Relationship Trading in OTC Markets*

TERRENCE HENDERSHOTT†
Haas School of Business
University of California, Berkeley
Berkeley, CA 94720

DAN LI‡
Board of Governors of the Federal Reserve System
Washington, DC 20551

DMITRY LIVDAN§
Haas School of Business
University of California, Berkeley
Berkeley, CA 94720

NORMAN SCHÜRHOFF¶
Faculty of Business and Economics and Swiss Finance Institute
Université de Lausanne
CH-1015 Lausanne, Switzerland

February 14, 2018

*We thank Darrell Duffie, Jerôme Dugast, Burton Hollifield, Haoxiang Zhu, and seminar audiences at California Polytechnic University, Carnegie Mellon University, FED Board, Higher School of Economics Moscow, IDC Herzliya, Imperial College, Michigan State University, National Bank of Belgium, NY FED, Rice University, University of Houston, University of Illinois-Chicago, University of Oklahoma, University of Wisconsin at Madison, University of Victoria, WU Vienna, and the 2015 Toulouse Trading and post-trading conference, 2016 AFA, 2016 NBER AP Meeting, 2016 SFI Research Day, 12th Annual Central Bank of France Workshop on the Microstructure of Financial Markets, 2016 LAEF Conference on OTC Markets, 2017 OTC Markets and Their Reform Conference in Rigi for valuable comments and feedback.

†Tel.: (510) 643-0619; fax: (510) 643-1412; *email*: hender@haas.berkeley.edu

‡Tel.: (202) 452-2626; *email*: dan.li@frb.gov

§Tel.: (510) 642-4733; fax: (510) 643-1412; *email*: livdan@haas.berkeley.edu

¶Tel.: +41 (21) 692-3447; fax: +41 (21) 692-3435; *email*: norman.schuerhoff@unil.ch. Schürhoff is Research Affiliate of the CEPR. He gratefully acknowledges research support from the Swiss Finance Institute and the Swiss National Science Foundation under Sinergia project CRSII1_154445/1, "The Empirics of Financial Stability."

Relationship Trading in OTC Markets

Abstract

Relations are known to matter in banking, but not in trading. We examine the network of trading relations between insurers and dealers in the over-the-counter corporate bond market. Comprehensive regulatory data show that every third insurer uses a single dealer while a small fraction of insurers has networks of up to 40 dealers. Execution prices are non-monotone in network size, strongly declining for non-exclusive relations with more dealers but then increasing once networks exceed 20 dealers. To understand these facts we build a model of decentralized trade in which insurers trade off the benefits of repeat business against dealer competition. The model quantitatively fits the distribution of insurers' network sizes and how prices depend on insurers' trading frequency. Counterfactual analysis shows that proposed regulations to unbundle or abolish client-dealer relations may decrease welfare.

JEL Classification: G12, G14, G24

Key words: Over-the-counter market, corporate bond, trading relationship, trading cost, liquidity, financial network, unbundling

Insurance companies are vital for a well-functioning society. They provide coverage in the form of compensation resulting from loss, damage, injury, treatment or hardship in exchange for premium payments. To minimize costs, insurance companies invest in a variety of financial assets. Corporate bonds comprise almost 70% of their investments with a total value close to \$3.8 trillion. To facilitate all types of on-demand compensations, including the very large ones, insurance companies need to be able to liquidate their holdings fast and without incurring large transaction costs (Koijen and Yogo, 2015). However, corporate bonds are traded on over-the-counter decentralized markets with more than 400 active broker dealers where insurance companies have to search for best execution. These markets are considered to be poorly functioning due to lack of transparency and fragmentation imposing search frictions (Duffie, Garleanu, and Pedersen, 2005, 2007; Weill, 2007; Vayanos and Wang, 2007). Little empirical evidence exists on whether insurers shop around randomly or build long-term relations with dealers to cope with the market frictions and how the heterogeneity in trading needs across insurers impacts their dealer choices and the execution terms they receive.

We study insurers' choice of trading networks and the corresponding execution prices in the corporate bond market. Regulatory data provide information on the trading relations between more than 4,300 insurers and their dealers for all transactions from 2001 to 2014. We first empirically examine insurers' choice of trading networks and how these relate to transaction prices. Insurers form small, but persistent dealer networks. Figure 1 provides two examples of insurer-dealer trading relations over time. Panel A shows buys and sells for an insurer repeatedly trading exclusively with a single dealer. Panel B shows an insurer trading with multiple dealers over time. Roughly 30% of insurers trade with a single dealer annually. A small fraction of insurers trades with up to 40 dealers in a year. The overall degree distribution follows a power law with exponential tail starting at about 10 dealers. We estimate trading costs as a function of network size N. Costs are non-monotone in N; costs decline with N for small networks and then increase once N exceeds 20 dealers.

Our evidence provides important insights into which models of trade in OTC markets better describe the empirical evidence on client-dealer relations and trading costs. In random search models clients repeatedly search for best execution without forming a finite network of dealers (Duffie et al. 2005, 2007; Lagos and Rocheteau, 2007, 2009; Gavazza, 2016). The empirical fact that insurers form finite dealer networks suggests that adding dealers must be costly for insurers. Traditional models of strategic search, e.g., Stigler (1961), assume each additional dealer imposes a fixed cost on insurers. Insurers add dealers to improve prices up to the point where the marginal benefit equal the fixed cost. This leads to prices improving monotonically in network size, which is inconsistent with the empirical non-monotonicity of trading costs as a function of network size.

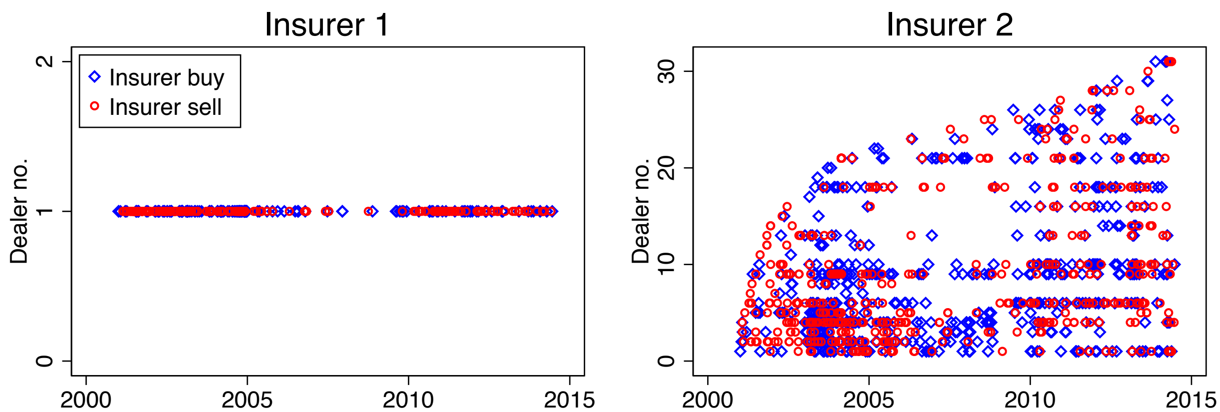


Figure 1: Example of dealer-client trading relations

The figure shows the buy (blue squares) and sell (red circles) trades of two insurance companies with different dealers. We sort the dealers on the vertical axis by the first time they trade with the corresponding insurance company.

To rationalize the empirical evidence we build a model of decentralized trade in which insurers establish relationships with multiple dealers and trade off the benefits of repeat business against inter-dealer competition. We model the full relationship between insurers and dealers as having two components. The first component is the repeat trading interactions between each insurer and the dealers. The insurer repeatedly buys and sells bonds from dealers in her network. This future repeat trading is taken into account by the insurer and the dealers when negotiating terms of each transaction. The second component of the relationship captures all other business between insurers and dealers unrelated to bond trading. It includes transacting in other securities, the ability to purchase newly issued securities, as well as other soft dollar and non-monetary transfers such as investment research.

In our model a single console bond trades on an inter-dealer market which clients can only access through dealers. Dealers have search intensity λ and upon trading with a client transact at the competitive inter-dealer bid and ask prices. Clients initially start without a bond but stochastically receive trading shocks with intensity η which cause them to simultaneously contact N dealers to buy; η and, hence, N vary across clients. The client's effective search intensity is $\Lambda = N\lambda$. The first dealer to find the bond captures all benefits from the transaction. Thus, our trading mechanism is identical to repeated winner-takes-all races (Harris and Vickers (1985)). The bond's purchase price is set by Nash bargaining. Once an owner, the client stochastically receives a liquidity shock forcing her to sell the bond. The mechanics of the sell transaction are the same as the buy transaction. Both dealers and clients derive value from repeat transactions, leading to price improvement for more frequent clients in Nash bargaining.

Existing OTC models provide predictions about network size or prices, but not both. Random search models assume investors may contact every other counterparty. Other models allow investors to choose specific networks or markets, but exogenously fix the structure of those networks. For instance, in Vayanos and Wang (2007) investors chose to search for a counterparty between two markets for the same asset: a large market with faster execution but higher transaction costs, and a small market with slower execution but lower transaction costs. Neklyudov and Sambalaibat (2016) use a similar setup as in Vayanos and Wang (2007) but with investors choosing between dealers with either large or small interdealer networks instead of asset markets. By contrast, both network sizes and transaction prices are endogenously determined in the equilibrium of our model.

Our analysis delineates the trade-offs leading to optimal network size and prices. Clients trade off the value of repeat relations with dealers against the benefits of competition among dealers. The benefits of intertemporal dealer competition lead to better prices and faster execution. However, the value of repeat relations declines in the number of dealers. Dealers compensate for losses from repeat business by charging higher spreads. Eventually the costs of having larger network outweigh the benefits and the dealers' spread starts to increase with the network size. This corresponds to the empirical non-monotonicity in trading costs with respect to the network size. The value of repeat relations diminishes more slowly with the addition of dealers for clients with larger trading intensity as dealers compete for larger repeat business. Therefore, these larger clients use more dealers and get better execution as benefits from having larger repeat business outweigh the costs of having larger network.

The model can quantitatively match the cross-section of insurers' observed network sizes and transaction prices. Doing so requires structural estimation of the model parameters not directly observable in the data, Θ. The clients' trading intensity, η, is the one parameter for which we observe its cross-sectional distribution, $p(\eta)$, in the data. Insurer n's trading shock intensity η_n can be estimated by the average number of bond purchases per year over the sample period. We estimate η_n separately for each insurer using multiple years of trade data, which enables us to construct $p(\eta)$. The model provides the optimal network size, $N^*(\Theta, \eta_n)$, for each client n as a function of its trading intensity η_n. The model predictions allow us estimate the unobservable model parameters Θ by matching the model-implied distribution of N_n^*'s, $p(N^*)$, to their empirical counterparts.

Using the structurally estimated parameters along with the distribution of trading intensities quantitatively reproduces both the empirical distribution of network sizes and the dependence of trading costs on network size found in the data. The model estimates reasonable unobserved parameters: Dealers can find the bond within a day or two, and insurers' average holding period ranges from two to four weeks. The sunk cost of obtaining quotes from a dealer is small, but investors' estimated willingness to pay for immediacy is high. Allowing dealers' bargaining power to decrease with insurers' trading intensity improves the model's fit. Dealers' bargaining power when investors are forced to sell is high and relatively insensitive to the insurers' trading frequency. In contrast, dealers' bargaining power when buying declines significantly with the insurers' trading frequency.

The regulatory authorities in the U.S. and Europe have put different emphasis on how to address the frictions and improve the functioning of OTC markets. Regulatory initiatives in the U.S., including the Dodd-Frank Act, encourage dealer competition through fostering of multilateral electronic trading platforms, with which client-dealer long-term relations would become obsolete. In Europe, the Markets in Financial Instruments Directive (MiFID) II introduces fundamental changes to the market structure by requiring the unbundling of trading and non-trading dealer services, which are supposed to be priced and sold separately. Little evidence exists either empirically or theoretically on the welfare implications of the proposed regulations on dealers and their clients.

Our model provides a convenient setting to quantify the effects of the proposed regulatory changes on trading costs and the value of repeat relations. We perform counterfactual analysis in which we either reduce the probability of repeat trading with the same dealer or we eliminate the non-trade value of relationships. Our results indicate that all types of insurers will incur higher transaction costs when the probability of repeat trading with the same dealer is reduced—all else equal—through, e.g., anonymous electronic trading facilities.[1] The impact on various types of insurance companies, including life, health, and property and casualty insurers, is different. Insurers that trade more frequently and, therefore, have larger networks will see a smaller increase in transaction costs than insurers who trade less frequently and tend to trade repeatedly with 1 to 5 dealers.

Decoupling trade and non-trade insurer-dealer business, as stipulated by MiFID II regulations, will reduce the optimal network sizes and decrease transaction costs for all insurers except for the least active ones. This is due to the fact that trading networks are currently in excess of what is required to minimize trading costs. More importantly, the proposed regulations can potentially reduce trading activity and decrease insurers' welfare, especially for those that trade infrequently. In addition, the regulations may increase the disparities between dealers and foster a "dog-eat-dog" selection among the dealers.

[1] Our counterfactual analysis takes the current decentralized market structure as given. It cannot be used to answer what would happen if, say, the trading mechanism was centralized. The latter is an interesting and important topic left for future research.

Relation to literature: Our paper complements the empirical literature on the microstructure of OTC markets and its implications for trading, price formation, and liquidity. Edwards, et al. (2005), Bessembinder et al. (2006), Harris and Piwowar (2007), Green et al. (2007) document the magnitude and determinants of transaction costs for investors in OTC markets. Our paper deepens the understanding of OTC trading costs by using the identities of all insurers along with their trading networks and execution costs. These help explain the substantial heterogeneity in execution costs observed in these studies. O'Hara et al. (2015) and Harris (2015) examine best execution in OTC markets without formally studying investors' optimal network choice.

There exists an empirical literature on the value of relationships in financial markets. Similar to our findings, Bernhardt et al. (2004) show that on the London Stock Exchange broker-dealers offer greater price improvements to more regular customers.[2] Bernhardt et al. (2004) do not examine the client-dealer networks and in their centralized exchange setting quoted prices are observable. Afonso et al. (2013) study the overnight interbank lending OTC market and find that a majority of banks in the interbank market form long-term, stable and concentrated lending relationships. These have a significant impact on how liquidity shocks are transmitted across the market. Afonso et al. (2013) do not formally model the network and do not observe transaction prices. DiMaggio et al. (2015) study inter-dealer relationships on the OTC market for corporate bonds while our focus is on the client-dealer relations.[3]

The role of the interdealer market in price formation and liquidity provision are the focus in Hollifield et al. (2015) and Li and Schürhoff (2015). These studies explore the heterogeneity across dealers in their network centrality and how they provide liquidity and what prices they charge. By contrast, we focus on the heterogeneity across clients and how trading intensity affects their networks and transaction prices.

The search-and-matching literature is vast. Duffie et al. (2005, 2007) provide a prominent treatment of search frictions in OTC financial markets, while Weill (2007), Lagos and Rocheteau (2007, 2009), Feldhütter (2011), Neklyudov (2014), Hugonnier et al. (2015), Üslü (2015) generalize the economic setting. These papers do not focus on repeat relations and do not provide incentives to investors to have a finite size network. Gavazza (2016) structurally estimates a model of trading in decentralized markets with two-sided one-to-one search and

[2] For a comprehensive theoretical model of loyalty see Board (2011).

[3] Our paper also relates to a growing literature studying trading in a network, e.g., Gale and Kariv (2007), Gofman (2011), Condorelli and Galeotti (2012), Colliard and Demange (2014), Glode and Opp (2014), Chang and Zhang (2015), Atkeson et al. (2015), Babus (2016), Babus and Hu (2016), and Babus and Kondor (2016). These papers allow persistent one-to-one dealer-client relationships, while the main focus of our model is on clients' networks.

bilateral bargaining using aircraft transaction data. At the market-wide level Gavazza (2016) quantifies the effects of market frictions on prices and allocations. We use the structural estimation of our one-to-many search-and-match model to quantify the effects of client-dealer relations on execution quality in the OTC market for corporate bonds.

Directed search models allow for heterogeneous dealers and investors, as well as arbitrary trade quantities. These typically rely on a concept of competitive search equilibrium proposed by Moen (1997) for labor market. Examples include Guerrieri, Shimer, and Wright (2010), Guerrieri and Shimer (2014), and Lester et al. (2015). These papers explain assortative matching between clients and dealers and show how heterogeneity affects prices and liquidity. However the matching technology employed by these papers is one-to-one, thus limiting the network size to a single dealer.

