conditionally independent, so their overall informativeness is maximized by maximizing the individual informativeness of each neighbor's action. Second, an individual's action is most informative to her neighbors when it is optimal to herself.

I then identify several distinct reasons for why efficiency of information aggregation is likely to fail in general circumstances, including a weak converse to the positive efficiency result. On any network that is not a clique tree, there exist some diffusion process and private signals such that the long-run play induced by empiricist learning rules and smooth decision rules is strictly dominated by some symmetric stationary play. This result generalizes the intuition of two examples that speak to two key properties of clique trees: conditional independence of neighbor's actions and local alignment of interests. In one example when the actions of two neighbors are conditionally correlated, a suboptimal play by one neighbor might break this correlation in a way that improves their overall informativeness. In another example when the flow of information to an individual affects both the quality of an observed neighbor's action and whether her action is observed by another neighbor, a suboptimal decision given the average quality of the action of the former neighbor may be optimal to the latter neighbor. Moreover, these forces similarly present challenges to Pareto efficiency when individuals can trade favors across rounds. While efficiency comparison in my paper focuses on environments without biases, it is also noted that exposure to biases in opposite directions may help with information aggregation.

There are several directions for future research. One direction is to formalize the extension of my convergence result to richer settings, thus providing a learning foundation for a rich class of games of information aggregation. More specifically, my model presents a generalizable two-step framework. In the first step, one would build a model where the game of interest is played repeatedly, and construct learning rules that ensure convergence of learning about relevant elements of the environment. In the second step, one would then analyze the asymptotic play of the multi-period model given individuals' asymptotic learning, thus providing a prediction for the outcome of the one-period game. For example, if the two-step framework is adapted to the context of social learning, it would predict the same outcome as the perfect Bayesian equilibria studied in papers that assume common knowledge of the underlying objective environment. That is, the two-step framework would achieve the same predictions as these papers, while making only minimal assumptions on individuals' knowledge of the environment.

Finally, if my model is taken seriously as a model of how people interact on social platforms, it has the following implications. First, access to independent private news is important for an individual to learn the reliability of information from her neighbors. Second, since social networks are unlikely to be clique trees, social platforms are unlikely to achieve social optimality of information aggregation. Future research could build on these baseline findings and the various challenges to efficiency illustrated in this paper to study, for example, policies that improve observability of information paths and the implications of persistent biases in the context of social platforms.

A Proofs

A.1 Proof that local models are well-defined

To show that every local model is well defined, I show that when i acts in round 2 or later, she could hear from any subset of her neighbors and possibly also receive a private signal.

Lemma A1. *Take any $E \in \bar{\mathcal{E}}_i$. For every nonempty set $M_i \subseteq N_i$ and $r \in \{2, ..., R\}$, $\Pr^E(M_i(\tau) = M_i \cup \{i\}, r_i(\tau) = r) > 0$ and $\Pr^E(M_i(\tau) = M_i, r_i(\tau) = r) > 0$.*

Proof. Construct $\tau \in \{1, ..., R, \infty\}^n$ with $\tau_i = r, \tau_j = r - 1$ for all $j \in M_i$ and $\tau_j = \infty$ for all $j \notin M_i \cup \{i\}$. Then $M_i(\tau) = M_i \cup \{i\}$ and $r_i(\tau) = r$. Similarly, construct $\tau' \in \{1, ..., R, \infty\}^n$ with $\tau'_j = r - 1$ for all $j \in M_i$ and $\tau'_j = \infty$ for all $j \notin M_i$. Then $M_i(\tau') = M_i$ and $r_i(\tau') = r$. The claim follows from the assumption that the distribution of the signal-timing vector has full support. □

A.2 Proof of Lemma 1

To see the second equation of part 2, recall that private signals are independent of the diffusion process and across individuals, conditional on the state of the world. Take any nonempty set $M_i \subseteq N_i$ and $r \in \{2, ..., R\}$. For any $\tau \in \{1, ..., R, \infty\}^n$ such that $M_i(\tau) = M_i \cup \{i\}$ and $r_i(\tau) = r$, the actions of i's neighbors in round $r-1$ are not affected by i's private signal. That is, for such τ,

$$\Pr^E(a_{M_i}, s_i|\theta, \tau) = \Pr^E(a_{M_i}|\theta, \tau)\Pr^E(s_i|\theta, \tau) = \Pr^E(a_{M_i}|\theta, \tau)\Pr^E(s_i|\theta).$$

Then,

$$\begin{aligned} & f_i^E(a_{M_i}, s_i|\theta, M_i \cup \{i\}, r) \\ = & \sum_{\tau: M_i(\tau) = M_i \cup \{i\}, r_i(\tau) = r} \Pr^E(a_{M_i}, s_i|\theta, \tau)\Pr^E(\tau|M_i(\tau) = M_i \cup \{i\}, r_i(\tau) = r) \\ = & \Pr(s_i|\theta) f_i^E(a_{M_i}|\theta, M_i \cup \{i\}, r). \end{aligned}$$

It follows that

$$\begin{aligned} f_i^E(a_{M_i}, s_i|M_i \cup \{i\}, r) &= \sum_{\theta \in \{-1,1\}} \Pr(\theta)\Pr(s_i|\theta) f_i^E(a_{M_i}|\theta, M_i \cup \{i\}, r) \\ &= \sum_{\theta \in \{-1,1\}} \Pr(s_i, \theta) f_i^E(a_{M_i}|\theta, M_i \cup \{i\}, r). \end{aligned}$$

Furthermore, the matrix $(\Pr^E(s_i, \theta))_{s_i, \theta \in \{-1,1\}} = \frac{1}{2}(\Pr^E(s_i|\theta))_{s_i, \theta \in \{-1,1\}}$ is full rank since $q_i > 1/2$. This completes the proof of part 1 of the lemma.

Finally, the first equation of part 2 relies on the assumption that the arrivals of private signals are independent across individuals. Conditional on i not having received her private signal by round $r-1$, the actions of those neighbors of i that act in round $r-1$ depend only on the arrivals and realizations of the private signals of individuals other than i. Formally, take any two signal-timing vectors τ and τ' such that $\tau_j = \tau'_j$ for all $j \neq i$, $\tau_i \geq r, \tau'_i \geq r$ and $M_i = M_i(\tau) \backslash \{i\}$, it holds that $\Pr^E(a_{M_i}|\theta, \tau, s) = \Pr^E(a_{M_i}|\theta, \tau', s)$ for all $a_{M_i} \in \{-1, 1\}^{|M_i|}$.

This implies

$$
\begin{aligned}
&f_i^E(a_{M_i}|\theta, M_i, r) \\
=&\frac{\sum_{\tau:M_i(\tau)=M_i}\sum_s H_{-i}(\tau_{-i})H_i(\tau_i)\mathrm{Pr}^E(s|\theta)\mathrm{Pr}^E(a_{M_i}|\theta,\tau,s)}{\sum_{\tau:M_i(\tau)=M_i} H_{-i}(\tau_{-i})H_i(\tau_i)} \\
=&\frac{\mathrm{Pr}^E(\tau_i > r)\sum_{\tau_{-i}:M_i(\tau_{-i},\infty)=M_i}\sum_s H_{-i}(\tau_{-i})\mathrm{Pr}^E(s|\theta)\mathrm{Pr}^E(a_{M_i}|\theta,(\tau_{-i},\infty),s)}{\mathrm{Pr}^E(\tau_i > r)\sum_{\tau_{-i}:M_i(\tau_{-i},\infty)=M_i} H_{-i}(\tau_{-i})} \\
=&\frac{\mathrm{Pr}^E(\tau_i = r)\sum_{\tau_{-i}:M_i(\tau_{-i},r)\setminus\{i\}=M_i}\sum_s H_{-i}(\tau_{-i})\mathrm{Pr}^E(s|\theta)\mathrm{Pr}^E(a_{M_i}|\theta,(\tau_{-i},r),s)}{\mathrm{Pr}^E(\tau_i = r)\sum_{\tau_{-i}:M_i(\tau_{-i},r)\setminus\{i\}=M_i} H_{-i}(\tau_{-i})} \\
=&f_i^E(a_{M_i}|M_i\cup\{i\}, r, \theta).
\end{aligned}
$$

A.3 Proof of Proposition 1

Notice that by Lemma 1, there is a one-to-one mapping $\Phi : (f_i^E(.|M_i\cup\{i\}, r))_{M_i,r} \mapsto f_i^E$ between the stationary unconditional distributions of an individual's observation (when she receives her private signal) and her local model. Moreover, this mapping is linear.

By the Strong Law of Large Numbers,

$$
\mathrm{Pr}^E\left(\lim_{t\to\infty}\|(\hat{f}_i(.|M_i\cup\{i\}, r)(h_i^t))_{M_i,r} - (f_i^E(.|M_i\cup\{i\}, r))_{M_i,r}\| = 0\right) = 1.
$$

The linearity of Φ then implies that for every $\epsilon > 0$, there exists $\epsilon' > 0$ such that

$$
\begin{aligned}
1 &= \mathrm{Pr}^E\left(\lim_{t\to\infty}\mathbf{1}\left\{\Phi(B^{\epsilon'}((\hat{f}_i(.|M_i\cup\{i\}, r)(h_i^t))_{M_i,r})) \subseteq B^{\epsilon}(\Phi((f_i^E(.|M_i\cup\{i\}, r))_{M_i,r}))\right\} = 1\right) \\
&= \mathrm{Pr}^E\left(\lim_{t\to\infty}\mathbf{1}\left\{F_i^{\epsilon'}(h_i^t) \subseteq B^{\epsilon}(f_i^E)\right\} = 1\right).
\end{aligned}
$$

Since $\lim_{t\to\infty}\gamma_i(h_i^t)(F_i^{\epsilon'}(h_i^t)) = 1$ by the definition of empiricist learning rules, part 1 of the proposition follows.

For part 2, construct a stationary belief-updating rule $\bar{\beta}_i^* : h_{it} \to b_i \in \Delta(\{-1, 1\})$ such that for all observation h_{it} and all $\theta \in \{-1, 1\}$,

$$
\bar{\beta}_i^*(h_{it})(\theta) = \frac{\mathrm{Pr}^E(h_{it}|\theta_t = \theta)}{\mathrm{Pr}^E(h_{it}|\theta_t = \theta) + \mathrm{Pr}^E(h_{it}|\theta_t = -\theta)}.
$$

It follows from part 1 that

$$
\mathrm{Pr}^E\left(\lim_{t\to\infty}\|\beta_i(\tilde{h}_i^{t-1}, h_{it})(\theta_t) - \bar{\beta}_i^*(h_{it})\| = 0\right) = 1.
$$

Recall that σ_i^η is a technical modification of the symmetric optimal stationary decision rule $\bar{\sigma}_i^*$, converging to $\bar{\sigma}_i^*$ as $t \to \infty$. Moreover, the decaying smoothing parameter of σ_i^η ensures that convergence of beliefs leads to convergence of play. Thus,

$$\Pr^E\left(\lim_{t\to\infty}\left|\sigma_i^\eta(\beta_i(\tilde{h}_i^{t-1}, h_{it}))(\theta_t) - \bar{\sigma}_i^*(\bar{\beta}_i^*(h_{it}))(\theta_t)\right| = 0\right) = 1.$$

This implies part 2 of the proposition.

A.4 The inductive step in Theorem 1

Lemma A2. *Fix environment $E = (G, H, (q_i)_{i\in N}, (\beta_i)_{i\in N}, (\sigma_i^\eta)_{i\in N})$ where for each i, β_i is an empiricist belief-updating rule and σ_i^η is a smooth decision rule. Consider an individual i, a nonempty subset $M_i \subseteq N_i$ and a round $r \in \{2, ..., R\}$. If there exists some stationary function $\bar{\psi}$ such that for every $j \in M_i$,*

$$\Pr^E\left(\lim_{t\to\infty} \mathbf{1}\{M_i(\tau_t) = M_i \cup \{i\}, r_i(\tau_t) = r\}\|\psi_j(\omega_t, \tilde{h}^{t-1}, t) - \bar{\psi}_j(\omega_t)\| = 0\right) = 1,$$

then whenever i's observation set is M_i or $M_i \cup \{i\}$ and i's action round is r, i's belief and play converge. That is,

1) for some unconditional distribution $f_i(.|M_i \cup \{i\}, r)$, it holds for all $\epsilon > 0$ that

$$\Pr^E\left(\lim_{t\to\infty} \gamma_i(h_i^t)\left(\{f_i' \in \mathcal{F}_i : \|f_i'(.|M_i \cup \{i\}, r) - f_i(.|M_i \cup \{i\}, r)\| > \epsilon\}\right) = 0\right) = 1;$$

2) for some stationary belief-updating rule $\bar{\beta}_i^*$,

$$\Pr^E\left(\lim_{t\to\infty} \sum_{a_{M_i}, s_i} \left[\begin{array}{c}\mathbf{1}\{h_{it} = (a_{M_i}, s_i, r) \text{ or } h_{it} = (a_{M_i}, r)\} \\ \times\|\beta_i(\tilde{h}_i^{t-1}, h_{it}) - \bar{\beta}_i^*(h_{it})\|\end{array}\right] = 0\right) = 1;$$

3) and for some stationary function $\bar{\psi}_i^*$ of i's play,

$$\Pr^E\left(\lim_{t\to\infty} \sum_{a_{M_i}, s_i} \left[\begin{array}{c}\mathbf{1}\{M_i(\tau_t) = M_i \text{ or } M_i(\tau_t) = M_i \cup \{i\}, \text{ and } r_i(\tau_t) = r\} \\ \times\|\psi_i(\omega_t, \tilde{h}^{t-1}, t) - \bar{\psi}_i^*(\omega_t)\|\end{array}\right] = 0\right) = 1.$$