The remainder is organized as follows. Section 1 describes the data. Sections 2 and 3 document our findings on insurer trading activity, dealer networks, and trading costs. In Section 4 we develop an OTC model with repeat trade and execution competition. We test its predictions and provide counterfactuals in Section 5. Section 6 concludes.

1 Data

Insurance companies file quarterly reports of trades of long-term bonds and stocks to the National Association of Insurance Commissioners (NAIC). For each trade the NAIC data include the dollar amount of transactions, par value of the transaction, insurer code, date of the transaction, the counterparty dealer name, and the direction of the trade for both parties, e.g., whether the trade was an insurance company buying from a dealer or an insurance company selling to a dealer. The NAIC data preclude intraday analysis as the trades do not include time stamps of the trades. To focus on secondary trading we only include trades more than 60 days after issuance and trades more than 90 days to maturity.

Our final sample covers all corporate bond transactions between insurance companies and dealers reported in NAIC from January 2001 to June 2014. We supplement the NAIC data with a number of additional sources. Bond and issuer characteristics come from the Mergent Fixed Income Security Database (FISD) and Lipper eMAXX. Insurer financial characteristics come from A.M. Best and SNL Financial. The final sample contains 506 thousand insurer buys and 497 thousand insurer sells.

Table 1 reports descriptive statistics for the corporate bond trades (Panel A) and insurers (Panel B) in our 2001-2014 sample. There are 4,324 insurance companies in our sample. Insurance companies fall into three groups based on their product types: (i) Health, 617 companies (14% of the sample); (ii) Life, 1,023 companies (24% of the sample); (iii) P&C

2,684 companies (62% of the sample). Health insurance companies account for 16.3% of trades and 4.4% of yearly trading volume. They trade on average with 6.59 dealers each year. Life insurance companies account for the majority 46.9% of trades and 70.4% of yearly trading volume. They trade on average with 8.06 dealers each year. P&C insurance companies comprise 36.8% of trades and 25.2% of yearly trading volume. They trade on average with 4.81 dealers each year.

The distribution of trading activity is skewed with the top ten insurance companies accounting for 6.3% of trades and 14.3% of trading volume. They use almost 30 dealers which is much higher than the sample average of 5.83 dealers per insurer. The top 100 insurers account for 27.8% of trades as well as for 45.3% of trading volume. The 3,000 smallest insurers use on average 3.76 dealers.

Order splitting is not very common. Larger transactions are cheaper to execute, thus eliminating incentives for order splitting. On just 1.2% of all insurer-days (13K out of 1.1M trades) an insurer trades the same bond multiple times.

Insurers trade in a variety of corporate bonds. The average issue size is quite large at $917 million and is similar across insurer's buys and sells. The average maturity is nine years for insurer buys and eight years for insurer sells. Bonds are on average 2.88 years old with sold bonds being a little older at 3.09 years. Finally, 75% of all bonds trades are in investment grade while only 1% are in unrated with the remainder being high yield. Privately placed bond trades form a small minority of our sample at 8%.

The risk-based-capital (RBC) ratio measures an insurer's capital relative to the riskiness of its business. The higher the RBC ratio, the better capitalized the firm. Insurer size is reported assets. The cash-to-asset ratio is cash flow from the insurance business operations divided by assets.

Overall, there exists a large degree of heterogeneity on the client side. Insurance companies buy and sell large quantities of different corporate bonds and execute these transactions with the number of dealers ranging on average from one to as many as 40.

2 Empirical Evidence on Insurer Trading Networks

This section empirically characterizes insurers' trading intensity and the size of their trading networks. We investigate the determinants of both the extensive margin, i.e., the number of trades, and intensive margin, i.e., the total dollar volume traded, of insurer trading in a given year. Both margins reveal that insurers have heterogeneous trading needs.

Table 1: Descriptive statistics

The table reports descriptive statistics for trades (Panel A) and insurers (Panel B) in our sample from 2001 to 2014. Panel A reports the average across all trades over the sample period. Panel B reports the yearly average across insurers.

	Panel A: Trades		
	All trades	Insurer buys	Insurer sells
No. of trades (k)	1,003	506	497
Trade par size ($mn)	1.80	1.73	1.87
Bond issue size ($mn)	916.66	921.37	911.87
Bond age (years)	2.88	2.67	3.09
Bond remaining life (years)	8.54	8.94	8.13
Private placement (%/100)	0.08	0.08	0.07
Rating (%/100)			
IG	0.74	0.76	0.72
HY	0.25	0.23	0.28
Unrated	0.01	0.01	0.01
	Panel B: Insurers (N= 4,324)		
	Volume ($mn)	No. of trades	No. of dealers
All insurers	17.32	9.52	5.83
Insurer type			
Health (617, 14%)	10.66	21.74	6.59
Life (1,023, 24%)	103.00	37.71	8.06
P&C (2,684, 62%)	14.08	11.29	4.81
Insurer activity			
Top 10	2,111.88	517.92	29.89
11-100	509.49	233.04	22.07
101-1000	75.66	46.22	11.56
1001+	3.80	4.24	3.76
Insurer characteristics:	Mean (SD)		
Insurer size	4.97 (0.90)		
Insurer RBC ratio	3.36 (0.35)		
Insurer cash-to-assets	3.49 (10.79)		
Life insurer	0.24 (0.42)		
P&C insurer	0.62 (0.48)		
Insurer rated A-B	0.37 (0.38)		
Insurer rated C-F	0.01 (0.07)		
Insurer unrated	0.53 (0.39)		

2.1 Insurer trading activity

We start with univariate analysis. The majority of insurers do not trade often at the annual frequency. About 30% of insurers trade just once per year while 1% of the insurers make at least 25 trades per year. This is consistent with the evidence from Table 1 that while the top 100 insurers constitute just 0.23% of the total sample, they account for as much as 32% of all trades in our sample. The mean number of trades per year is 19, with a median of 14, with several insurers making more than 1,000 trades in some years and up to the maximum of 2,200 trades in a year.

Figure 2 shows the distribution in the average number of trades per year across insurers. A large fraction of insurers do not trade in a given month and we therefore report an annual figure. The annual distributions follow a power law with $p(X) \propto .27 \times X^{-1.21}$ for all insurer

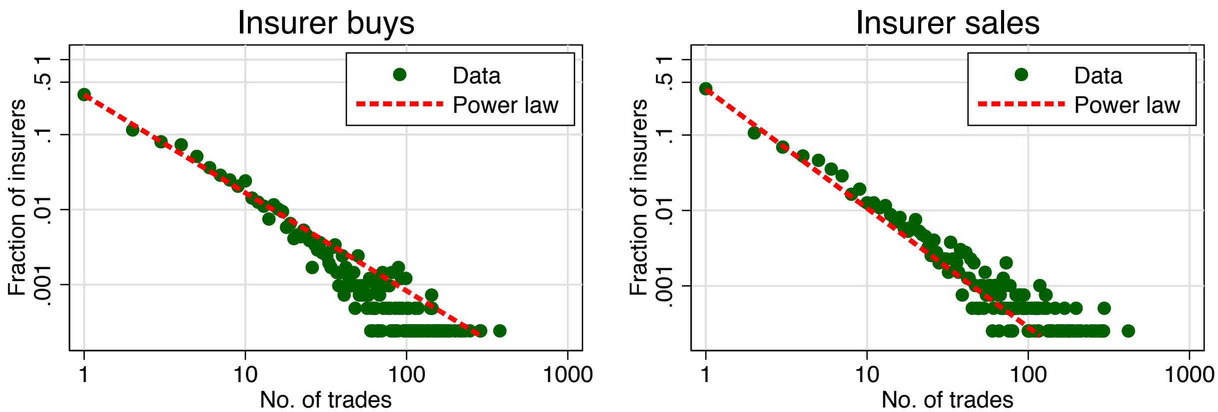


Figure 2: Insurer trading activity

The figure shows the distribution in the number of insurer buys per year (left) and insurer sales (right). We use a log-log scale.

trades combined. The power law is $.34 \times X^{-1.31}$ for insurer buys (depicted in Panel A) and $.40 \times X^{-1.58}$ insurer sales (Panel B). Visually the two power law distributions for insurer buys and sales in Figure 2 look similar. This suggests insurers buy and sell at similar rates, even though these rates vary significantly across insurers.

We next examine what characteristics explain the heterogeneity in trading intensities. Table 2 documents the determinants of the intensive margins (trading volume in $bn, column (1)) and extensive margins (number of trades, column (2)) of the annual trading by insurance companies using pooled regressions with time fixed effects. The specification consists of the trade par size, insurer and bond characteristics, as well as the variation in the trade size and bond characteristics across all trades of the insurer during the year. The variables capturing the variation in characteristics capture complexity in insurers' portfolios and their need for more frequent rebalancing and for dealer specialization. Insurer characteristics include its size, cash-to-assets ratio, type, RBC ratio, and rating. Bond and trade characteristics include size, age, maturity, rating, a private placement dummy, and trade size. Insurer size, RBC ratio, and the dependent variables are log-transformed. All regressors are averaged across all trades of the insurer during the period and lagged by one year.

Our evidence is consistent with insurance companies having persistent portfolio rebalancing needs. Logarithms of both measures of trade intensity are persistent; the coefficient on the lagged log-volume is 0.67 and the coefficient on the lagged log-number-of-trades is 0.76. Both coefficients are statistically significant at 1% levels.

Insurer trading strongly correlates with insurer size, type, and quality, with bond types and bond varieties as these variables explain 79% of the variation in annual trading volume

Table 2: Insurers' trading activity

The determinants of insurance company trading activity are reported. We measure trading activity by the total dollar volume traded in a given year and, alternatively, by the number of trades over the same time horizon. All dependent variables are log-transformed by 100*log(1+x). All regressors are averaged across all trades of the insurer during the period and lagged by one time period. Estimates are from pooled regressions with time fixed effects. Standard errors are adjusted for heteroskedasticity and clustering at the insurer and time level. Significance levels are indicated by * (10%), ** (5%), *** (1%).

Determinant	(1) Volume ($mn)	(2) No. of trades
Insurer size	21.95***	14.51***
Insurer RBC ratio	-1.53	-4.68***
Insurer cash-to-assets	0.27***	0.26***
Life insurer	4.96***	0.27
P&C insurer	-1.06	-3.89**
Insurer rated A-B	5.13**	5.80***
Insurer rated C-F	1.97	-0.43
Insurer unrated	6.39***	5.79**
Trade par size	-3.62***	-3.22***
Bond issue size	-0.00	0.00
Bond age	-0.79***	-1.25***
Bond remaining life	-0.04	-0.05
Bond high-yield rated	4.62***	4.65***
Bond unrated	-6.67	-12.81*
Bond privately placed	-5.37	-3.56
Variation in trade size	4.50***	1.52***
Variation in issue size	0.00	0.00
Variation in bond age	0.37	0.39
Variation in bond life	0.52***	0.65***
Variation in bond rating	-0.35	0.04
Variation in rated-unrated	39.60*	6.16
Variation in private-public	1.62	-1.60
No varieties traded	9.23***	13.64***
Lagged volume	0.67***	
Lagged no. of trades		0.76***
Year fixed effects	Yes	Yes
R^2	0.789	0.646
N	30,029	30,029

and 65% variation in annual number of trades. A ten-fold increase in insurer's size increases trading volume by $2.2 billion. Larger insurance companies and insurers with higher cash-to-assets ratio also trade more often and submit larger orders. Insurers with higher RBC ratios trade less often than insurers with low RBC ratios. Both margins of trading increase with the insurer's rating, i.e., insurers with the lowest rating (C-F) trade less than higher-rated insurers. Life insurers tend to submit larger orders.

Both margins of bond turnover increase as bond ratings decline; lower rated bonds are traded more often and in larger quantities. Insurers tend to trade privately placed bonds less since potentially they just own fewer of them than publicly placed bonds. Both margins of bond turnover decline with par size and bond age indicating that the majority of insurers are long-term investors. Neither measure of trade intensity depends on bond issue size and remaining life as their coefficients are not statistically significant.

Finally, both trading volume and the number of trades decline if an insurer trades more bond varieties. However, a specific variety can have an opposite effect on the trading intensity. For instance, both measures of trading intensity increase with variation in bond rating and bond life. This is consistent with insurers increasing trading intensity when rebalancing their portfolios, i.e. shifting from high-yield to investment-grade bonds or from younger to older bonds.

Overall, the analysis reveals large heterogeneity in trading intensities across different insurers. Trading intensity depend on the variety of bonds traded, bond specific characteristics, and the insurer type and quality. We now turn to how these characteristics affect the insurers' choice of dealer network.

2.2 Properties of insurer networks

The previous section's results demonstrate how insurer characteristics explain the intensity of their trading. There is large degree of heterogeneity in the insurers' trading intensity, with some insurers trading twice per day while others trade just once per year. This section studies how many dealers they trade with over time and how persistent are these networks. While insurers should have heterogeneous demands for their dealer network depending on their trading intensity, how concentrated and persistent their trading is reveals basic network formation mechanisms.

We start with the examples of the insurer-dealer relationship depicted in Figure 1. These show that insurers do not trade with a dealer randomly picked from a large pool of corporate bonds dealers. Instead, insurers buy from the same dealers that they sell bonds to and they engage in long-term repeat, but non-exclusive, relations. We analyze how representative are the examples in Figure 1 and how insurer characteristics determine their network size.

Figure 3 plots the degree distribution across insurers by year, i.e., the fraction of insurers trading on average across all years with the given number of dealers, using a log-log scale. The figure shows insurers trade with up to 31 dealers every year, with some trading with as many as 40 but these represent less than 1/10,000 of the sample. Exclusive relations are dominant with almost 30% of insurers trading with a single dealer in a given year.[4] The degree distributions in Figure 3 follow a power law with exponential tail starting at about

[4]Because of the number of insurers that do not trade in a year Figures 2 and 3 are not directly comparable. Figure 2 plots the average number of trades per year, rounded to the next integer and winsorized from below at 1. For about half of the insurer-years there is no trade by an insurer. Figure 3 does not impute a zero for a given year if the insurer does not trade. Hence, 27.7% of insurers use a single dealer in a year when they trade, while 14.8% of insurers trade just once in a year. Thus, many insurers who trade more than once in a year use a single dealer.

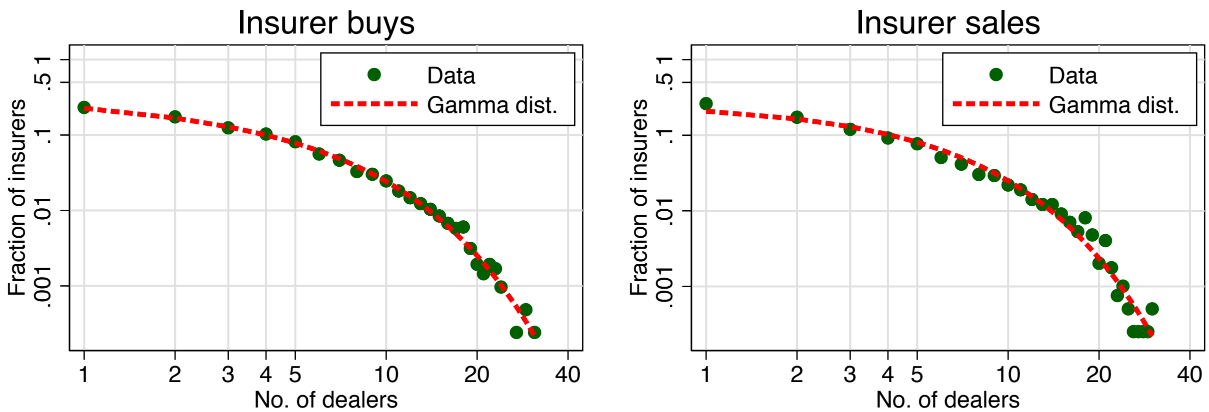


Figure 3: Size of insurer-dealer trading networks

The figure shows the degree distribution for insurer-dealer relations by year for insurer buys (left) and insurer sales (right). We use a log-log scale.