Proof. From $\bar{\psi}_{M_i}$, construct for each $a_{M_i} \in \{-1, 1\}^{|M_i|}$ and $s_i \in \{-1, 1\}$,

$$f_i(a_{M_i}, s_i, r|M_i \cup \{i\}, r) = \Pr^E(\bar{\psi}_{M_i}(\omega) = a_{M_i}, s_i|M_i(\tau) = M_i \cup \{i\}, r_i(\tau) = r).$$

By Lemma 1, for each nonempty set $M_i \subseteq N_i$ and action round $r \in \{2, ..., R\}$, there is a one-to-one linear mapping $\phi^{M_i,r} : f_i(.|M_i \cup \{i\}, r) \mapsto (f_i(.|\theta, M_i \cup \{i\}, r), f_i(.|\theta, M_i, r))_{\theta \in \{-1,1\}}$.

Construct a hypothetical empirical distribution that takes draws from $f_i(.|M_i \cup \{i\}, r)$ such that for all $a_{M_i} \in \{-1,1\}^{|M_i|}$, $s_i \in \{-1,1\}$ and history ω^t of system states,

$$\tilde{f}_i(a_{M_i}, s_i|M_i \cup \{i\}, r)(\omega^t) = \frac{\sum_{t'=1}^{t} \mathbf{1}\{\bar{\psi}_{M_i}(\omega_{t'}) = a_{M_i}, s_{it'} = s_i, r_i(\tau_{t'}) = r\}}{\sum_{t'=1}^{t} \mathbf{1}\{M_i(\tau_{t'}) = M_i \cup \{i\}, r_i(\tau_{t'}) = r\}}$$

if the denominator is positive, otherwise set $\tilde{f}_i(a_{M_i}, s_i|M_i \cup \{i\}, r)(\omega^t) = 1/(2^{|M_i|+1})$.

By the inductive hypothesis,

$$\Pr^E \left(\lim_{t\to\infty} \| \hat{f}_i(.|M_i \cup \{i\}, r)(h_i^t) - \tilde{f}_i(.|M_i \cup \{i\}, r)(\omega^t) \| = 0 \right) = 1.$$

Moreover, by the Strong Law of Large Numbers,

$$\Pr^E \left(\lim_{t\to\infty} \| \tilde{f}_i(.|M_i \cup \{i\}, r)(\omega^t) - f_i(.|M_i \cup \{i\}, r) \| = 0 \right) = 1.$$

It follows that

$$\Pr^E \left(\lim_{t\to\infty} \| \hat{f}_i(.|M_i \cup \{i\}, r)(h_i^t) - f_i(.|M_i \cup \{i\}, r) \| = 0 \right) = 1.$$

By an argument analogous to the proof of Proposition 1 and using the linearity of the mapping $\phi^{M_i,r}$, it can then be shown that γ_i asymptotically puts zero probability on local models that induce an unconditional distribution at observation set $M_i \cup \{i\}$ and action round r of some ϵ-distance away from that induced by f_i. That is, part 1 of this lemma holds.

Construct $\bar{\beta}_i^*$ such that for every $a_{M_i} \in \{-1,1\}^{|M_i|}$, $s_i \in \{-1,1\}$ and $\theta \in \{-1,1\}$,

$$\bar{\beta}_i^*(a_{M_i}, s_i, r)(\theta) = \frac{f_i(a_{M_i}, s_i|\theta, M_i \cup \{i\}, r)}{f_i(a_{M_i}, s_i|\theta, M_i \cup \{i\}, r) + f_i(a_{M_i}, s_i| -\theta, M_i \cup \{i\}, r)}$$

and

$$\bar{\beta}_i^*(a_{M_i}, r)(\theta) = \frac{f_i(a_{M_i}|\theta, M_i, r)}{f_i(a_{M_i}|\theta, M_i, r) + f_i(a_{M_i}| -\theta, M_i, r)}.$$

Such a stationary belief-updating rule satisfies part 2 of the lemma.

To see part 3, recall that smooth decision rule σ_i^η converges to the symmetric optimal stationary decision rule $\bar{\sigma}_i^*$ and ensures that convergence of beliefs leads to convergence of play. Construct a stationary function of i's play, $\bar{\psi}_i^*$, as following. If ω induces for i

observation (a_{M_i}, s_i, r), set

$$\bar{\psi}_i^*(\omega) = \begin{cases} -1 & \text{if } \zeta_i \leq \bar{\sigma}_i^*(\bar{\beta}_i^*(a_{M_i}, s_i, r))(-1) \\ 1 & \text{otherwise.} \end{cases}$$

If ω induces for i observation (a_{M_i}, r), set

$$\bar{\psi}_i^*(\omega) = \begin{cases} -1 & \text{if } \zeta_i \leq \bar{\sigma}_i^*(\bar{\beta}_i^*(a_{M_i}, r))(-1) \\ 1 & \text{otherwise.} \end{cases}$$

Then $\bar{\psi}^*$ satisfies part 3 of the lemma. □

Finally, notice that the case when i only receives her private signal is similar to the base case of $r = 1$. In this case when $h_{it} = (s_{it}, r)$, the individual holds correct belief about the current state, that is, for all $\theta \in \{-1, 1\}$,

$$\beta_i(\tilde{h}_i^{t-1}, h_{it})(\theta) = \bar{\beta}_i^*(s_{it}, r)(\theta) = q_i^{\mathbf{1}\{s_i=\theta\}}(1 - q_i)^{\mathbf{1}\{s_i \neq \theta\}},$$

regardless of history $\tilde{h}_i^{t-1}$. Moreover, smooth decision rules imply that asymptotically, an individual i facing such history reports her private signal. For all ω that induces observation (s_i, r) for individual i, set $\bar{\psi}_i^*(\omega) = s_i$.

Together with parts 2 and 3 of Lemma A2, this final argument completes the inductive step in the proof of parts 1 and 2 of Theorem 1. To see part 3 of Theorem 1, collect part 1 of Lemma *A2* over all possible observation sets and action rounds of each individual i.

A.5 Proof of the equivalence in Definition 7

It is easy to see that for any given objective environment, a profile of symmetric belief-updating rules and a profile of symmetric decision rules induce a symmetric feasible stationary function of play. For the other direction, fix an objective environment and a symmetric feasible stationary function of play $\bar{\psi}$. Since $\bar{\psi}$ is feasible, it can be induced by some profile $(\bar{\beta}_i')_{i \in N}$ of stationary belief-updating rules and some profile $(\bar{\sigma}_i')_{i \in N}$ of stationary decision rules. Consider an observation (a_{M_i}, r) of individual i, under symmetry of $\bar{\psi}$,

$$\begin{aligned} &\{(\tau, s, \zeta) : (\tau, s, \zeta) \text{ induces observation } (a_{M_i}, r) \text{ for } i\} \\ =&\{(\tau, s, \zeta) : (\tau, -s, \mathbf{1} - \zeta) \text{ induces observation } (-a_{M_i}, r) \text{ for } i\}. \end{aligned}$$

Call this set $\tilde{\Omega}$. By the symmetry of $\bar{\psi}_i$,

$$\begin{aligned}\bar{\sigma}_i'(\bar{\beta}_i'(a_{M_i}, r))(1) &= \Pr(\bar{\psi}_i(\tau, s, \zeta) = 1 | (\tau, s, \zeta) \in \tilde{\Omega}) \\ &= \Pr(\bar{\psi}_i(\tau, -s, \mathbf{1} - \zeta) = -1 | (\tau, s, \zeta) \in \tilde{\Omega}) \\ &= \bar{\sigma}_i'(\bar{\beta}_i'(-a_{M_i}, r))(-1).\end{aligned}$$

Similarly, $\bar{\sigma}_i'(\bar{\beta}_i'(a_{M_i}, s_i, r))(1) = \bar{\sigma}_i'(\bar{\beta}_i'(-a_{M_i}, -s_i, , r))(-1)$ for every history (a_{M_i}, s_i, r).

Now, construct symmetric functions $(\bar{\beta}_i)_{i\in N}$ and $(\bar{\sigma}_i)_{i\in N}$ from $(\bar{\beta}_i')_{i\in N}$ and $(\bar{\sigma}_i')_{i\in N}$. For each pair $(a_{M_i}, -a_{M_i})$, set $\bar{\beta}_i(a_{M_i}, r) = \bar{\beta}_i'(a_{M_i}, r)$ and $\bar{\sigma}_i(\bar{\beta}_i(a_{M_i}, r)) = \bar{\sigma}_i'(\bar{\beta}_i'(a_{M_i}, r))$; then set $\bar{\beta}_i(-a_{M_i}, r) = \mathbf{1} - \bar{\beta}_i(a_{M_i}, r)$ and $\bar{\sigma}_i(\bar{\beta}_i(-a_{M_i}, r)) = \bar{\sigma}_i'(\bar{\beta}_i'(-a_{M_i}, r))$. Perform similar construction for cases when i receives her private signal. For observations that never arise, the beliefs induced by the belief-updating rules at such observations can be determined arbitrarily and symmetrically. Similarly for beliefs that never arise, the decision rules can be determined arbitrarily and symmetrically at such beliefs. By construction $(\bar{\beta}_i)_{i\in N}$ and $(\bar{\sigma}_i)_{i\in N}$ induce $\bar{\psi}$ and they are symmetric.

A.6 Proof of two technical properties of clique trees

The first property follows immediately from the existence of a unique shortest path connecting any two individuals in G. To see the second property, suppose by contradiction that for some $i \in N, j \in N_i$ and $k \in N_j \backslash (N_i \cup \{i\})$, there exists some $k' \in IF_{k\backslash j}$ but $k' \notin IF_{j\backslash i}$. This means that $l(k', i) < l(k', j) + l(j, i)$, that is, the shortest path connecting k' to i does not go through j. This path and the shortest path that goes from k' to i through j form a cycle. This cycle includes k, but it is not complete because k and i are not linked. This contradicts that G is a clique tree.

A.7 Proof of Lemma 3

By property 1 of Lemma 2, $\{IF_{j\backslash i}\}_{j\in N_i}$ are mutually exclusive. Consider the case that $i \notin M_i$ and enumerate individuals in M_i by $j_1, ..., j_{|M_i|}$. By the conditional independence of private signals and the independence of the instrumental variables, $\Pr(\bar{\psi}_{M_i}|\theta, \tau) = \prod_{j\in M_i} \Pr(\bar{\psi}_j|\theta, \tau)$ for any τ such that $M_i(\tau) = M_i$ and $r_i(\tau) = r$.

Next, notice that the event that i observes from M_i and acts in round r can be rewritten as a join of independent events

$$\{\tau : M_i(\tau) = M_i, r_i(\tau) = r\} = \left(\cap_{j\in M_i}\left\{\tau : \min_{k\in IF_{j\backslash i}} l(j,k) + \tau_k = r-1\right\}\right) \\ \cap \left(\cap_{j\in N_i\backslash M_i}\left\{\tau : \min_{k\in IF_{j\backslash i}} l(j,k) + \tau_k \geq r\right\}\right) \cap \{\tau : \tau_i > r\}.$$

For all $j \in M_i$, let $\mathcal{A}_{ij} = \{\tau_{IF_{j\backslash i}} : \min_{k\in IF_{j\backslash i}} l(j,k) + \tau_k = r-1\}$. Also, let

$$\mathcal{A}_i^- = \{\tau_{N\backslash(\cup_{j\in M_i} IF_{j\backslash i})} : \tau_i > r \text{ and } \min_{k\in IF_{j\backslash i}} l(j,k) + \tau_k \geq r \text{ for all } j \in N_i\backslash M_i\}.$$

It follows from the above decomposition that

$$\begin{aligned} &\Pr(\tau|M_i(\tau) = M_i, r_i(\tau) = r) \\ =&\Pr\left(\tau_{N\backslash(\cup_{j\in M_i} IF_{j\backslash i})}|\tau_{N\backslash(\cup_{j\in M_i} IF_{j\backslash i})} \in \mathcal{A}_i^-\right) \prod_{j\in M_i} \Pr(\tau_{IF_{j\backslash i}}|\tau_{IF_{j\backslash i}} \in \mathcal{A}_{ij}). \end{aligned}$$

Thus,

$$\begin{aligned} &\Pr(\bar{\psi}_{M_i}(\omega)|M_i(\tau) = M_i, r_i(\tau) = r) \\ =&\sum_{\tau: M_i(\tau)=M_i, r_i(\tau)=r} \Pr(\bar{\psi}_{M_i}(\omega)|\theta,\tau)\Pr(\tau|M_i(\tau) = M_i, r_i(\tau) = r) \\ =&\sum_{\tau_{IF_{j_1\backslash i}}\in\mathcal{A}_{ij_1}} \cdots \sum_{\tau_{IF_{j_{|M_i|}\backslash i}}\in\mathcal{A}_{ij_{|M_i|}}} \\ &\sum_{\tau_{N\backslash(\cup_{j\in M_i} IF_{j\backslash i})}\in\mathcal{A}_i^-} \begin{pmatrix} \prod_{j\in M_i}\Pr(\bar{\psi}_j(\omega)|\theta,\tau_{IF_{j\backslash i}}, M_i(\tau) = M_i, r_i(\tau) = r) \\ \times\prod_{j\in M_i}\Pr(\tau_{IF_{j\backslash i}}|\tau_{IF_{j\backslash i}} \in \mathcal{A}_{ij}) \\ \times\Pr\left(\tau_{N\backslash(\cup_{j\in M_i} IF_{j\backslash i})}|\tau_{N\backslash(\cup_{j\in M_i} IF_{j\backslash i})} \in \mathcal{A}_i^-\right) \end{pmatrix} \\ =&\prod_{j\in M_i}\Pr(\bar{\psi}_j(\omega)|M_i(\tau) = M_i, r_i(\tau) = r). \end{aligned}$$

The proof for the case when $i \in M_i$ is similar.