10 dealers. This is consistent with insurers building networks that they search randomly within. Fitting the degree distribution to a Gamma distribution by regressing the log of the probabilities of each N on a constant, the logarithm of N, and N yields the following coefficients:

$$\begin{aligned} \text{For all insurer trades combined:} \quad & p(N) \propto N^{.15}e^{-.20N}, \\ \text{For insurer buys:} \quad & p(N) \propto N^{-.12}e^{-.22N}, \\ \text{For insurer sales:} \quad & p(N) \propto N^{.01}e^{-.24N}. \end{aligned} \tag{1}$$

Table 3 reports the determinants of insurer dealer network sizes using pooled regressions with time fixed effects. We measure the size of the trading network by the number of dealers that an insurance company trades with in a given year. We log-transform all dependent variables by $100 * log(1 + x)$ and average all regressors across all trades by the same insurer during the year and lag them by one time period. We perform the estimation on the whole sample (Column (1)) and, in order to examine how these vary with insurer's size, on subsamples of small and large insurers based on asset size. We classify an insurer as small if it falls in the bottom three size quartiles and, respectively, as large if it falls in the top quartile of the size distribution.

Column (1) indicates that insurer size and type, bond characteristics, and bond varieties matter for the size of the dealer network. Large insurers, which Table 2 shows have larger trading intensity, trade with more dealers. Insurers with demand for larger bond variety have larger networks even controlling for their size (column (2) and (3)). Higher quality insurers,

Table 3: Size of insurers' trading network

The table reports the determinants of the size of insurers' trading network. We measure the size of the trading network by the number of different dealers that an insurance company trades with in a given year. See caption of Table 2 for additional details. Standard errors are adjusted for heteroskedasticity and clustering at the insurer and time level.

Determinant	(1) All insurers	(2) Small insurers	(3) Large insurers
Insurer size	9.95***	7.93***	5.04***
Insurer RBC ratio	-3.58***	-3.64***	0.11
Insurer cash-to-assets	0.15***	0.15***	0.11**
Life insurer	-0.13	0.75	-2.77**
P&C insurer	-2.60***	-1.20	-3.59***
Insurer rated A-B	4.12***	6.00***	0.13
Insurer rated C-F	-1.54	-0.18	-5.42***
Insurer unrated	3.05*	3.98*	-2.86**
Trade par size	-1.92***	-1.23***	-1.24***
Bond issue size	-0.00	-0.00	0.00
Bond age	-0.91***	-0.77***	-1.10***
Bond remaining life	-0.08	-0.11*	0.10
Bond high-yield rated	1.23	-2.93*	2.56
Bond unrated	-13.74***	-12.50***	-7.73
Bond privately placed	-3.85	1.10	-10.22
Variation in trade size	0.51	-0.10	0.81*
Variation in issue size	0.00	0.00	0.00
Variation in bond age	0.50**	0.56**	0.16
Variation in bond life	0.50***	0.41***	0.11
Variation in bond rating	0.05	0.19	0.36*
Variation in rated-unrated	3.60	-11.38	-24.50
Variation in private-public	1.20	-4.33	3.17
No varieties traded	5.88***	2.71***	12.50**
Lagged no. of dealers	0.75***	0.62***	0.75***
Year fixed effects	Yes	Yes	Yes
R^2	0.614	0.350	0.593
N	30,029	18,033	11,996

i.e., insurers with higher cash-to-assets ratio and higher ratings, have larger networks, but this matters only for smaller insurers as column (2) indicates. This is potentially due to it being cheaper for a dealer to set up a credit account for higher quality insurers. These factors matter more for small insurers because they face larger adverse selection problems in forming permanent links with dealers. Insurers with greater variety in the bond they trade have larger networks. Overall these findings suggest insurers' network choice is endogenous and dependent on multiple factors. Competition and specialization jointly determine investors' trade choices.

Table 3 shows persistence in the size of the network with the coefficient on the lagged network size being 0.75 (column (1)). This result is mostly due to large insurers, because this coefficient equals only 0.62 for small insurers. Table 4 examines this in more detail by reporting statistics for the frequency with which insurers adjust their network size. We compute the likelihood that an insurer uses a certain number of dealers in a given year and

Table 4: Persistence in insurers' trading network

The table reports switching probabilities, $p(\text{No. of dealers in } t+1 | \text{No. of dealers in } t)$, for using a network size conditional on the insurer's past behavior for each year.

	No. of dealers next year			
No. of dealers this year	1	2-5	6-10	>10
1	0.61	0.30	0.06	0.03
2-5	0.20	0.54	0.20	0.06
6-10	0.06	0.31	0.40	0.24
>10	0.01	0.07	0.17	0.75

compare it to the corresponding number in the following year. The transition probabilities are reported in Table 4. Trading relations are persistent from year to year. This is especially true for exclusive relations as the probability of staying with a single dealer each year is equal to 0.61. Insurers with more than one dealer are unlikely to switch to a single dealer as the annual switching probabilities are equal to 0.20 for insurers with 2 to 5 dealers and 0.06 for insurers with 6 to 10 dealers. Insurers with the largest networks (> 10 dealers) tend to maintain large networks over time, with a 75% probability of staying with a large network. The distribution of insurers shown in Figure 3 together with the stable network sizes are difficult to reconcile with a "pure" random search model à la Duffie et al. (2005, 2007).

One potential concern is that the relationship between the size of the insurer and the size of its network arises mechanically. Small insurers trade once or twice per year and thus need only a single dealer, while large insurers need to execute many trades and, therefore, use many dealers. As a consequence, small exclusive networks are pervasive simply because they are used by small insurers. To alleviate this concern we plot in Figure 4 the number of insurer buys (left) and sales (right) per dealer and year as a function of the network size. The red line is a fitted non-linear lowess model between $log(\#\text{ of trades per dealer})$ and $log(N)$. The relationship has an inverse hump-shape. Insurers executing four trades per year are equally likely to use a single dealer or split them between two dealers. In addition, a large number of insurers execute between 10 and 100 trades with a single dealer. A dealer can facilitate repeated trading with the same insurer by offering other business unrelated to bond trading. This can include trading in other securities, access to new issues, investment research, and others. Since our data does not allow to control for these non-trade benefits from repeat relations, we use a structural model to quantify them.

The next section studies the relation between client-dealer networks and execution costs.

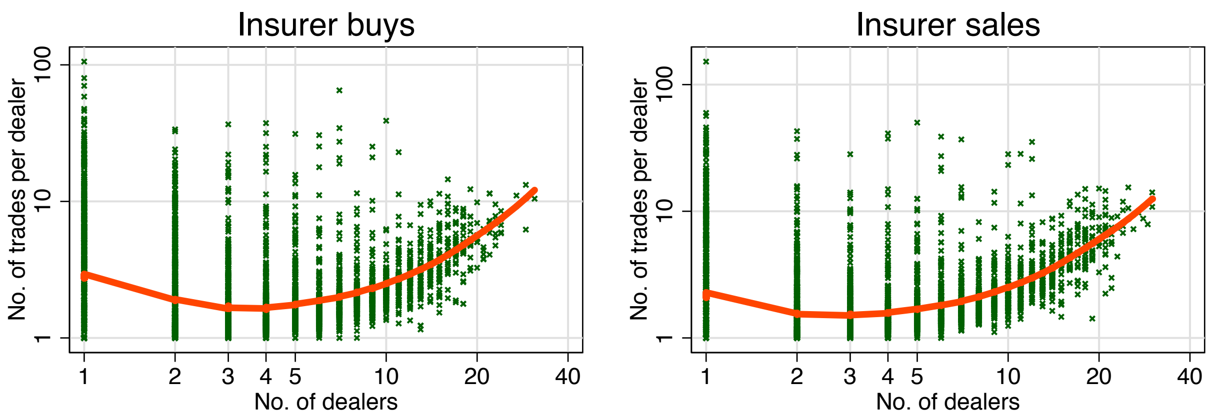


Figure 4: Insurer trading activity and network size

The figure shows the number of insurer buys per year per dealer (left) and insurer sales per year per dealer (right). The red line is a fitted non-linear model. We use a log-log scale.

3 Insurer trading costs and networks

Tables 2 and 3 suggest that bond characteristics impact insurers' trading intensity. To control for bond, time, and bond-time variation, we compare transaction prices to daily bond-specific Bank of America-Merrill Lynch (BAML) bid (sell) quotes. BAML is the largest corporate bond dealer, transacting with more than half of all insurers for almost 10% of both the trades and volume. The BAML (bid) quotes can be viewed as representative quotes for insurer sales and enable us to measure prices relative to a transparent benchmark price. The BAML quotes essentially provide bond-time fixed effects, which would be too numerous to estimate in our sample. Our relative execution cost measure in basis points is defined as

$$\text{Execution cost (bp)} = \frac{\text{BAML Quote} - \text{Trade Price}}{\text{BAML Quote}} * (1 - 2 * \mathbf{1}_{Buy}) * 10^4, \qquad (2)$$

where $\mathbf{1}_{Buy}$ is an indicator for whether the insurer is buying or selling. Because some quotes may be stale or trades misreported, leading to extreme costs estimates, we winsorize the distribution at 1% and 99%.

Execution costs depend on the bond being traded, time, whether the insurers buys or sells, the insurer's characteristics, dealer identity and characteristics, and the insurer network size. To examine the relationship-specific effects on execution costs we control for bond and timefixed effects. In principle if the BAML perfectly controls for bond-time effects, the additional bond and time fixed effects are unnecessary.

The relationship component of transaction costs depends on the properties of the insurers' networks. Figure 3 and equation (1) indicate that the degree distribution for insurer-dealer relations follows a Gamma distribution. Therefore, we include both the size of the network, N, and its natural logarithm, $ln(N)$, as explanatory variables. We control for seasonality using time fixed effects, α_t, and for unobserved heterogeneity using either bond characteristics or bond fixed effects, α_i. The other explanatory variables consist of either insurers' or dealers' characteristics, or both. We estimate the following panel regression for execution costs in bond i at time t:

$$\text{Execution cost}_{it} = \alpha_i + \alpha_t + \beta N + \gamma lnN + \theta X_{it} + \epsilon_{it}. \tag{3}$$

The set of explanatory variables X includes characteristics of the bond, as well as features of the insurer and dealer.

Table 5 provides trading cost estimates from panel regressions. We adjust standard errors for heteroskedasticity and cluster them at the insurer, dealer, bond, and day level. The coefficient on insurer buy captures the average bid-ask spread of roughly 40 basis points. Column (1) of Table 5 shows that execution costs decline with insurer network size. An insurer with an additional dealer has trading cost 0.22 basis point lower. Large insurers pay on average lower execution costs. An insurer with 10 times as many assets has trading cost 3.72 basis points lower. Better capitalized insurers (higher RBC ratio) get better prices. Column (2) adds the logarithm of N to the specification reported in Column (1). The coefficient on N switches from -0.22 reported in Column (1) to 0.32, while the coefficient on the logarithm of N is -6.29. Both coefficients are statistically significant at 1%. This result indicates that the execution costs are non-monotone in the network size. Improvements in execution quality from having a larger dealer network are exhausted at $N = \frac{6.29}{0.22} \approx 20$. Clients with networks of 40 dealers and 10 dealers pay, on average, the same bid-ask spread of 40 basis points. This finding goes against the traditional wisdom that inter-dealer competition improves prices. It is also inconsistent with classic static strategic network formation models (e.g., Jackson and Wolinsky (1996)). In these models a client trades off fixed costs of adding an extra dealer against better execution due to increased dealer competition thus making price a monotonically decreasing function of the network size. In the next section we use this and other network-related empirical evidence to motivate an alternative strategic model of finite network formation in which clients and dealer share the benefits of repeated interactions.

Columns (3) and (4) replace bond and dealer fixed effects with bond and dealer characteristics. NYC-located dealers offer better prices to all insurers and more diversified dealers charge, on average, higher prices. Bond characteristics matter for execution costs as insur-

Table 5: Execution costs and investor-dealer relations

The table reports the determinants of execution costs. Execution costs are expressed in basis points relative to the Merrill Lynch quote at the time of the trade. Standard errors are adjusted for heteroskedasticity and clustered at the insurer, dealer, bond, and day level. See caption of Table 2 for additional details.

Determinant	(1)	(2)	(3)	(4)
Insurer no. of dealers	-0.22***	0.32***	0.32***	0.32***
ln(Insurer no. of dealers)		-6.29***	-6.51***	-6.55***
Insurer size	-3.72***	-3.59***	-3.52***	-3.95***
Insurer RBC ratio	-3.51***	-4.19***	-4.68***	-5.40***
Insurer cash-to-assets	-0.04**	-0.04**	-0.04**	-0.03*
Life insurer	4.43***	4.47***	5.73***	7.21***
P&C insurer	1.72**	1.73**	1.99**	2.82***
Insurer rated A-B	-0.47	0.01	-0.18	-0.50
Insurer rated C-F	11.29*	11.13*	10.90*	12.18*
Insurer unrated	0.74	0.62	0.55	0.14
Insurer buy	39.56***	39.27***	39.71***	40.17***
Trade size × Buy	-0.26***	-0.25***	-0.20**	-0.18*
Trade size × Sell	0.53***	0.50***	0.48***	0.52***
Bond issue size			-0.00***	-0.00***
Bond age			0.58***	0.63***
Bond remaining life			0.81***	0.82***
Bond HY rated			4.54***	4.16***
Bond unrated			-5.96***	-6.53***
Bond privately placed			3.38***	3.26***
Dealer size				-5.37***
NYC dealer				-6.66***
Primary dealer				2.43
Dealer leverage				-5.68**
Dealer diversity				0.41***
Dealer dispersion				0.07
Local dealer				0.23
Dealer distance				-0.20
Dealer leverage missing				-6.41**
Dealer dispersion missing				6.09
Bond fixed effects (16,823)	Yes	Yes	No	No
Dealer fixed effects (401)	Yes	Yes	Yes	No
Day fixed effects (3,375)	Yes	Yes	Yes	Yes
R^2	0.154	0.155	0.103	0.098
N	918,279	918,279	918,987	891,875

ers receive worse prices for special bonds and better prices for bonds with larger issue size. Insurers get better prices on unrated bonds.

The next section uses our evidence on networks and execution costs to motivate our model of the OTC markets.

4 Model

The model is stylized but still rich enough to allow for the structural estimation of its primitives from the NAIC data. We view the full relationship between insurers and dealers as having two components. The first component is the repeat trading interactions between each insurer and the dealers. The insurer repeatedly buys and sells bonds from dealers in her network. This future repeat trading is taken into account by the insurer and the dealers when negotiating terms of each transaction. Since our data can speak directly about it, we model this component explicitly. The second component of the relationship captures all other business between insurers and dealers unrelated to bond trading. In practice it includes transacting in other securities, the ability to purchase newly issued securities, as well as other soft dollar and non-monetary transfers such as investment research. Since we do not directly observe these non-trade relations in our data, we utilize a reduced form approach to model them. We quantify both components of the insurer-dealer relationship when structually estimating the model from the NAIC data.

4.1 Setup and Solution

The economy has a single risk-free perpetual bond paying a coupon flow C. The risk-free discount rate is constant and equal to r, so that the present value of the bond is $\frac{C}{r}$. To model client-dealer repeat trade interactions, we keep several attractive features of Duffie et al. (2005) type models such as liquidity supply/demand shocks on the client side and random search with constant intensity. Following Lester et al. (2015), the bond trades on a competitive market accessible only to dealers.[5] Unlike the frictionless inter-dealer market in Lester et al. (2015), in our model dealers face search frictions as in Duffie et al. (2005). We implicitly assume that the structure of insurer-dealer networks does not affect the interdealer market. Dealers, therefore, buy bonds at an exogenously given price M^{ask} from other dealers and sell it to other dealers at an exogenously given price M^{bid}. A bid-ask spread $M^{ask} - M^{bid} \geq 0$ reflects trading costs or cost of carry.

In the model we use a more generic term "client" instead of insurer. Clients act competitively with respect to other clients by not internalizing other clients' trades in their own decisions.[6] Each client chooses a network of dealers, N, without knowledge of other clients' decisions. When a client wants to buy/sell a bond, she simultaneously contacts all N dealers in her network. Upon being contacted each dealer starts searching the competitive dealer market for a seller/buyer with a search intensity λ.[7] All dealers in the client's network search independently of each other. Therefore, the effective rate at which a client with N dealers in

[5]The interdealer market obviates the need to track where the entire stock of the asset is held at every moment in time.

[6]While we do not have a direct evidence on the information flows across the insurers and between the insurers and dealers regarding trades, it is unlikely that such information is shared even voluntary.

[7]Bessembinder et al. (2016) show that corporate bond dealers increasingly hold less inventory and facilitate trade via effectively acting as brokers by simultaneously buying and selling the same quantity of the same bond.

her network finds a counterparty equals λN. When the client receives a subsequent trading shock all dealers in the network are contacted to reverse the initial transaction.