A.8 Proof of Lemma 4

When $r = 2$, it must be that $M_j(\tau) = \{j\}$ and thus $\tilde{q}_j^{M_i,r} = \Pr(\bar{\psi}_j(\omega) = \theta|\theta, M_j(\tau) = \{j\}, r_j(\tau) = 1) = q_j$. Now consider the case that $r \geq 3$. Since $j \in M_i(\tau)$, $(M_j(\tau)\backslash\{j\}) \cap (N_i \cup \{i\}) = \varnothing$. By the second property of Lemma 2 and the full support assumption of H, any observation set of j that satisfies the above condition is feasible, that is, $\Pr(M_j(\tau) =$

$M_j, M_i(\tau) = M_i, r_i(\tau) = r) > 0$ for all $M_j \subseteq (N_j \backslash (N_i \cup \{i\})) \cup \{j\}$ and $r \geq 3$. Then the quality of j's action observed by i at observation set (M_i, r) is

$$\tilde{q}^{M_i,r} = \sum_{M_j \subseteq (N_j \backslash (N_i \cup \{i\})) \cup \{j\}} \begin{pmatrix} \Pr(M_j(\tau) = M_j | M_i(\tau) = M_i, r_i(\tau) = r) \\ \times \Pr(\bar{\psi}_j(\omega) = \theta | \theta, M_j(\tau) = M_j, M_i(\tau) = M_i, r_i(\tau) = r) \end{pmatrix}.$$

It remains to show that for all $M_j \subseteq (N_j \backslash (N_i \cup \{i\})) \cup \{j\}$,

$$\begin{aligned} &\Pr(\bar{\psi}_j(\omega) = \theta | \theta, M_j(\tau) = M_j, M_i(\tau) = M_i, r_i(\tau) = r) \\ =&\Pr(\bar{\psi}_j(\omega) = \theta | \theta, M_j(\tau) = M_j, r_j(\tau) = r - 1). \end{aligned}$$

For such M_j, let

$$\tilde{\mathcal{A}}_{ij} = \{\tau_{\cup_{k \in M_j} IF_{k \backslash j}} : \exists \tau' \text{ s.t. } \tau'_{\cup_{k \in M_j} IF_{k \backslash j}} = \tau_{\cup_{k \in M_j} IF_{k \backslash j}}, M_j(\tau') = M_j, M_i(\tau') = M_i, r_i(\tau') = r\},$$
$$\tilde{\mathcal{A}}_{j} = \{\tau_{\cup_{k \in M_j} IF_{k \backslash j}} : \exists \tau' \text{ s.t. } \tau'_{\cup_{k \in M_j} IF_{k \backslash j}} = \tau_{\cup_{k \in M_j} IF_{k \backslash j}}, M_j(\tau') = M_j, r_j(\tau') = r - 1\}.$$

By Lemma 2, $\cup_{k \in M_j} IF_{k \backslash j} \subseteq IF_{j \backslash i}$. Therefore, conditional on $M_j(\tau) = M_j$ and $r_j(\tau) = r - 1$, the observation set and action round of i depends further only on $\tau_{IF_{j' \backslash i}}$ for $j' \in (N_i \cup \{i\}) \backslash \{j\}$. This means that $\tilde{\mathcal{A}}_{ij} = \tilde{\mathcal{A}}_j$. Furthermore, conditional on j hearing from M_j in round $r - 1$, $\bar{\psi}_j$ depends only on $(\tau_{j'}, s_{j'}, \zeta_{j'})$ of $j' \in \cup_{k \in M_j} IF_{k \backslash j}$ and on ζ_j. Then,

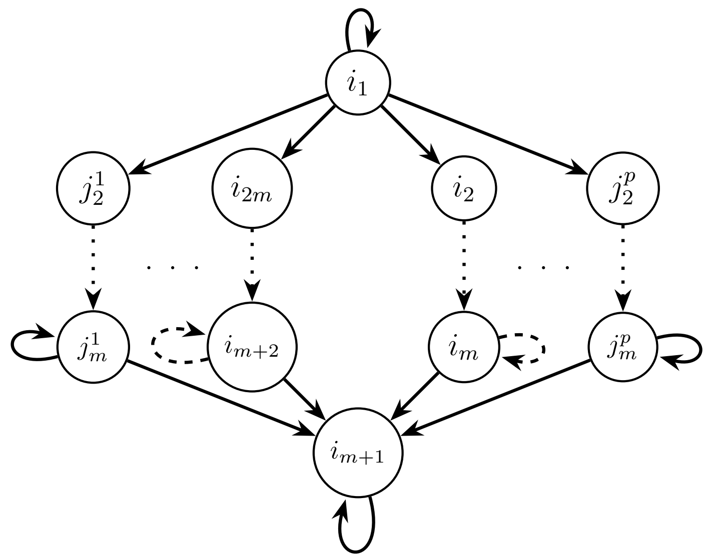
$$\begin{aligned} &\Pr(\bar{\psi}_j(\omega) | M_i(\tau) = M_i, M_j(\tau) = M_j, r_i(\tau) = r) \\ =& \sum_{\tau_{\cup_{k \in M_j} IF_{k \backslash j}} \in \tilde{\mathcal{A}}_{ij}} \begin{pmatrix} \Pr(\tau_{\cup_{k \in M_j} IF_{k \backslash j}} | M_j(\tau) = M_j, M_i(\tau) = M_i, r_i(\tau) = r) \\ \times \Pr(\bar{\psi}_j(\omega) | \tau_{\cup_{k \in M_j} IF_{k \backslash j}}, M_j(\tau) = M_j, M_i(\tau) = M_i, r_i(\tau) = r) \end{pmatrix} \\ =& \sum_{\tau_{\cup_{k \in M_j} IF_{k \backslash j}} \in \tilde{\mathcal{A}}_{j}} \begin{pmatrix} \Pr(\tau_{\cup_{k \in M_j} IF_{k \backslash j}} | M_j(\tau) = M_j, r_j(\tau) = r - 1) \\ \times \Pr(\bar{\psi}_j(\omega) | \tau_{\cup_{k \in M_j} IF_{k \backslash j}}, M_j(\tau) = M_j, r_j(\tau) = r - 1) \end{pmatrix} \\ =& \Pr(\bar{\psi}_j(\omega) | M_j(\tau) = M_j, r_j(\tau) = r - 1). \end{aligned}$$

This completes the proof of Lemma 4.