Each client pays a sunk cost K per transaction.[8] The cost K is any cost of a client contacting dealers, e.g., the time required to make each phone call and any fixed costs required to hire more in-house traders. Clients trading more frequently incur K more often. While K does not depend directly on network size, clients who choose a larger network trade faster, thus, incurring K more often. The cost K enter clients' value functions thus affecting their reservation values in bargaining. Clients' search mechanism can be viewed as a winner-takes-all race with the dealer first to find the bond winning the race. The prize is the spread $P^b - M^{ask}$ when the client buys and $M^{bid} - P^s$ when the client sells, where P^b (P^s) is the price at which the client buys (sells) the bond from (to) the dealer.

Clients transition through ownership and non-ownership based on liquidity shocks. At these transitions clients act as buyers and sellers. The discounted transition probabilities and transaction prices link valuations across the owner, non-owner, buyer, and seller states.

A client i starting as a non-owner with valuation $\widehat{V}^{no}$ is hit by stochastic trading shocks to buy with intensity η_i. η_i varies across clients and its distribution can be directly inferred from the data. Section 2.1 characterizes the compound distribution of trading activity. For the sake of clarity we are going to drop the subscript i throughout our theoretical analysis.

The client contacts her network of N dealers leading to her transiting to a buyer state with valuation $\widehat{V}^b$. In steady state valuations in these two states are related by

$$\widehat{V}^{no} = \underbrace{\widehat{V}^b \frac{\eta}{r+\eta}}_{\text{Value from Trading}} + \underbrace{V^r \frac{r}{r+\eta}}_{\text{Non-Trade Value}} . \tag{4}$$

Here $V^r(\eta, N)$ captures exogenously given relation-specific non-trade flows to the client from her dealers. We are largely agnostic regarding $V^r(\eta, N)$ as a function of the trading intensity and network size, η and N, and let the functional form be determined by the data-driven estimation. We assume, however, that $V^r(\eta, N)$ is a monotonic function of N and satisfies the following Inada condition $\lim\limits_{N\to\infty} \frac{V^r(\eta,N)}{N} = 0$, which guarantees that the optimal network size is always finite. Below we will show that $V^r(\eta, N)$ does not directly affect transaction prices, but does impact clients' choice of network size.

The buyer purchases the bond from her network at the expected price $\mathrm{E}[P^b]$ and transitions into being an owner with valuation $\widehat{V}^o$. In steady state $\widehat{V}^b$ satisfies the Bellman

[8] The costs of additional dealers could alternatively be modeled as per dealer or per dealer per transaction. Per dealer costs consist of costs of forming a credit relationship and any other costs of maintaining the relationship independent of the number of trades. Such per dealer costs will immediately lead to clients with larger trading intensity using more dealers.

equation linking it to $\widehat{V}^o$:

$$\widehat{V}^b = \frac{1}{1+rdt}[\lambda N dt(\widehat{V}^o - \mathrm{E}[P^b] - K) + (1 - \lambda N dt)\widehat{V}^b + rV^r dt] \tag{5}$$

yields

$$\widehat{V}^b = (\widehat{V}^o - \mathrm{E}[P^b] - K)\frac{\lambda N}{r+\lambda N} + V^r \frac{r}{r+\lambda N}. \tag{6}$$

While clients are owners, they receive a coupon flow C and have valuation $\widehat{V}^o$. Non-owners do not receive the coupon flow. With intensity κ an owner receives a liquidity shock forcing her to become a seller with valuation $\widehat{V}^s$. In steady state valuations in these two states are linked according to the Bellman equation

$$\widehat{V}^o = \frac{1}{1+rdt}[dtC + \kappa dt V^s + (1-\kappa dt)\widehat{V}^o + rV^r dt], \tag{7}$$

yielding the expression for $\widehat{V}^o$:

$$\widehat{V}^o = \frac{C}{r+\kappa} + \underbrace{\widehat{V}^s \frac{\kappa}{r+\kappa}}_{\text{Value From Future Sale}} + V^r \frac{r}{r+\kappa}, \tag{8}$$

where the second term captures the value from future sales. The liquidity shock received by the owner reduces the value of the coupon to $C(1-L)$ until she sells the bond. After receiving the liquidity shock she contacts her dealer network expecting to sell the bond for $\mathrm{E}[P^s]$. Upon selling she becomes a non-owner, completing the valuation cycle. Valuations $\widehat{V}^s$ and $\widehat{V}^{no}$ are related by

$$\widehat{V}^s = \frac{1}{1+rdt}[dtC(1-L) + \lambda N dt(\mathrm{E}[P^s] + \widehat{V}^{no} - K) + (1-\lambda N dt)\widehat{V}^s + rV^r dt], \tag{9}$$

which can be solved for $\widehat{V}^s$ as

$$\widehat{V}^s = \frac{C(1-L)}{r+\lambda N} + (\mathrm{E}[P^s] + \widehat{V}^{no} - K)\frac{\lambda N}{r+\lambda N} + V^r \frac{r}{r+\lambda N}. \tag{10}$$

This sequence of events continues in perpetuity and, therefore, we focus on the steady state of the model.

The above valuation equations depend upon the expected transaction prices. The realized transaction prices are determined by bilateral Nash bargaining. The clients' reservation values are determined by the differences in values between being an owner and non-owner and a buyer and seller. Similarly, the dealers' reservation values arise from their transaction cycle. Each dealer acts competitively, i.e., without taking into account the effect of her actions on the actions of other dealers. In addition, we assume that each dealer internalizes only trade-specific value of her relation with each client.[9] When a client contacts her dealer network, each dealer simultaneously starts looking for the bond at rate λ and expects to pay the inter-dealer ask price M^{ask} for the bond. The value to the dealer searching for the bond satisfies

$$U^b = \frac{1}{1+rdt}[\lambda dt(P^b - M^{ask}) + \lambda N dt U^o + (1 - \lambda N dt)U^b], \tag{11}$$

thus yielding

$$U^b = \underbrace{(P^b - M^{ask})\frac{\lambda}{r+\lambda N}}_{\text{Transaction Profit/Loss}} + \underbrace{U^o\frac{\lambda N}{r+\lambda N}}_{\text{Value of Future Business}}. \tag{12}$$

The last term in the expression for U^b captures the expected value of the future business with the same client which happens with frequency λN. This client, who is now the owner of the bond, becomes a seller with intensity κ and contacts dealers in her network to sell the bond. This generates a value $U^o = U^s\frac{\kappa}{r+\kappa}$ per dealer, where U^s represents the valuation of the dealer searching to sell the bond. The dealer expects to resell the bond at rate λ for the inter-dealer bid price M^{bid} and earn a markup of $M^{bid} - P^s$. She also anticipates that with intensity λN the same client will approach her in the future to buy back the bond. Future business from the same client generates $U^{no} = U^b\frac{\eta}{r+\eta}$ in value to the dealer, thus leading to the following Bellman equation for U^s:

$$U^s = \frac{1}{1+rdt}[\lambda dt(M^{bid} - P^s) + \lambda N dt U^{no} + (1 - \lambda N dt)U^s], \tag{13}$$

which can be solved to obtain

$$U^s = \underbrace{(M^{bid} - P^s)\frac{\lambda}{r+\lambda N}}_{\text{Transaction Profit/Loss}} + \underbrace{U^{no}\frac{\lambda N}{r+\lambda N}}_{\text{Value of Future Business}}. \tag{14}$$

Valuations U^{no} and U^o lead to price improvement for repeat business.

As in most OTC models, prices are set by Nash bargaining resulting in prices that are the bargaining-power (w) weighted average of the reservation values of client and dealer:

$$P^b = (\widehat{V}^o - \widehat{V}^b)w + (M^{ask} - U^o)(1-w), \tag{15}$$

$$P^s = (\widehat{V}^s - \widehat{V}^{no})w + (M^{bid} + U^{no})(1-w). \tag{16}$$

[9] We do not require the dealer to internalize the client's non-trade value V^r of the relationship and omit it. This could happen, for instance, if the dealer is compensated to maximize trading profits and is not incentivized to maximize enterprise value. We also do not need it to fit the data.

The above equations assume that the dealer loses all future business from the client if the bilateral negotiations fail. Upon dropping a dealer the client maintains her optimal network size by forming a new link with another randomly picked identical dealer. Thus, by agreeing rather than not, the dealer receives U^o. As a consequence, dealers face intertemporal competition for future clients. This is a novel assumption missing from the existing models of OTC markets.

Each client's valuation, $\widehat{V}^k$, $k \in \{b, o, s, no\}$, can be written as a sum of its trade-specific, V^k, and the relation-specific, V^r, values

$$\widehat{V}^k = V^k + V^r. \tag{17}$$

Substituting (17) into relations (4), (6), (20), and (10) yields the following relations for trade-specific client valuations:

$$\begin{aligned} V^{no} &= V^b \frac{\eta}{r+\eta}, \\ V^b &= (V^o - \mathrm{E}[P^b] - K)\frac{\lambda N}{r+\lambda N}, \\ V^o &= \frac{C}{r+\kappa} + V^s \frac{\kappa}{r+\kappa}, \\ V^s &= \frac{C(1-L)}{r+\lambda N} + (\mathrm{E}[P^s] + V^{no} - K)\frac{\lambda N}{r+\lambda N}. \end{aligned} \tag{18}$$

Correspondingly, transaction prices depend only on clients' trade-specific valuations:

$$P^b = (V^o - V^b)w + (M^{ask} - U^o)(1-w), \tag{19}$$

$$P^s = (V^s - V^{no})w + (M^{bid} + U^{no})(1-w). \tag{20}$$

The valuations and prices provide ten equations and ten unknowns. Proposition 1 in the Appendix provides the closed-form solutions (35) and (36) for transaction buy prices P^b and, respectively, sell prices P^s.

Bargaining power could differ for buys and sells. For ease of exposition, we equate them here. In the subsequent structural estimation we allow for different bargaining powers when the insurer is looking to buy a bond, w^b, and when the insurer is selling a bond, w^s.

4.2 Discussion

Expressions (35) and (36) are nonlinear functions of the model primitives and N, which renders the analysis difficult. However, we can verify that prices are well-behaved functions of the network size, N, in the large network limit, $N \to \infty$. $N \to \infty$ implies $\frac{\lambda N}{r+\lambda N} \to 1$ and clients' search friction in terms of time is zero. Dealers' valuations, in this case denoted with subscript $N \to \infty$, satisfy the system of equations $U^b_{N\to\infty} = U^s_{N\to\infty}\frac{\kappa}{r+\kappa}$ and $U^s_{N\to\infty} = U^b_{N\to\infty}\frac{\eta}{r+\eta}$. These only have a trivial solution $U^s_{N\to\infty} = U^b_{N\to\infty} = 0$, implying that dealers compete away all rents from future relations with clients. Clients have valuations $V^o_{N\to\infty} - V^b_{N\to\infty} = P^b_{N\to\infty} + K$ and $V^s_{N\to\infty} - V^{no}_{N\to\infty} = P^s_{N\to\infty} - K$ yielding the following expressions for transaction prices:

$$P^b_{N\to\infty} = M^{ask} + \frac{w}{1-w}K, \tag{21}$$

$$P^s_{N\to\infty} = M^{bid} - \frac{w}{1-w}K. \tag{22}$$

Dealers receive no relationship-based rents, but they charge clients a spread $\frac{w}{1-w}2K$ per roundtrip transaction over the inter-dealer spread. The coefficient $\frac{w}{1-w}$ indicates that the non-zero bargaining power enables dealers to extract some value from a client. This is because every dealer charges the same price thus making clients' threat of ending the relationship and forming a new link not credible (Diamond's (1971) paradox). Equations (21) and (22) illustrate the importance of making $K > 0$ in the model. If K is equal to zero a client can choose an infinitely large network. This increases dealer competition such that the client trades at the inter-dealer prices, effectively becoming a dealer herself. Therefore, the finite network is never optimal if K is equal to zero.

Overall, dealers' surplus comes from both the immediate value of trade (the spread) and the value of future transactions. As the probability of transacting with the same client in the future declines with the size of the client's network, dealers may charge higher spreads, $P^b - M^{ask}$ and $M^{bid} - P^s$, when the network size becomes larger. This result is similar to Vayanos and Wang (2007) where an asset with more buyers and sellers has lower search times and worse prices relative to its identical-payoff counterpart with fewer buyers and sellers.

We verify that the model yields a finite network size by considering a limiting case of large search intensity, $\lambda \to \infty$. In this case a single dealer can instantaneously find the bond and the optimal size of the network is one. Taking limits in equations (12) and (14) yields

$$U^{no}_{\lambda\to\infty} = \frac{1}{N}\frac{\frac{\eta}{r+\eta}}{1-\frac{\kappa}{r+\kappa}\frac{\eta}{r+\eta}}\left[P^b_{\lambda\to\infty} - M^{ask} + \frac{\eta}{r+\eta}(M^{bid} - P^s_{\lambda\to\infty})\right], \tag{23}$$

$$U^{o}_{\lambda\to\infty} = \frac{1}{N}\frac{\frac{\kappa}{r+\kappa}}{1-\frac{\kappa}{r+\kappa}\frac{\eta}{r+\eta}}\left[M^{bid} - P^s_{\lambda\to\infty} + \frac{\kappa}{r+\kappa}(P^b_{\lambda\to\infty} - M^{ask})\right].$$

Solving for transaction prices from (15-16) and using that $(V^o - V^b)_{\lambda\to\infty} = P^b_{\lambda\to\infty} + K$,

$(V^s - V^{no})_{\lambda\to\infty} = P^s_{\lambda\to\infty} - K$ we obtain

$$P^b_{\lambda\to\infty} = M^{ask} + \frac{w}{1-w}K - U^o_{\lambda\to\infty}, \qquad (24)$$
$$P^s_{\lambda\to\infty} = M^{bid} - \frac{w}{1-w}K + U^{no}_{\lambda\to\infty}.$$

Expressions (24) show that $U^o_{\lambda\to\infty}$ and $U^{no}_{\lambda\to\infty}$ represent the repeat relation buy discount and sell premium, respectively. Equations (23) show that both $U^o_{\lambda\to\infty}$ and $U^{no}_{\lambda\to\infty}$ are strictly decreasing with the size of the network N. Thus, the client optimally chooses a single dealer.

In order to find the optimal network size N^* in the general case, we maximize the total valuation of the "first-time" owner, i.e., the client buying the asset, $\widehat{V}^b$. This is because the client has to take possession of the asset in the first place. Maximizing $\widehat{V}^b$ accounts for all trade and relation benefits the client receives from a larger network. Proposition 2 given in the Appendix demonstrates that, for a given client type described by the model primitives $\{L, K, w, \kappa, \lambda, \eta\}$, there may exist an optimal network size $N^* = N(L, K, w, \kappa, \lambda, \eta)$.

Next we provide intuition for why the model yields finite size networks greater than one. Without loss of generality we assume that V^r is a monotonically increasing function of the network size N.[10] Because equation (40) is nonlinear in N the existence of the solution is not guaranteed for an arbitrary parameter set $\{L, K, w, \kappa, \lambda, \eta\}$. Given the complexity of the problem it is convenient to consider the optimal network resulting solely from the repeated trading, i.e., N^{**} maximizing V^b from $\frac{dV^b}{dN} = 0$. At this point we do not take a stand on whether V^r is increasing or decreasing function of N. $\frac{dV^r}{dN} > (<)0$ implies that $N^{**} < (>)N^*$ thus making N^{**} the lower(upper) bound of N^*. As a result a sufficient condition for existence of a finite N^{**} also applies to N^*. In addition, our analysis helps to separate the trade-specific and the relation-specific network formation mechanisms.

The derivative $\frac{dV^b}{dN} = \frac{\lambda N}{r+\lambda N}\left(\frac{rV^b}{\lambda N^2} + \frac{dV^o}{dN} - \frac{dP^b}{dN}\right)$ must be positive on $[1, N^{**})$ and equal to zero at $N = N^{**}$. $\frac{dV^b}{dN}$ has three terms. The first one is due to the direct effect of the larger network on the speed of execution as faster execution improves buyer's valuation. The second term reflects the marginal effect of the network size on the owner's value, while the third term represents the effect of the network size on the transaction buy price. We will focus on the case of a buy transaction price, P^b, improving with the network size, $\frac{dP^b}{dN} < 0$. As both the speed of execution and P^b improve with N, the value of being the owner and, consequently, the seller of the bond must decline with N, $\frac{dV^o}{dN} = \frac{\kappa}{r+\kappa}\frac{dV^s}{dN} < 0$, but not faster

[10]It is straightforward to derive the corresponding results when V^r monotonically decreases with N.

than the combined value improvement from increasing N

$$\left|\frac{dV^o}{dN}\right| \leq \frac{rV^b}{\lambda N^2} - \frac{dP^b}{dN},\ N \in [1, N^{**}]. \tag{25}$$

Using relation (10) the sufficient condition for $\frac{dV^s}{dN} < 0$ can be written as

$$V^s + \frac{\lambda N^2}{r}\left(\frac{dP^s}{dN} + \frac{\eta}{r+\eta}\frac{dV^b}{dN}\right) \leq \frac{C(1-L)}{r}. \tag{26}$$

Inequality (26) implies that in order for the seller's value to be decreasing with N, V^s plus the present value of all future marginal increases in V^b with N is no greater than the value of holding the discounted asset net of the present value of all marginal sale price declines from increasing the size of the network. Finally, by differentiating equation (15) with respect to N we obtain the necessary condition for P^b to be improving with N

$$\left|\frac{dU^o}{dN}\right| < \frac{w}{1-w}\left(\frac{dV^b}{dN} - \frac{dV^o}{dN}\right), \tag{27}$$

where we have used that $\frac{dU^o}{dN} < 0$. We conjecture that there exists an interval of network sizes belonging to $[1, N^*]$, on which $\frac{dV^o}{dN} < 0$ and the inequality (27) is satisfied for a range of values of model's primitives. While this conjecture cannot be proven analytically, it is clear that when dealers have high bargaining power, w is close but not equal to 1, the right hand side of (27) can be made larger than its left hand side. Therefore, the optimal network size $N^{**} > 1$, and consequently $N^* \geq N^{**}$, exists and P^b improves with $N \in [1, N^{**}]$ if inequalities (25-27) are simultaneously satisfied.

It then follows from the expression (16) that the sell price must decline with the network size, $\frac{dP^s}{dN} = (\frac{dV^s}{dN} - \frac{\eta}{r+\eta}\frac{dV^b}{dN})w + \frac{dU^{no}}{dN}(1-w) < 0$, as U^{no} is monotonically decreasing with N and V^b is monotonically increasing with N on $[1, N^{**})$. As a result the buyer's value is maximized by trading off more frequent buys at a discounted buy price against more frequent sells at dealer marked up sell price. The effect of N on buy and sell transaction prices is not, however, symmetric due to the time lag between each buy and sell as well as the different effect of relations on P^b and P^s. Therefore, instead of focusing on the individual transaction prices, we investigate the sign of the marginal effect of increasing the network size on the bid-ask spread $SP \equiv P^b - P^s$

$$\frac{dSP}{dN} = -w\left(\frac{r}{r+\kappa}\frac{dV^s}{dN} + \frac{r}{r+\eta}\frac{dV^b}{dN}\right) - (1-w)\left(\frac{dU^o}{dN} + \frac{dU^{no}}{dN}\right). \tag{28}$$

For the network sizes below the trade-specific optimal network size N^{**} we must have that $\frac{dV^b}{dN} > 0$ while the sign of $\frac{dV^s}{dN}$ can be either positive (when $\frac{dP^s}{dN} > 0$) or negative (when $\frac{dP^s}{dN} < 0$). Both values dealers derive from long-term relations with clients monotonically decline with the size of the network, $\frac{dU^{o,no}}{dN} < 0$. Therefore the bid-ask spread improves with the size of the network as long as the following inequality is satisfied

$$\frac{r}{r+\eta}\frac{dV^b}{dN} > -\frac{r}{r+\kappa}\frac{dV^s}{dN} - \frac{1-w}{w}\left(\frac{dU^o}{dN} + \frac{dU^{no}}{dN}\right). \tag{29}$$

At $N = N^*$ we have that $\frac{dV^b}{dN}|_{N=N^*} < 0$ thus implying that $\frac{dSP}{dN}|_{N=N^*} > 0$ as long as $\frac{dV^s}{dN}|_{N=N^*} \leq 0$. This implies that $\frac{dSP}{dN}$ switches its sign from negative to positive, i.e., the inequality (29) is violated, at some $\widetilde{N} < N^*$. Likewise, at $N = N^{**}$ we have that $\frac{dV^b}{dN}|_{N=N^{**}} = 0$ thus leading to $\frac{dSP}{dN}|_{N=N^{**}} > 0$ as long as $\frac{dV^s}{dN}|_{N=N^{**}} \leq 0$, thus implying that the sign of $\frac{dSP}{dN}$ switches from negative to positive at at some $\widetilde{N} < N^{**} < N^*$.

A finite optimal size of the trade-specific network follows from several important trade-offs. When a client adds another dealer to her network the speed of execution, measured by $\frac{\lambda N}{r+\lambda N}$, improves. The additional dealer also improves the bid-ask spread but the improvement declines with the size of the network. As the size of the network becomes large, the value of the relationship, shared by more dealers, declines by so much that dealers have to compensate by charging higher markups and the bid-ask spread starts to widen. Therefore, the spread incorporates the indirect effects additional dealers have on both client and dealer reservation values. At this point the client starts trading off the benefit due to increased execution speed against the cost due to wider bid-ask spread when deciding to add another dealer to the network. The optimal size of the network is set when the benefits from transacting with the larger number of dealers equal the costs. When the relation-specific non-traded value, $V^r(\eta, N)$, is added the client is willing to accept even wider bid-ask spreads in exchange for greater non-trade-relation-specific benefits. However, as marginal non-traded value improvement from a larger network declines with the size of the network, the overall optimal network size N^* is reached when relation specific benefits together with benefits from transacting faster are balanced by the wider bid-ask spread.

4.3 Empirical predictions and identification

While there is no client heterogeneity in the model, each client creates her own access to the interdealer market and the interdealer market can link clients together. Therefore, the model's empirical predictions can be considered via comparative statics with respect to the client's trading intensity, η. Section2.1 has shown that η follows a power law with expoential tail in the cross-section of insurers. Based on these empirical findings, we examine how larger trading intensity impacts clients' choice of network size and the prices they receive. The predictions are complex as network size and transaction prices are interrelated.

Transaction prices: There exist several competing channels through which active insurers receive better prices than inactive insurers, as documented in Table 5. The first channel for price improvement is that active insurers have more repeat trade business with the dealers who internalize the benefits from future business and grant price improvements. A second, complementary explanation is that large insurers have higher bargaining power than small insurers vis-a-vis the dealers. A third explanation is that the trading activity in corporate bonds is correlated with the value of other businesses they conduct with the dealers.

The intuition for the price improvement channel from future business is as follows. It is best to discuss the effect of η on prices while keeping the size of the network fixed. When $\eta = 0$ a client owning the bond will never be an owner again after she sells. Therefore, all of the dealers' surplus comes from the price at which the sale occurs and none from future trades. In this case, a dealer's buy reservation value is lower, resulting in a wider bid-ask spread. When η is large, dealers derive significant value from repeated trades, leading to smaller bid-ask spreads. This implies that, conditional on network size, clients with greater trading intensity (higher η) get better execution than clients with smaller trading intensity (lower η). This is consistent with the empirical findings in Table 2 that insurers' trading activity increases in insurer size and that trading costs decline in insurer size in Table 5.

Network size: Increasing the network size increases buyers' valuation, $\widehat{V}^b$, through improved speed of execution,[11] greater inter-temporal dealer competition,[12] and increased relation-specific non-trade value. The optimal network size, N^*, as a function of η balances these gains against a lower buyer valuation due to the intra-temporal dealer competition leading to the loss of repeat trading business.

The optimal network size is increasing in η for a non-empty parameter range over which the gains from a larger network increase faster in η than the losses ($\frac{\partial^2 \widehat{V}^b}{\partial N \partial \eta}|_{N=N^*} < 0$). Intuitively, dealers' profit from repeated trade improves with the increasing trading intensity and dealers offer better execution. Clients with greater trading intensity benefit from a larger network which improves the speed of execution and may generate larger non-trading relation-specific value. Therefore, consistent with Table 3, the model can generate larger insurers with more frequent trading needs having larger dealer networks.

[11] Speed is proportional to $\widehat{V}^b$ which is increasing with η.

[12] The buy transaction price may or may not be improving with N, $\frac{dP^b}{dN} \gtrless 0$. However, the cross derivative $\frac{\partial^2 P^b}{\partial N \partial \eta}$ is positive.

Network sizes and prices in the cross-section: Figure 5 illustrates for different values of η—when network sizes and prices are jointly determined—how sensitive to the model parameters are the model-implied network sizes (Panel A) and trading costs (Panel B). We implement the sensitivity analysis for each parameter in (32) using the following procedure. The model is fitted to the data by minimizing the distance between data and model prediction for each η and the parameter set $\theta = (\theta_L, \theta_K, \theta^j_{w^i}, \theta_\kappa, \theta_\lambda, \theta_V)$ is estimated that delivers the best fit.[13] We then take a parameter from θ, reduce it by 10% while keeping the other parameters fixed, resolve the model, and then plot the solution (solid line) against the best fit (dahed line).

Figure 5 clearly demonstrates that there exists a lot of variability in how each structural parameter affects the size distribution and transaction costs—thus allowing the identification. Some parameters, such as K, have strong impact on both network size and trading costs. By contrast, L has a very similar impact on trading costs as do the non-trade value parameters θ_V, but they have different impact on the insurers' networks. Similarly, κ and λ impact networks in the same way but transaction costs in different ways.

Ultimately, the model must be structurally estimated to test its ability to fit the cross-sectional distribution of network sizes from Figure 3 as well as the cross-sectional distribution of execution costs as a function of the network size from Table 5.

5 Model Estimation and Policy Analysis

The empirical findings in equation (1) and Table 5 suggest that in the cross-section trading networks follow a power law with exponential tail and that trading costs are non-monotone in network size. While the model can produce such non-monotonicity, it is unclear if it can quantitatively match the empirical relationship between network size and trading costs.

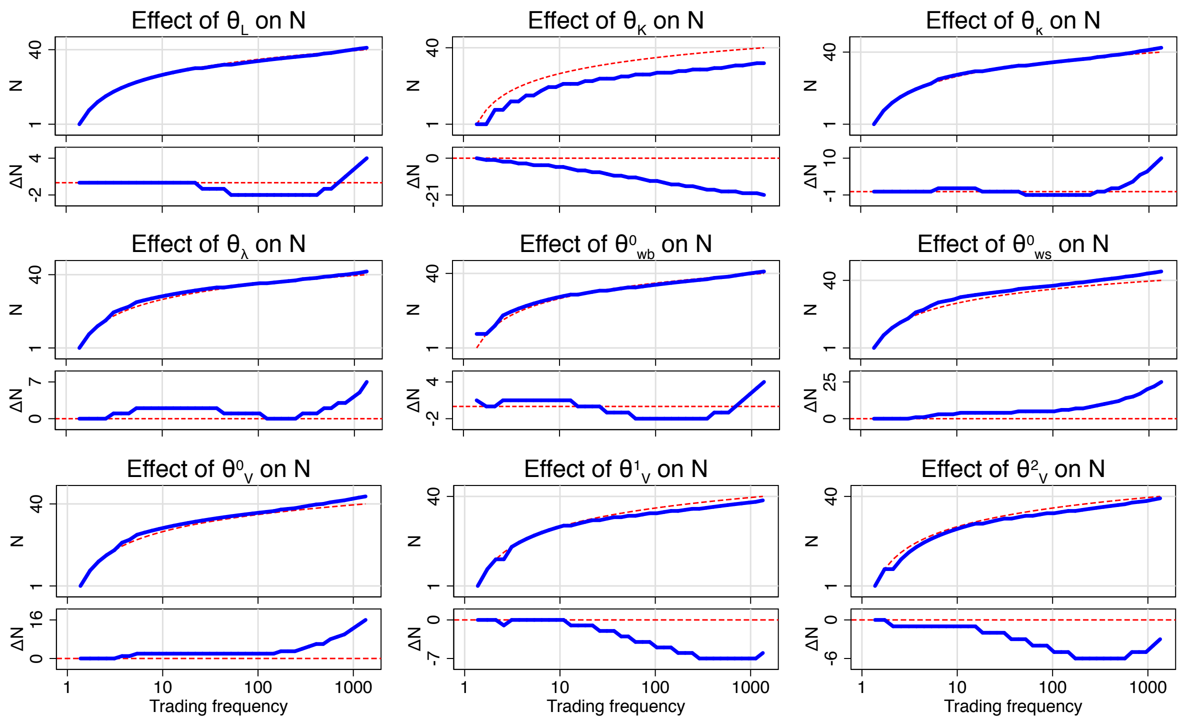
Settings such as labor markets and marriages markets involve one-to-one matching. These literatures structurally estimate models to examine their quantitative fit to the data (for example, Poste-Vinay and Robin (2002) and Eckstein and Van den Berg (2007) for labor search and Hitsch et al. (2010) and Choo (2015) for marriage search). Gavazza (2016) estimates a one-to-one search-and-bargaining model of a decentralized market using aircraft transaction data. We follow this approach for our one-to-many network formation model.[14]

Figures 2 and 3 present insurers' heterogeneous trading intensities and network sizes.

[13]Details of the estimation are reported in the next subsection.

[14]The identification of the model shares several similarities with Gavazza (2016). As in Gavazza (2016) the identification of unobserved parameters relies on key moments in the data. Unlike the aggregate moments in Gavazza (2016), we utilize heterogeneity in insurers' trading intensities and networks to facilitate our estimation.

Panel A: Network sizes N



Panel B: Trading cost $c(N)$

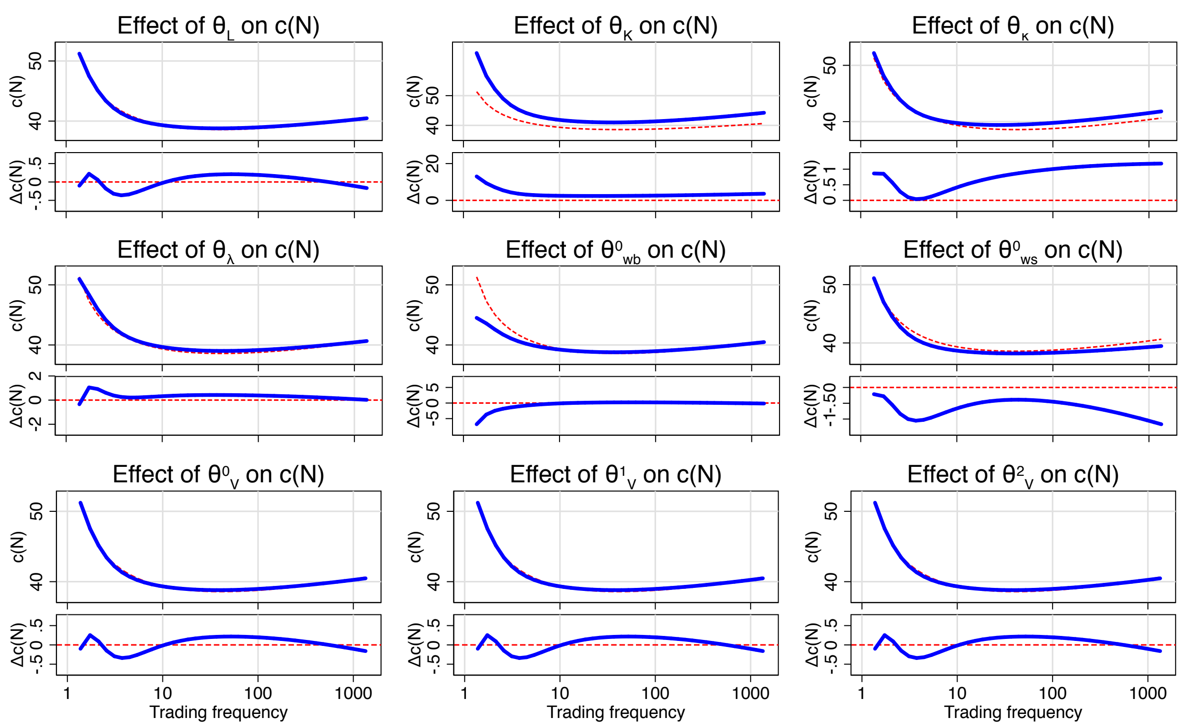


Figure 5: Sensitivity to model parameters and identification

The figure illustrates the sensitivity of the model-implied network size distribution (Panel A) and trading costs (Panel B) to model parameters. The top panel in each plot shows the network sizes N/trading costs $c(N)$ at the estimated parameters (dashed line) compared parameters that are shifted down by 10% (solid line). The bottom panel in each plot shows the difference in network sizes ΔN/trading costs $\Delta c(N)$ between estimated and shifted parameters, illustrating the sensitivity of network sizes/trading costs to each parameter (across plots) for different values of trading frequency η on the horizontal axis.