A.9 Proof of Theorem 3

Take a graph G. Denote by $\mathcal{C}(i_1, ..., i_L)$ a cycle of size L where $\{i_l, i_{l+1}\} \in G$ for all $l = 1, ..., L$ and $i_{L+1} = i_1$. This cycle is complete if any two individuals of this cycle are linked. A chord of a cycle is a link between two non-consecutive individuals. A cycle of size at least four is chordless if any two non-consecutive individuals are not linked. Furthermore,

Figure 7: The key diffusion in a generalization of Example 1



$\Pr(\tau_{i_{m+1}} = m+1) \approx 1$ when p is even

a graph is chordal if all cycles of size at least four have a link between two non-consecutive individuals.

Suppose that G is not a clique tree. Then G must have at least a cycle that is not complete. Take a smallest incomplete cycle $\mathcal{C}(i_1, ..., i_L)$ and refer to this subgraph of G as G'. There are three cases.

Case 1: G' is a chordless cycle of size $L = 2m$ for $m \geq 2$.

Consider a potential path $(i_1, j_2, ..., j_{m'}, i_{m+1})$ from i_1 to i_{m+1} different from the paths $(i_1, i_2, ..., i_m, i_{m+1})$ and $(i_1, i_{2m}, ..., i_{m+2}, i_{m+1})$. Let I^* collect individuals in $\{i_1, ..., i_{m+1}\}$ that appear in the first path. If $m' < m$, there must exist $i_{m_1}, i_{m_2} \in I^*$ such that $m_2 \geq m_1 + 2$. Then for some m'_1 and m'_2 such that $2 \leq m'_1, m'_2 \leq m'$, there is a cycle of size strictly less than L, $\mathcal{C}(i_{m_1}, i_{m_1+1}, ..., i_{m_2-1}, i_{m_2}, j_{m'_2}, j_{m'_2-1}, ..., j_{m'_1+1}, j_{m'_1})$. Moreover, this cycle is incomplete because i_{m_1} and i_{m_2} are not linked. This contradicts that G' is the smallest incomplete cycle of G. If $m' = m$ and I^* contains individuals other than i_1 and i_{m+1}, then a similar argument provides a contradiction to G' being the smallest incomplete cycle. This concludes that the shortest paths connecting i_1 and i_{m+1} include only paths of length m with no joint elements other than i_1 and i_{m+1}. Suppose that there are $p \geq 0$ such paths other than $(i_1, i_2, ..., i_m, i_{m+1})$ and $(i_1, i_{2m}, ..., i_{m+2}, i_{m+1})$. Denote such a path by $(i_1, j_2^{p'}, ..., j_m^{p'}, i_{m+1})$ for $p' = 1, ..., p$.

Consider first the case when p is even. Construct a vector of signal qualities with $q_{i_1} =$

$0.5+\epsilon$, $q_{i_m} = q_{i_{m+2}} = 0.8$ and $q_{i_{m+1}} = q_{j_m^{p'}} = 0.75$ for all $p' = 1, ..., p$. Construct the distribution of the signal-timing vector so that $\Pr(\tau_{i_1} = 1) \approx 1$, $\Pr(\tau_{i_m} = m) = \Pr(\tau_{i_m} = \infty) = \Pr(\tau_{i_{m+2}} = m) = \Pr(\tau_{i_{m+2}} = \infty) \approx 0.5$, $\Pr(\tau_{i_{m+1}} = m+1) = \Pr(\tau_{j_m^{p'}} = m) \approx 1$ for all $p' = 1, ..., p$, and $\Pr(\tau_i = \infty) \approx 1$ for all other i.

Figure 7 illustrates the subgraph of all shortest paths connecting i_1 and i_{m+1}, and plots the key diffusion for this example when p is even. In this diffusion, i_1 receives his private signal in round 1. He then passes this signal, $a_{i_1} = s_{i_1}$, onto i_2, i_{2m} and $j_2^{p'}$ for $p' = 1, ..., p$. In round 2, these individuals then simply pass a_1 onto i_3, i_{2m-1} and $j_3^{p'}$ for $p' = 1, ..., p$ respectively. This goes on until round m when $j_m^{p'}$ for $p' = 1, ..., p$ receive their private signals, which are more informative than a_{i_1}. They thus choose $a_{j_m^{p'}} = s_{j_m^{p'}}$. In round m, i_{m+2} and i_m may or may not receive their private signals. If i_{m+2} receives his private signal then he chooses $a_{i_{m+2}} = s_{i_{m+2}}$, otherwise $a_{i_{m+2}} = a_{i_{m+3}} = s_{i_1}$. Similarly, if i_m receives his private signal then he chooses $a_{i_m} = s_{i_m}$, otherwise $a_{i_m} = a_{i_{m-1}} = s_{i_1}$. In round $m+1$, i_{m+1} receives his private signal and hears from $J \cup \{i_{m+2}, i_m\}$, where $J = \{j_m^{p'} : p' = 1, ..., p\}$. Let $M_i = J \cup \{i_{m+2}, i_m, i_{m+1}\}$.

Notice that when $a_{i_{m+2}} \neq a_{i_m}$, the optimal decision of i_{m+1} depends only on his private signal and the actions of individuals in J. The key is whether consensus posts by i_{m+2} and i_m could overturn a decision based solely on i_{m+1}'s private signal and actions of her neighbors in J. The informativeness of $(s_{i_{m+1}}, a_J) = (s_{i_{m+1}}, s_J)$ is summarized by

$$\frac{\Pr(s_{i_{m+1}}, a_J|\theta, M_i(\tau) = M_i, r_i(\tau) = m+1)}{\Pr(-s_{i_{m+1}}, -a_J|\theta, M_i(\tau) = M_i, r_i(\tau) = m+1)} = \frac{\Pr(s_{i_{m+1}}, s_J|\theta)}{\Pr(-s_{i_{m+1}}, -s_J|\theta)} = \left(\frac{0.75}{0.25}\right)^{\sum_{j \in J \cup \{i_{m+1}\}} \mathbf{1}(s_j = \theta) - \mathbf{1}(s_j = -\theta)}.$$

If i_{m+2} and i_m follow the actions of i_{m+3} and i_{m-1} respectively when they do not receive private signals, the informativeness of the consensus posts by i_{m+2} and i_m as they reach i_{m+1} is summarized by

$$\frac{\Pr(a_{i_{m+2}} = a_{i_m} = \theta|\theta, M_i(\tau) = M_i, r_i(\tau) = m+1)}{\Pr(a_{i_{m+2}} = a_{i_m} = -\theta|\theta, M_i(\tau) = M_i, r_i(\tau) = m+1)} = \frac{(0.25)(0.5+\epsilon) + (0.5)(0.5+\epsilon)(0.8) + (0.25)(0.8)^2}{(0.25)(0.5-\epsilon) + (0.5)(0.5-\epsilon)(0.2) + (0.25)(0.2)^2} \approx 2.62 < \frac{0.75}{0.25}.$$

Thus, i_{m+1} optimally follows the majority of $\{s_{i_{m+1}}, a_{j_m^1}, ..., a_{j_m^p}\}$, ignoring the posts of i_{m+2} and i_m. If instead, i_{m+2} and i_m randomize equally between 1 and -1 when not receiving private signals, the informativeness of their consensus posts to i_{m+1} is summarized by

$$\begin{aligned}&\frac{\Pr(a_{i_{m+2}} = a_{i_m} = \theta|\theta, M_i(\tau) = M_i, r_i(\tau) = m+1)}{\Pr(a_{i_{m+2}} = a_{i_m} = -\theta|\theta, M_i(\tau) = M_i, r_i(\tau) = m+1)}\\=&\frac{((0.5)(0.5) + (0.5)(0.8))^2}{((0.5)(0.5) + (0.5)(0.2))^2} = 3.45 > \frac{0.75}{0.25}.\end{aligned}$$

In this alternative play, their consensus posts can overturn the majority of $\{s_{i_{m+1}}, a_{j_m^1}, ..., a_{j_m^p}\}$ when the winning margin is one. Thus, this change strictly benefits i_{m+1}.

When p is odd, modify the above construction by letting $\Pr(\tau_{i_{m+1}} = \infty) \approx 1$. In the individually optimal play, i_{m+1} follows the majority of $\{a_{j_m^1}, ..., a_{j_m^p}\}$. In the alternative play where i_{m+2} and i_m randomize equally between 1 and -1 when not receiving private signals, i_{m+1} strictly benefits from following these neighbors' consensus posts when the majority of $\{a_{j_m^1}, ..., a_{j_m^p}\}$ disagree with them but the majority is won by a margin of one.

Case 2: G' is a chordless cycle of size $L = 2m+1$ for $m \geq 2$.

Notice that $(i_1, i_2, ..., i_m, i_{m+1})$ is the unique shortest path connecting i_1 and i_{m+1}. The reason is that if there was an alternative path of length less than or equal to m connecting i_1 and i_{m+1}, an incomplete cycle of size at most $2m$ would be formed by elements of these two paths. This would contradict that G' is the smallest incomplete cycle. Similarly, $(i_1, i_{2m+1}, ..., i_{m+3}, i_{m+2})$ is the unique shortest path connecting i_1 and i_{m+2}.

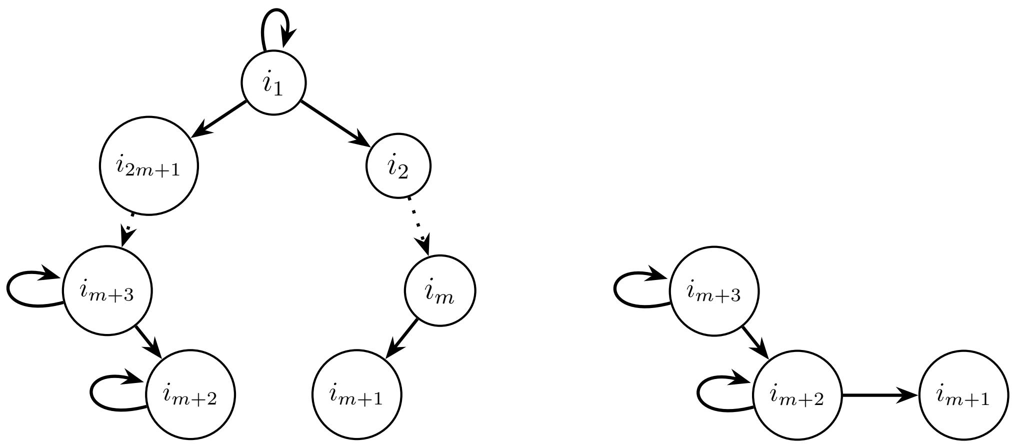
Construct a vector of signal qualities with $q_{i_1} = 0.9$, $q_{i_{m+3}} = 0.5 + \epsilon$ and $q_{i_{m+2}} = 0.7$. Construct a distribution H over the signal-timing vector such that $\Pr(\tau_{i_1} = 1) = \Pr(\tau_{i_1} = \infty) \approx 0.5$, $\Pr(\tau_{i_{m+3}} = m) \approx 1$, $\Pr(\tau_{i_{m+2}} = m+1) \approx 1$ and $\Pr(\tau_i = \infty) \approx 1$ for all $i \in N \backslash \{i_1, i_{m+3}, i_{m+2}\}$. Figure 8 plots two key diffusions of this example, with each occurring with probability approximately 0.5. In the left panel, i_1 receives his private signal in round 1 and has it spread to i_m and i_{m+3} in round m. Both of these individuals pass on this information to i_{m+1} and i_{m+2} respectively. In round $m+1$, individual i_{m+2} would consider between the action of i_{m+3} and her own private signal, but her decision will not be observed by i_{m+1}, who already takes an action in the same round. In the right panel, i_{m+3} receives his private signal in round m and passes that onto i_{m+2}, who will consider this piece of information with his own private signal to make a decision later observed by i_{m+1}.

Under the constructed diffusion process,

$$\begin{aligned}&\Pr(a_{i_{m+3}} = \theta|\theta, M_{i_{m+2}}(\tau) = \{i_{m+3}, i_{m+2}\}, r_{i_{m+2}}(\tau) = m+1)\\\approx&0.5(0.9) + 0.5(0.5+\epsilon) = 0.7 + 0.5\epsilon > q_{i_{m+2}}.\end{aligned}$$

It follows that at observation set $M_{i_{m+2}}(\tau) = \{i_{m+3}, i_{m+2}\}$ and action round $r_{i_{m+2}}$, individual

Figure 8: The key diffusions in a generalization of Example 2



$a_{i_{m+2}} = a_{i_{m+3}} = s_1, i_{m+2} \nrightarrow i_{m+1}$ $\qquad$ $a_{i_{m+2}} = a_{i_{m+3}} = s_{i_{m+3}}, i_{m+2} \rightarrow i_{m+1}$

i_{m+2} optimally chooses $a_{i_{m+2}} = a_{i_{m+3}}$. This gives

$$\Pr(a_{i_{m+2}} = \theta | \theta, M_{i_{m+1}}(\tau) = \{i_{m+2}\}, r_{i_{m+1}}(\tau) = m+2) = \Pr(s_{i_{m+3}} = \theta | \theta) = 0.5 + \epsilon.$$

If instead i_{m+2} chooses $a_{i_{m+2}} = s_{i_{m+2}}$, then

$$\Pr(a_{i_{m+2}} = \theta | \theta, M_{i_{m+1}}(\tau) = \{i_{m+2}\}, r_{i_{m+1}}(\tau) = m+2) = \Pr(s_{i_{m+2}} = \theta | \theta) = 0.7.$$

Such change strictly benefits i_{m+1} at a small expense of i_{m+2}.