To estimate the model's distribution of network sizes we infer the distribution of trading intensities, η_n, $n = 1, ..., \mathcal{N}$, across insurance companies n. Section 2.1 characterizes the compound distribution of trading activity. If trading shocks occur at Poisson times, the intensity η_n of the shocks can be estimated by the expected number of buy trades per year. This yields the maximum likelihood estimator $\widehat{\eta_n} = \frac{1}{T}\sum_{t=1}^{T} X_{nt}$, where X_{nt} is the number of bond purchases by insurer n in year t. To utilize the multiple years of trade data, we perform the estimation separately for each insurer. This yields a cross-sectional distribution of trading intensities, which we index by $p(\eta)$. Section 2.1 and Figure 2 show the distribution of the insurer trades per year follows approximately a power law. For insurer buys, this distribution is best described by $p(\eta) = 0.34 \times \eta^{-1.31}$.

The other model parameters $\Theta = (L, K, \kappa, \lambda, w^s, w^b)$ are not directly observable in the data, so they are estimated structurally. Section 2.2 and Figure 3 show the degree distribution of client-dealer relations follows a mixed power-exponential law. The distance between the empirical distribution and the model-generated network sizes N^* depends on the parameters Θ. We fit this by minimizing the probability-weighted distance between the data and model:

$$\sum_{N=1}^{max(N)} p(N)[N - N^*(\Theta, \eta(p(N)))]^2, \tag{30}$$

where $p(N)$ is the empirical network-size distribution and $N^*(\Theta, \eta)$ is the model-implied network size for the set of parameters Θ given η. Inverting the power law distribution for trading intensities from Panel A of Figure 2 yields $\eta(p) = (\frac{p}{.34})^{-\frac{1}{1.31}}$. It then follows from equation (1) that for insurer buys $p(N) = .28e^{-.22N}N^{-0.12}$. Therefore, by substituting $p(N)$ into the expression for $\eta(p)$ we obtain the mapping of trading intensities into the network sizes, $\eta(p(N)) = (\frac{.28}{.34})^{-\frac{1}{1.31}} e^{\frac{.22}{1.31}N} N^{\frac{0.12}{1.31}}$. We use this relation in (30).

The model also predicts how percentage trading costs, $(P^b - P^s)/.5(P^b + P^s)$, depend on the parameters Θ and η. Empirically, the estimated relation between trading costs and network size is in column (4) of Table 5 as $c(N) = 51 + 0.32N - 6.29\, lnN$. The probability-weighted distance between the empirical and the model-generated trading costs c^* depends on the parameters Θ and is:

$$\sum_{N=1}^{max(N)} p(N)[c(N) - c^*(\Theta, \eta(p(N)))]^2. \tag{31}$$

A minimum-distance estimator chooses the model parameters to minimize the sum of the distances (30) and (31). To fit the data the model's optimal N^* is constrained to be an integer. This makes the the objective functions (30) and (31) non-smooth. To accommodate this we optimize using simulated annealing with fast decay and fast reannealing.

To ensure that the estimated parameters remain in their natural domains, i.e., positive or between zero and one, in the estimation we transform model parameters as follows:

$$L = \Phi(\theta_L) \in (0,1), \quad K = e^{\theta_K} \geq 0, \quad \kappa = e^{\theta_\kappa} \geq 0, \quad \lambda = e^{\theta_\lambda} \geq 0, \tag{32}$$
$$w^s = \Phi(\theta^0_{w^s} + \theta^1_{w^s} ln\, \eta) \in (0,1), \quad w^b = \Phi(\theta^0_{w^b} + \theta^1_{w^b} ln\, \eta) \in (0,1),$$

where $\Phi \in (0,1)$ is the normal cdf. Our results are robust to alternative transformations in place of Φ and e.

As discussed above, insurers' trading shocks are heterogeneous and directly observable in the data. In the model κ measures clients' selling intensity after having bought. Because the empirical number of buys and sells are equal, and buys and sells follow a similar power law, there is no evidence of heterogeneity in κ. Therefore, κ does not vary across insurers. Because we fit the model to the two observable outcomes of network size and trading costs we limit the number of model parameters that vary across insurers by assuming L and K are constant. The competitive interdealer spread is set to match the average execution costs of insurers by choosing $C = 1$, $r = .1$, $M^{ask} = (1 + 0.0018)\frac{C}{2r}$, and $M^{bid} = (1 - 0.0018)\frac{C}{2r}$. Because insurers can potentially access all dealers λ is constant across insurers.

We set the functional form of the non-trade value to be Cobb-Douglas

$$V^r = e^{\theta^0_V} \eta^{\theta^1_V} N^{\theta^2_V}, \; \theta^1_V + \theta^2_V < 1, \tag{33}$$

with a constant "technology" $e^{\theta^0_V}$ and elasticities to trading intensity and network size given by θ^1_V and θ^2_V respectively. The Cobb-Douglas functional form captures the complementarity between trading intensity, measured by η, and network size, while precluding infinitely large non-trade value.

Finally, the evidence in Table 5 shows trading costs depend on insurer size even after controlling for network size. The most straightforward way to accommodate this is to allow dealer bargaining power to vary across insurers, e.g., large active insurers have higher effective bargaining power than small inactive insurers. We report the model fit under both constant (specification 1) and variable (specification 2) bargaining power by holding w^b and w^s constant and allowing w^b and w^s to be functions of η, respectively.

5.1 Estimated model parameters, discussion, and model fit

The model can be solved assuming the client maximizes her value as a buyer or as a seller. To simplify exposition we use her value as a buyer. Panel A of Table 6 reports the estimates for the transformed model parameters from (32). Panel B provides the original model parameters $\Theta = (L, K, \kappa, \lambda,\ w^s, w^b)$. All parameters are significantly different from zero and appear well identified as the standard deviations of the residuals are small. Specification 2 allows bargaining power to vary across insurers and generates a substantially lower minimum distance between the model and empirical distributions than does the uniform bargaining power in specification 1.

The estimated liquidity shock parameter L in both specifications is 100% of the flow income from the bond, suggesting a high willingness to pay for immediacy. The cost K of obtaining quotes for each transaction is small, 0.01 (specification 1) and 1.56 (specification 2) basis points per trade, respectively. The estimated selling shock intensity κ is between 13.51 (specification 1) and 19.95 (specification 2), a holding period from two and a half to four weeks. Longer holding periods, smaller κ, reduce the value of repeat relations and increase network size. Without heterogeneity in κ the model needs relatively short holding periods to reproduce the large fraction of insurers using a single dealer.

The dealers' search efficiency λ is estimated to be 385 in specification 1, corresponding to dealers taking two-thirds of a day $(250/\lambda)$ to locate a bond. The dealers' search efficiency λ is estimated to be 166 (1.5 days) in specification 2, which allows bargaining power to depend on η. Given overall corporate bond trading frequencies, these estimates seem reasonable.

Dealers' bargaining power on the buy and the sell sides in specification 1 is large, 98% and 96%, suggesting that dealers capture most of the trade surplus in this specification. In specification 2, when bargaining power depends on insurers' type, dealers' average bargaining power remains large and fairly symmetric across buys, 85%, and sells, 89%. Dealers bargaining power on sales is relatively insensitive to insurer trading frequency. In contrast, $\theta^1_{w^b} = -1.09$ indicates that dealers' bargaining power when buying declines significantly with insurers trading intensity η.[15] Dealers' bargaining power with insurers with trading intensity $\eta = 1$ is close to one. As insurer trading intensity increase to five, dealers' bargaining power falls to about one half. Dealers have almost zero bargaining power when the largest insurers are buying. This could arise from insurers being buy-and-hold investors whereby a dealer can more easily locate a seller than a buyer. The heterogeneity in bargaining across insurers and across buy and sell transactions suggests richer modeling of the price setting process may enable deeper understanding of OTC trading.

The estimated non-trade value, V^r, is quite small in specification 1. The estimated non-trade value, V^r, is fifteen times larger in specification 2 and it is monotonically increasing with the size of the network and trading intensity. The decreasing returns to scale parameter

[15] The asymmetry in buy-sell bargaining power's sensitivity to η arises from our choice to maximum the clients' valuation as a buyer.

Table 6: Estimated model parameters

The table reports the estimated model parameters $\theta = (\theta_L, \theta_K, \theta^j_{w^i}, \theta_\kappa, \theta_\lambda, \theta_V)$ in Panel A. Estimates are from the minimum-distance estimation. Standard errors are reported in parenthesis and scaled by 100. Panel B reports the implied values for the model parameters $K = e^{10*\theta_K}$, $L = \Phi(\theta_L)$, $\kappa = e^{10*\theta_\kappa}$, $\lambda = e^{10*\theta_\lambda}$, $w^i = \Phi(\theta^0_{w^i} + \theta^1_{w^i} ln\, \eta)$ for $i = s, b$, where $\Phi \in (0, 1)$ is the normal cdf, and non-trade value $V^r = e^{\theta^0_V} \eta^{\theta^1_V} N^{\theta^2_V}$.

	(1) w^b, w^s constant	(2) $w^b(\eta), w^s(\eta)$
Panel A: Parameter estimates		
θ_K	-1.23 (0.50)	-0.65 (0.04)
θ_L	5.02 (43.48)	4.18 (0.98)
θ_κ	0.26 (0.01)	0.30 (0.15)
θ_λ	0.60 (0.38)	0.51 (0.10)
$\theta^0_{w^b}$	2.99 (0.68)	2.55 (0.18)
$\theta^1_{w^b}$		-1.09 (0.24)
$\theta^0_{w^s}$	1.93 (0.94)	1.32 (0.23)
$\theta^1_{w^s}$		0.02 (0.11)
θ^0_V	-3.61 (0.80)	-1.75 (0.19)
θ^1_V	0.09 (0.76)	0.30 (0.20)
θ^2_V	-0.51 (0.93)	0.15 (0.19)
Minimum distance	1,729.80	27.03
S.D. residuals	4.68	0.58
S.D. residuals network	1.35	0.40
S.D. residuals prices	4.48	0.42
Panel B: Implied model parameters		
K*1,000	0.01	1.56
L	1.00	1.00
κ	13.51	19.95
λ	384.99	165.88
w^b (S.D.)	0.98 (0.00)	0.85 (0.05)
w^s (S.D.)	0.96 (0.00)	0.89 (0.01)
Non-trade value V^r (S.D.)	0.02 (0.00)	0.30 (0.02)

restriction is satisfied. The trade intensity elasticity is twice the network size elasticity.

Figures 6 and 7 visually examine the quality of the model's fit. They each provide four plots illustrating the fits for their respective specification. The top left graph plots the number of dealers as a function of their trading intensity both in the data (circles) and in the model (dashed line). Similar to Figure 3, the top right graph plots the degree distribution for insurer-dealer relations in the data (circles) compared to the model-implied distribution under the estimated parameters (dashed line). In both cases the distance between the two lines captures the model fit. To further visualize the goodness of model's fit, the bottom left graph plots the degree distribution for insurer-dealer relations in the data against the model-implied distribution under the estimated parameters. Each circle is labeled with the corresponding network size. The deviation from the 45-degree line measures model fit.

The specification with uniform dealer bargaining power fits the network distribution up to 16 dealers. In the bottom left graph in Figure 6 the blue circles are on the 45-degree line for all values of N less than 16. However, the goodness of its fit deteriorates for larger

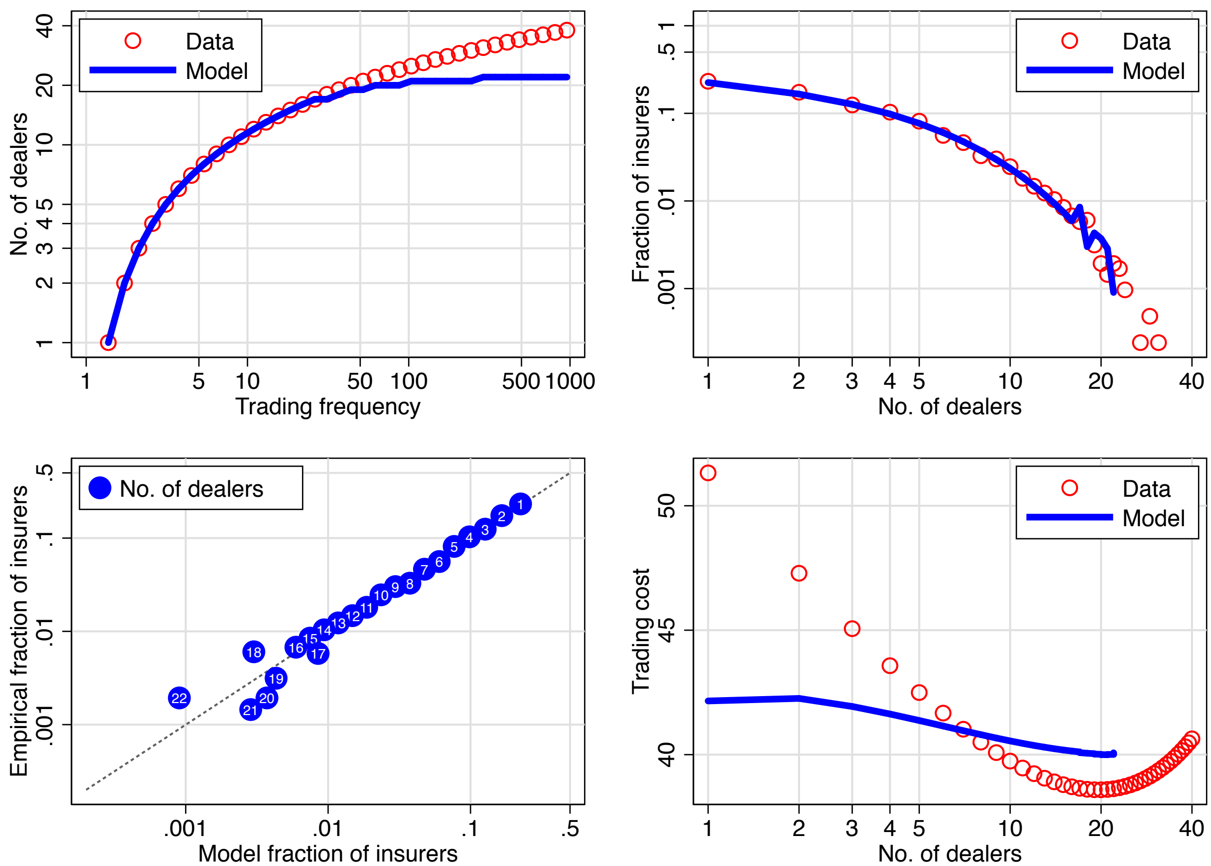


Figure 6: Model fit with constant bargaining power w^b, w^s

The figure shows the model fit for the base specification of constant dealers' bargaining power w. The top left plot shows the network size of insurer-dealer relations for insurer buys as a function of the insurers' trading frequency in the data (circles) compared to the model-implied distribution under the estimated parameters (solid line). The top right plot shows the degree distribution for network size in the data (circles) compared to the model-implied distribution under the estimated parameters (solid line). The bottom left plot compares the empirical probability of a given network size to the model-implied probability. The bottom right plot shows the bid-ask spread in the data as a function of the insurers' network size in the data (circles) compared to the model-implied bid-ask spread under the estimated parameters (solid line).

networks and the model does not generate networks larger than 22 dealers.

The bottom right corner graph in Figure 6 plots the empirical relation between trading costs and network size from column (4) of Table 5 (circles), $c(N^*) = 51 + 0.32N^* - 6.29\, lnN^*$, against its model-implied counterpart (dashed line).[16] The distance between the two lines is a measure for model fit, showing the parameter estimates from specification 1 do not well

[16]Because values reported in Table 5 are estimated on individual transactions and contains many control variables, it is simpler to plot the functional relationship rather than some other transformation of the underlying data.

describe the relationship between network sizes and trading costs in the data. The model's relation is too weak as the line is too flat. In addition, the non-monotonicity in the relation is barely visible. This suggests that uniform bargaining power limits the variation in the benefits and costs of having a larger network.

In the model network size impacts trading costs, insurers with different η choose different $N^*(\eta)$, and η impacts trading costs independent of the network size. Understanding how these interact in the model provides insight into why the model with uniform bargaining power weakly fits the relationship between network sizes and trading costs. The spread can be written as $SP(\eta(N^*), N^*)$. Its full derivative with respect of N^* is

$$\frac{dSP}{dN^*} = \frac{\partial SP}{\partial N^*} + \frac{\partial SP}{\partial \eta}\frac{d\eta}{dN^*}. \tag{34}$$

The top left graph in Figure 6 is the inverse of $\frac{d\eta}{dN^*}$. The top right graph in Figure 6 weighs this by the observed empirical distribution of trading intensities. Therefore, fitting the model requires fitting $\frac{d\eta}{dN^*}$ and $\frac{dSP}{dN^*}$, and $\frac{dSP}{dN^*}$ depends on $\frac{d\eta}{dN^*}$. The model determines $\frac{\partial SP}{\partial N^*}$ and $\frac{\partial SP}{\partial \eta}$.

The first term in (34) corresponds to the derivative of spread with respect to optimal network size in the model. As discussed earlier, at the optimum this derivative is positive. The second term in (34) corresponds to the effect of insurers with different η choosing different N^* and η impacting trading costs independent of the network size. The value of future business causes the bid-ask spread to improve with η, $\frac{\partial SP}{\partial \eta} < 0$, and the optimal network size is a non-decreasing function of η, $\frac{d\eta}{dN^*} \geq 0$, the second term in (34) is always negative. The opposing signs of the two terms explain how the model can produce the non-monotonic relationship between network size and trading costs in the data. To fit the trading cost-network size relationship in Table 5 the full derivative in (34) must be initially negative and then become positive at 22 dealers. The slope of the line for the model with uniform bargaining power in neither initially negative enough nor does it increase noticeably enough at 22.

The log-log plot in the top left of Figure 6 shows that fitting $\frac{d\eta}{dN^*}$ requires that $N^* = 1$ for a large range of η. If dealers place a high value on repeat business then spreads are very sensitive to increasing the number of dealers. In the model high dealer bargaining power, w, increases the value of repeat business. Therefore, the fit in specification 1 matches the large fraction of insurers with less than five dealers by setting the dealers' bargaining power close to one. Unfortunately, this same mechanism makes it costly to have larger dealer networks, so the model produces too few insurers with dealer networks larger than 16 and no networks larger than 22. Finally, high dealer bargaining power makes the relationship weak between spreads and network size. If dealer bargaining power declines with insurer trading intensity then larger networks may be optimal and the relationship between network size and trading costs may be larger. Specification 2 allows for bargaining power to vary with η.

The model fit improves drastically when dealers' bargaining power depends on η. All four plots in Figure 7 show a close correspondence between the model and data. The graphs show network sizes greater than 22 and the U-shape in the bid-ask spread starting at 20 dealers. In this specification dealers' bargaining power is greater than 0.8 for insurers with trading intensity less than five.[17] Dealer bargaining power for insurer buys is about 0.5 when trading intensity approaches 10. When η exceeds 30 dealer bargaining power for insurer buys is small. This enables the model to produce large network sizes for insurers with large η as the value dealers place on repeat business increases with η. In addition, bargaining power varying with η has a direct effect on trading costs. Larger insurers have larger bargaining power, lowering their trading costs. This strengthens the relationship between trading costs and network sizes.

To summarize, the theoretical model with the dealers' bargaining power declining with a trading intensity η fits the data well. In particular, it is important to have the dealer's bargaining power on the ask side to depend on η. That is, small inactive insurers have very weak outside options both when they buy and sell. By contrast, large active insurers have strong outside options when they buy and weak bargaining power when they sell.

5.2 Impact of policy changes

The parameter estimates can help quantify the value of both the repeat client-dealer trading and non-trade value V^r. We use the model and estimated parameters to construct counterfactuals that capture the impact of repeat trade business on client-dealer networks and prices. We compute counterfactual networks and bid-ask prices under the assumption that the dealers give only partial price improvement for the future repeat trading from the same insurer. In this case, the search and bargaining proceeds in the same way, but the dealers in the network are implicitly re-chosen with probability $\frac{1}{2}$ after every trade. This is counterfactual scenario 1. In this counterfactual scenario, the bid and ask price equations in the model are adjusted by dividing both U^o and U^{no} in (15) and (16) by two. To capture the impact of V^r on networks and transaction prices we compute counterfactual networks and bid-ask prices under the assumption that clients fully benefit from the repeat trading but receive no relation-specific value from dealers, $V^r = 0$. This is counterfactual scenario 2.

The upper left plot in Figure 8 corresponds to the top left plot in Figure 7 and shows

[17]Figure 2 shows that the majority of insurers have trading intensity less than five.

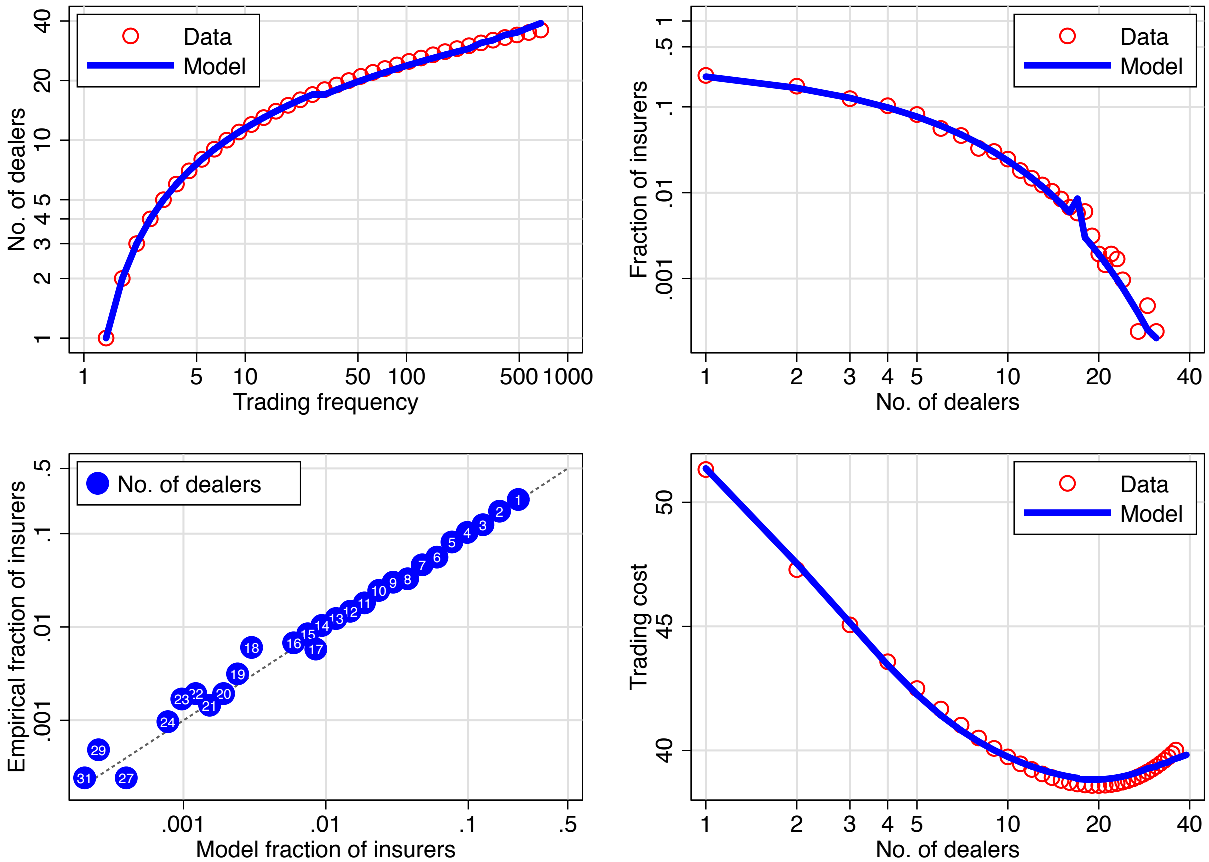


Figure 7: Model fit with client-specific bargaining power $w^b(\eta), w^s(\eta)$

The figure shows the model fit when dealers' bargaining power is a function of trading intensity, $w^b(\eta)$ and $w^s(\eta)$. Please see caption of Figure 6 for further details.

the optimal network sizes, N^*. Because the network sizes change under the counterfactuals, the upper right plot in Figure 8 shows the bid-ask spread as a function of trading intensity, η. All four graphs in Figure 8 use the estimated parameters from specification 2 in Table 6. Each graph shows the results from specification 2 (solid line), counterfactual 1 (crosses), and counterfactual 2 (circles).

Figure 8 highlights the fundamental trade-offs the insurer faces in our model. When dealers' repeat trade benefits are reduced (counterfactual 1) dealers charge wider bid-ask spreads. Dealers in smaller networks lose more repeat trade surplus than dealers in large networks because each dealer's per-trade loss is scaled by the network size. Smaller insurers with lower trading intensities have smaller networks. Consequently, the widening in bid-ask spread is slightly larger for insurers with lower trading intensities than for insurers who trade very frequently. Both large (high η) and small (low η) insurers respond to wider bid-ask spread by reducing the size of their networks. The magnitude of this effect is much smaller

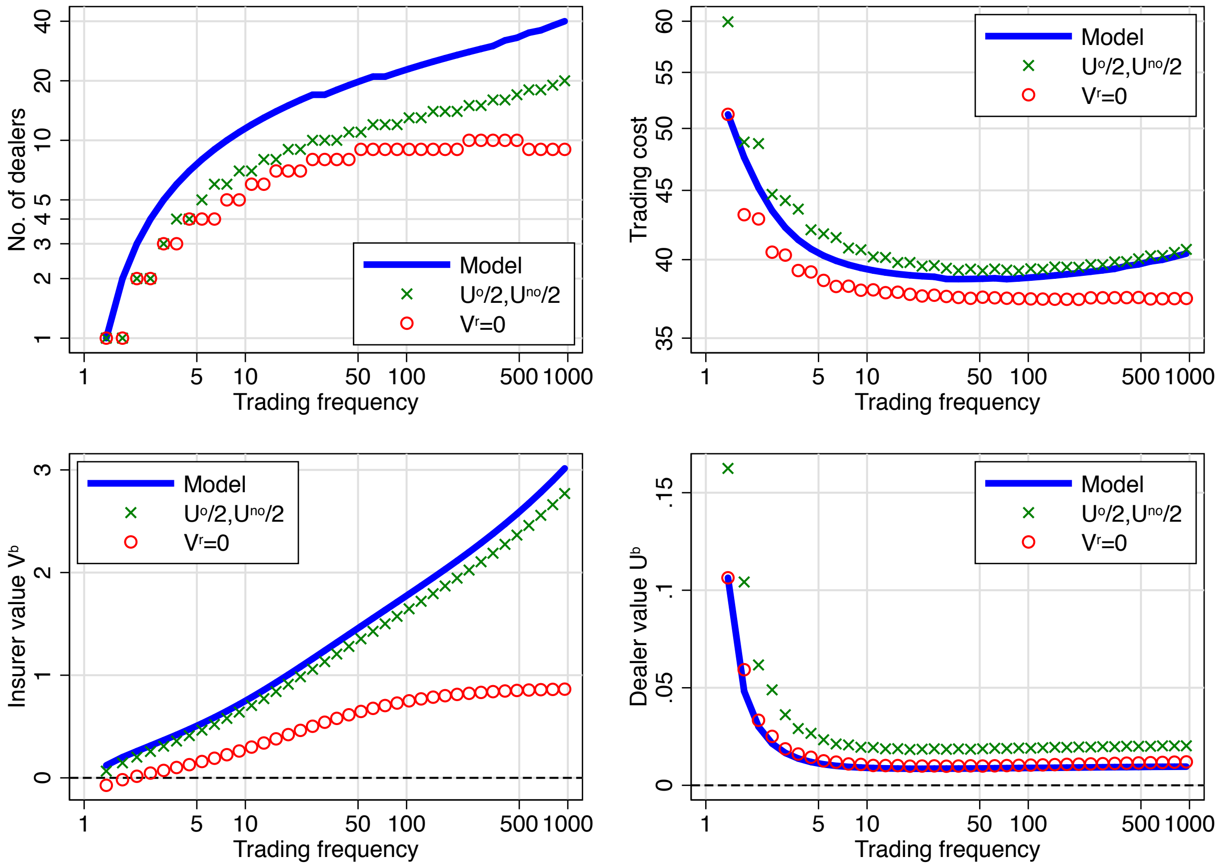


Figure 8: Counterfactual analysis

The left plot shows the insurer-dealer network size as a function of the insurers' trading frequency, η. The right plot shows the bid-ask spread as a function of η. The bottom left plot shows the insurers' value from trading, V^b, as a function of the insurers' trading frequency, η. The bottom right plot shows the dealers' value from trading, U^b, as a function of the insurers' trading frequency, η. Each plot shows the corresponding variable in under the estimated parameters from specification 2 in Table 6 (solid line) compared to the counterfactuals that would arise if: (i) dealers have suffered a permanent loss of 50% of the repeat trade business (crosses); (ii) the non-trade relationship value is zero (circles). We use log-log scale.

for insurers with low η as they already have small networks.

When the non-trade value V^r is set to zero in counterfactual 2 insurers reduce their network sizes. Because V^r is larger for insurers with larger trading intensities, they reduce the size of their networks more than insurers with low η. Dealers respond to the network size reduction by charging better transaction prices thus leading to narrower bid-ask spreads for insurers of all η-types. The magnitude of the bid-ask spread improvement is larger(smaller) for insurers with high(low) η since they reduce the size their network by more(less).

The U-shape in the bid-ask spread is not present without the non-trade relationship value V^r. This highlights the role V^r increasing in N^* plays in the non-monotonicity of relationship between network size and the bid-ask spread. As V^r increases with N^*, insurers have an incentive to choose larger networks than in the absence of the non-trade value because larger transaction costs are offset by the non-trade value. Because V^r increases with η as well, this incentive is especially strong for insurers with very large trading intensities and large networks, thus leading to the U-shape in the bid-ask spread for large networks.

MiFID II attempts to unbundle various aspects of relationship between clients and dealers. Trading and non-trading services are supposed to be priced and sold separately. In addition, there exists a large push by the U.S. and European regulators to improve transparency and liquidity in the corporate bond market by shifting trading to limit order book-type electronic platforms like MarketAxess, TrueMid, and others, easily accessible by both institutional and retail investors. All these initiatives will impact the transaction costs incurred by insurers but very little empirical and/or theoretical evidence exists to quantify these effects. Our counterfactuals provides useful insights about quantitative implications of both moving the bond trading from the OTC markets to the electronic limit order book where the probability of the repeat trading with the same dealer is reduced (counterfactual 1) and unbundling trading and non-trading dealer services (counterfactual 2).

The upper right panel of Figure 8 shows that when the probability of repeat trading with the same dealer is reduced (green crosses), all insurers will incur higher transaction costs. Insurers who trade more frequently and, therefore, have larger networks will see much smaller increase in the transaction costs than insurers who trade less frequently and tend to trade repeatedly with 1 to 5 dealers. For instance, insurers with 100 annual trades have on average 11 dealers in their network and will see less than 1 bps increase in transaction costs per bond per transaction. These insurers tend to be larger with more market power. In contrast, insurers with less than 10 trades annually have networks with up to 5 dealers and will on average pay 3 to 7 bps more per bond per transaction. These insurers also tend to be smaller and more vulnerable to bond market conditions.

When the non-trade value is set to zero (red circles), the transaction costs are improved for all insurers with the largest insurers trading most frequently getting the largest improvement. This is because these insurers getting the largest non-trade benefits from dealers at the expense of the bond transaction prices. Overall, our results indicate that decoupling trade and non-trade insurer-dealer business will decrease transaction costs for all but a few insurers.

The bottom left plot of Figure 8 shows that the proposed regulations would decrease the value from trading, V^b, for all types of insurers. Larger, more active insurers would be affectede differently than smaller, less active insurers. In the model, higher-η insurers incur the largest drop in V^b. These insurers would also reduce the size of their trading networks the most. However, insurers with lower-η already extract a small value from trading. As a consequence, under the new regulations the value they generate from trading may become negative, causing them to exit the market. The proposed regulations can potentially reduce trading activity and decrease insurers' welfare.