Case 3: G' is a chordal graph.

Since G' is an incomplete chordal graph, it must be of size at least four and have a chord. Without loss of generality, suppose that i_1 and i_l are linked for some $l \in \{3, ..., L-1\}$. Then $\mathcal{C}(i_1, i_2, ..., i_{l-1}, i_l)$ and $\mathcal{C}(i_l, i_{l+1}, ..., i_L, i_1)$ are two cycles smaller than G'. These two cycles must be complete because otherwise it contradicts that G' is a smallest incomplete cycle. Since G' is incomplete, this means that there exist $j \in \{i_2, ..., i_{l-1}\}$ and $k \in \{i_{l+1}, ..., i_L\}$ that are not linked. In fact, $\mathcal{C}(i_1, j, i_l, k)$ is a smallest incomplete cycle of G. Consider the subgraph of G consisting of all shortest paths connecting j and k. Denote by $J = \{j_1, ..., j_p\}$ the set of individuals in this subgraph other than i_1, j, i_l and k. Construct a vector of signal qualities with $q_j = 0.5 + \epsilon$, $q_{i_1} = q_{i_l} = 0.8$ and $q_k = q_{j_{p'}} = 0.75$ for all $p' = 1, ..., p$. Construct the distribution of the signal-timing vector so that $\Pr(\tau_j = 1) \approx 1$, $\Pr(\tau_{i_1} = 2) = \Pr(\tau_{i_1} = \infty) = \Pr(\tau_{i_l} = 2) = \Pr(\tau_{i_l} = \infty) \approx 0.5$ and $\Pr(\tau_{j_{p'}} = 2) \approx 1$ for all $p' = 1, ..., p$. Furthermore, let $\Pr(\tau_k = 3) \approx 1$ if p is even and let $\Pr(\tau_k = \infty) \approx 1$ if p is odd. Similarly to the general example constructed in Case 1, this environment fails strong efficiency because under the individually optimal play, the correlation in the actions of i_1 and i_l makes k disregard their consensus actions. The alternative play where either i_1 or i_l randomizes equally between 1 and -1 when not receiving their private signals increases the informativeness of their consensus actions sufficiently to strictly benefit k.

A.10 Proof of Proposition 2

Suppose to the contrary that such $\bar{\psi}$ exists. Take r to be the smallest round such that for some $i \in N$, $\Pr(\bar{\psi}_i(\omega) \neq \bar{\psi}^*_i(\omega)|r_i(\tau) = r) > 0$. Since the local environment of i does not change at any conditioning set (M_i, r) for $M_i \subseteq N_i \cup \{i\}$, i can not do strictly better at these conditioning sets than she does under $\bar{\psi}^*$. That is,

$$\Pr(\bar{\psi}^*_i(\omega) = \theta|\theta, M_i(\tau) = M_i, r_i(\tau) = r) \geq \Pr(\bar{\psi}_i(\omega) = \theta|\theta, M_i(\tau) = M_i, r_i(\tau) = r)$$

for all $M_i \subseteq N_i \cup \{i\}$. The equality holds if and only if $\Pr(\bar{\psi}^*_i(\omega) = \bar{\psi}_i(\omega)|M_i(\tau) = M_i, r_i(\tau)) = 1$. The reason is that when i is indifferent between $a_i = 1$ and $a_i = -1$, the only symmetric response is to randomize between the two actions with equal probabilities. The inequality holding as an equality for all $M_i \subseteq N_i \cup \{i\}$ would contradict the choice of r as the smallest round where play differs between $\bar{\psi}^*$ and $\bar{\psi}$. The inequality holding strictly for some $M_i \subseteq N_i \cup \{i\}$ would mean that i is strictly worse off in round r under $\bar{\psi}$, a contradiction.

References

Acemoglu, D., M. A. Dahleh, I. Lobel, and A. Ozdaglar (2011). Bayesian learning in social networks. *The Review of Economic Studies 78*(4), 1201–1236.

Alatas, V., A. Banerjee, A. G. Chandrasekhar, R. Hanna, and B. A. Olken (2016). Network structure and the aggregation of information: Theory and evidence from indonesia. *American Economic Review 106*(7), 1663–1704.

Banerjee, A. V. (1992). A simple model of herd behavior. *The quarterly journal of economics 107*(3), 797–817.

Battigalli, P., A. Francetich, G. Lanzani, and M. Marinacci (2019). Learning and self-confirming long-run biases. *Journal of Economic Theory 183*, 740–785.

Bikhchandani, S., D. Hirshleifer, and I. Welch (1992). A theory of fads, fashion, custom, and cultural change as informational cascades. *Journal of political Economy 100*(5), 992–1026.

Bohren, J. A. (2016). Informational herding with model misspecification. *Journal of Economic Theory 163*, 222–247.

Breza, E., A. G. Chandrasekhar, and A. Tahbaz-Salehi (2018). Seeing the forest for the trees? an investigation of network knowledge. Technical report, National Bureau of Economic Research.

Brown, G. W. (1951). Iterative solution of games by fictitious play. *Activity analysis of production and allocation 13*(1), 374–376.

Dasaratha, K., B. Golub, and N. Hak (2018). Social learning in a dynamic environment. *arXiv preprint arXiv:1801.02042*.

Eyster, E. and M. Rabin (2010). Naive herding in rich-information settings. *American economic journal: microeconomics 2*(4), 221–43.

Fudenberg, D. and D. M. Kreps (1993). Learning mixed equilibria. *Games and Economic Behavior 5*(3), 320–367.

Guarino, A. and P. Jehiel (2013). Social learning with coarse inference. *American Economic Journal: Microeconomics 5*(1), 147–74.

Klibanoff, P., M. Marinacci, and S. Mukerji (2005). A smooth model of decision making under ambiguity. *Econometrica 73*(6), 1849–1892.

Li, W. and X. Tan (2020). Locally bayesian learning in networks. *Theoretical Economics 15*(1), 239–278.

Lobel, I. and E. Sadler (2015). Information diffusion in networks through social learning. *Theoretical Economics 10*(3), 807–851.

McBride, M. (2006). Imperfect monitoring in communication networks. *Journal of Economic Theory 126*(1), 97–119.

McBride, M. (2008). Position-specific information in social networks: Are you connected? *Mathematical Social Sciences 56*(2), 283–295.

Sethi, R. and M. Yildiz (2016). Communication with unknown perspectives. *Econometrica 84*(6), 2029–2069.

Sethi, R. and M. Yildiz (2019). Culture and communication. *Available at SSRN 3263560*.

Smith, L. and P. Sørensen (2000). Pathological outcomes of observational learning. *Econometrica 68*(2), 371–398.

Wiseman, T. (2009). Sequential choice and non-bayesian observational learning. *International Game Theory Review 11*(03), 285–300.