As for the dealers, the bottom right plot of Figure 8 shows that the reduction in network sizes by insurers diminishes inter-dealer competition thus raising the value of trading for the dealers still reamining in the network. The profits from trading, U^b, would rise for the remaining dealers across all types of investors. Profits would rise the most for low-η insurers for which U^b is already the highest. The proposed regulations may thus increase the disparities between different dealers and foster a "dog-eat-dog" selection among the dealers.

6 Conclusion

Over-the-counter markets are pervasive across asset classes and often considered poorly functioning due to lack of transparency and fragmented trading imposing search frictions. Regulators have attempted to address these concerns through increased transparency, fostering of anonymous electronic trading platforms, and unbundling of trade and non-trade services. Changes in regulations and market structure will impact heterogeneous investors differently. However, there is limited theory closely linked to empirical work to guide these decisions.

We use comprehensive regulatory corporate bond trading data for all U.S. insurance companies to study how investors choose the size of their trading networks and how this impacts the transaction prices they receive in the current decentralized OTC market setting.

We document that 30% of insurers trade only with a single dealer and up to 100 times each year. There exist few insurers with networks of up to 40 dealers. The cross-sectional distribution in trading activity is a power law, while the network sizes follow a hybrid of a power law with an exponential tail. Trading costs decline with network size up to 20 dealers and increase for larger networks.

A parsimonious model of OTC trading in which insurers build one-to-many dealer relations can match the empirical regularities. Insurers trade off the value of repeat relations with dealers against the benefits of intra-temporal competition among dealers. The value of repeat relations declines in the number of dealers as the increased competition erodes the chance to transact. Dealers compensate for losses from repeat business by charging higher spreads. The value of repeat relations diminishes more slowly with the addition of dealers for clients with larger trading intensity as dealers compete for larger repeat business. Therefore, larger clients use more dealers and get better execution as benefits from having larger repeat business swamp the costs of having larger network. Dealers provide better prices to these larger clients because their repeat business is more valuable. Eventually the costs of having larger network outweigh the benefits and the spread starts to increase with the network size.

Using the structurally estimated model parameters we assess to impact of proposed regulations to unbundle and abolish client-dealer relations. We find that clients will incur higher transaction costs when repeat trading is reduced through, e.g., anonymous electronic trading facilities. Unbundling trade and non-trade services will decrease transaction costs for all insurers except for the least active ones and reduce optimal network sizes—fostering a "dog-eat-dog" selection among the dealers. More importantly, the proposed regulations may reduce trading activity and decrease welfare, especially for clients that trade infrequently.

Appendix

Data Filters

Table 7: Data filters

Filter	Full sample	Corp. bonds
1. All trades from original filings (includes all markets, all trades since 2001)	19.1	4.5
2. Remove all trades that do not involve a dealer (e.g., paydown, redemption, mature, correction)	6.6	3.1
3. Remove duplicates, aggregate all trades of the same insurance company in the same CUSIP on the same day with the same dealer	6.5	3.1
4. Map dealer names to SEC CRD number, drop trades without a name match, drop trades with a dealer that trades less than 10 times in total over the sample period	6.1	2.9
5. Drop if not fixed coupon (based on eMaxx data), drop if outstanding amount information is in neither eMaxx nor FISD	5.3	2.5
6. Drop if trade is on a holiday or weekend	5.2	2.5
7. Drop if counterparty is "various"	4.1	2.1
8. Drop trades less than 90 days to maturity or less than 60 days since issuance (i.e., primary market trade)	2.8	1.5
9. Merge with FISD data, keep only securities that are not exchangeable, preferred, convertible, issued by domestic issuer, taxable muni, missing the offering date, offering amount, or maturity, and offering amount is not less than 100K	1.0	1.0

Parts 3 and 4 of Schedule D filed with the NAIC contains purchases and sales made during the quarter, except for the last quarter. In the last quarter of each year, insurers file an annual report, in which all transactions during the year are reported. Part 3 of Schedule D reports all long-term bonds and stocks acquired during the year, but not disposed of, while Part 4 of Schedule D reports all long-term bonds and stocks disposed of. In addition, all long-term bonds and stocks acquired during the year and fully disposed of during the current year are reported in the special Part 5 of Schedule D. NAIC's counterparty field reports names in text, which can sometimes be mistakenly typed. The bank with the most variation in spelling is DEUTSCHE BANK. We manually clean the field to account for different spellings of broker-dealer names.

We compile the information in Parts 3, 4, and 5 of Schedule D to obtain a comprehensive set of corporate bond transactions by all insurance companies regulated by NAIC.

We apply various FISD-based data filters based on Ellul et al. (2011) to eliminate outliers and establish a corporate bond universe with complete data. The data filters are in Table 7 which summarizes the number of observations that is affected by each step. We exclude a bond if it is exchangeable, preferred, convertible, MTN, foreign currency denominated, puttable or has a sinking fund. We also exclude CDEB (US Corporate Debentures) bonds, CZ (Corporate Zero) bonds, and all government bond (including municipal bonds) based on the reported industry group. Finally, we also drop a bond if any of the following fields is missing: offering date, offering amount, and maturity. We restrict our sample to bonds with the offering amount greater than \$10 million, as issues smaller than this amount are very illiquid and hence are rarely traded. Ellul et al. (2011) have used \$50,000 which we find restrictive for our purpose. We windsorize the cash-to-asset ratio at 80 percent to remove extreme values.

Proofs

PROPOSITION 1: *Bid and ask prices are*

$$P^b = \left[\frac{C(1-L)}{\lambda N}\Psi_1 + \frac{C}{r+\kappa}\Psi_2 + K\Pi_1(\kappa,\eta)\right] w + \left[M^{ask}\Pi_2 + M^{bid}\Pi_3\frac{\kappa}{r+\kappa}\right](1-w), \tag{35}$$

$$P^s = \left[\frac{C(1-L)}{\lambda N}\Psi_3 - \frac{C}{r+\kappa}\Psi_4 - K\Pi_1(\eta,\kappa)\right] w + \left[M^{bid}\Pi_2 + M^{ask}\Pi_3\frac{\eta}{r+\eta}\right](1-w). \tag{36}$$

where the coefficients Ψ_1, Ψ_2, Ψ_3, Ψ_4, $\Pi_1(x,y)$, Π_2, and Π_3 are given below in the proof.

Proof: Define $\Lambda = \lambda N$, $\widetilde{\lambda} = \frac{\lambda}{r+\Lambda}$, $\widetilde{\Lambda} = \frac{\lambda N}{r+\lambda N}$, $\widetilde{\eta} = \frac{\eta}{r+\eta}$, and $\widetilde{\kappa} = \frac{\kappa}{r+\kappa}$. We can rewrite both the client's and the dealer's valuations as

$$\begin{aligned} V^b &= \frac{C}{r+\kappa}\widetilde{\Lambda} - (P^b+k)\widetilde{\Lambda} + V^s\widetilde{\kappa}\widetilde{\Lambda} = \frac{\widetilde{\Lambda}}{1-\widetilde{\kappa}\widetilde{\eta}(\widetilde{\Lambda})^2}\left[\frac{C(1-l)}{r+\Lambda}\widetilde{\kappa} + \frac{C}{r+\kappa} - (P^b+K) + (P^s-k)\widetilde{\kappa}\widetilde{\Lambda}\right], \\ V^s &= \frac{C(1-l)}{r+\Lambda} + (P^s-k)\widetilde{\Lambda} + V^b\widetilde{\eta}\widetilde{\Lambda} = \frac{\widetilde{\Lambda}}{1-\widetilde{\kappa}\widetilde{\eta}(\widetilde{\Lambda})^2}\left[\frac{C(1-l)}{\Lambda} + \frac{C}{r+\kappa}\widetilde{\eta}\widetilde{\Lambda} + (P^s-K) - (P^b+K)\widetilde{\eta}\widetilde{\Lambda}\right], \end{aligned}$$

and

$$\begin{aligned} U^s &= (M^{bid}-P^s)\widetilde{\lambda} + U^b\widetilde{\eta}\widetilde{\Lambda} = \frac{\widetilde{\lambda}}{1-\widetilde{\kappa}\widetilde{\eta}(\widetilde{\Lambda})^2}\left[(M^{bid}-P^s) + (P^b-M^{ask})\widetilde{\eta}\widetilde{\Lambda}\right], \\ U^b &= (P^b-M^{ask})\widetilde{\lambda} + U^s\widetilde{\kappa}\widetilde{\Lambda} = \frac{\widetilde{\lambda}}{1-\widetilde{\kappa}\widetilde{\eta}(\widetilde{\Lambda})^2}\left[(P^b-M^{ask}) + (M^{bid}-P^s)\widetilde{\kappa}\widetilde{\Lambda}\right]. \end{aligned}$$

After some algebra one obtains

$$\begin{aligned} V^o - V^b &= \frac{C}{r+\kappa} + V^s\widetilde{\kappa} - V^b \\ &= \frac{C}{r+\kappa} + \frac{\widetilde{\Lambda}}{1-\widetilde{\kappa}\widetilde{\eta}(\widetilde{\Lambda})^2}\left[\frac{C(1-l)}{\Lambda}\widetilde{\kappa} - \frac{C(1-l)}{r+\Lambda}\widetilde{\kappa} - \frac{C}{r+\kappa}(1-\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda}) + \right. \\ &\quad \left. +(P^b+K)(1-\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda}) + (P^s-K)\widetilde{\kappa}(1-\widetilde{\Lambda})\right], \\ V^s - V^{no} &= V^s - V^b\widetilde{\eta} \\ &= \frac{\widetilde{\Lambda}}{1-\widetilde{\kappa}\widetilde{\eta}(\widetilde{\Lambda})^2}\left[\frac{C(1-l)}{\Lambda}(1-\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda}) - \frac{C}{r+\kappa}\widetilde{\eta}(1-\widetilde{\Lambda}) + \right. \\ &\quad \left. +(P^b+K)\widetilde{\eta}(1-\widetilde{\Lambda}) + (P^s-K)(1-\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda})\right]. \end{aligned}$$

Substituting these expressions into (15) and (16) yields

$$
\begin{aligned}
P^b &= \frac{C}{r+\kappa}w \\
&+\frac{\widetilde{\Lambda}}{1-\widetilde{\kappa}\widetilde{\eta}(\widetilde{\Lambda})^2}\left[\frac{C(1-l)}{\Lambda}\widetilde{\kappa}-\frac{C(1-l)}{r+\Lambda}\widetilde{\kappa}-\frac{C}{r+\kappa}(1-\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda})+(P^b+K)(1-\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda})+(P^s-K)\widetilde{\kappa}(1-\widetilde{\Lambda})\right]w \\
&+M^{ask}(1-w)-\frac{\widetilde{\lambda}}{1-\widetilde{\kappa}\widetilde{\eta}(\widetilde{\Lambda})^2}\left[(M^{bid}-P^s)+(P^b-M^{ask})\widetilde{\eta}\widetilde{\Lambda}\right]\widetilde{\kappa}(1-w), \\
P^s &= \frac{\widetilde{\Lambda}}{1-\widetilde{\kappa}\widetilde{\eta}(\widetilde{\Lambda})^2}\left[\frac{C(1-l)}{\Lambda}(1-\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda})-\frac{C}{r+\kappa}\widetilde{\eta}(1-\widetilde{\Lambda})+(P^b+K)\widetilde{\eta}(1-\widetilde{\Lambda})+(P^s-K)(1-\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda})\right]w \\
&+M^{bid}(1-w)+\frac{\widetilde{\lambda}}{1-\widetilde{\kappa}\widetilde{\eta}(\widetilde{\Lambda})^2}\left[(P^b-M^{ask})+(M^{bid}-P^s)\widetilde{\kappa}\widetilde{\Lambda}\right]\widetilde{\eta}(1-w).
\end{aligned}
$$

These expressions can be rewritten as

$$
\begin{aligned}
P^bA_1(w) &= \left[\frac{C(1-l)}{\Lambda}\widetilde{\kappa}\widetilde{\Lambda}+\frac{C}{r+\kappa}\right](1-\widetilde{\Lambda})w+KA_2(\widetilde{\kappa},w) \\
&\quad +M^{ask}A_3(w)-M^{bid}\widetilde{\kappa}\widetilde{\lambda}(1-w)+P^s\widetilde{\kappa}A_4(w) \\
P^sA_1(w) &= \left[\frac{C(1-l)}{\Lambda}(1-\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda})-\frac{C}{r+\kappa}\widetilde{\eta}(1-\widetilde{\Lambda})\right]\widetilde{\Lambda}w-KA_2(\widetilde{\eta},w) \\
&\quad -M^{ask}\widetilde{\eta}\widetilde{\lambda}(1-w)+M^{bid}A_3(w)+P^b\widetilde{\eta}A_4(w),
\end{aligned} \tag{37}
$$

where we have defined

$$
\begin{aligned}
A_1(w) &\equiv 1-\widetilde{\Lambda}w-\widetilde{\kappa}\widetilde{\eta}(\widetilde{\Lambda})^2(1-w)+\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda}\widetilde{\lambda}(1-w), \\
A_2(x,w) &\equiv [1-\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda}-x(1-\widetilde{\Lambda})]\widetilde{\Lambda}w, \\
A_3(w) &\equiv [1-\widetilde{\kappa}\widetilde{\eta}(\widetilde{\Lambda})^2+\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda}\widetilde{\lambda}](1-w), \\
A_4(w) &\equiv (1-\widetilde{\Lambda})\widetilde{\Lambda}w+\widetilde{\lambda}(1-w), \\
A_5(w) &\equiv A_1(w)-\widetilde{\kappa}\widetilde{\eta}\frac{A_4(w)^2}{A_1(w)}.
\end{aligned} \tag{38}
$$

The system of equations (37) can be solved to yield expressions (35) and (36), where we have defined

$$
\begin{aligned}
\Psi_1 &\equiv \frac{\widetilde{\kappa}\widetilde{\Lambda}}{A_5(w)}\left[1-\widetilde{\Lambda}+(1-\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda})\frac{A_4(w)}{A_1(w)}\right], \\
\Psi_2 &\equiv \frac{(1-\widetilde{\Lambda})}{A_5(w)}\left[1-\widetilde{\kappa}\widetilde{\eta}\frac{A_4(w)}{A_1(w)}\widetilde{\Lambda}\right], \\
\Psi_3 &\equiv \frac{\widetilde{\Lambda}}{A_5(w)}\left[1-\widetilde{\kappa}\widetilde{\eta}\widetilde{\Lambda}+\widetilde{\kappa}\widetilde{\eta}(1-\widetilde{\Lambda})\frac{A_4(w)}{A_1(w)}\right], \\
\Psi_4 &\equiv \frac{\widetilde{\eta}(1-\widetilde{\Lambda})}{A_5(w)}\left[\widetilde{\Lambda}-\frac{A_4(w)}{A_1(w)}\right], \\
\Pi_1(x,y) &\equiv \frac{1}{A_5(w)}\left[A_2(x)-xA_2(y)\frac{A_4(w)}{A_1(w)}\right], \\
\Pi_2 &\equiv \frac{1}{A_5(w)}\left[A_3-\widetilde{\kappa}\widetilde{\eta}\frac{A_4(w)}{A_1(w)}\widetilde{\lambda}\right], \\
\Pi_3 &\equiv \frac{1}{A_5(w)}\left[A_3\frac{A_4(w)}{A_1(w)}-\widetilde{\lambda}\right].
\end{aligned} \tag{39}
$$

PROPOSITION 2: *The optimal size of the insurer's dealer network N^* is given by the following condition*

$$\frac{rV^b}{(\lambda N^*)^2}\left(1+\frac{\kappa}{r+\kappa}\frac{\eta}{r+\eta}\left(\frac{\lambda N^*}{r+\lambda N^*}\right)^2\right)=-\left(\frac{r+\lambda N^*}{\lambda N^*}-\frac{\eta}{r+\eta}\frac{\kappa}{r+\kappa}\frac{\lambda N^*}{r+\lambda N^*}\right)\frac{dV^r}{d\lambda N^*}+ \tag{40}$$
$$+\frac{\kappa}{r+\kappa}\frac{r}{(r+\lambda N^*)^2}\left(\frac{C(1-L)}{r}-P^s+K\right)+\frac{dP^s}{d\lambda N^*}\frac{\kappa}{r+\kappa}\frac{\lambda N^*}{r+\lambda N^*}-\frac{dP^b}{d\lambda N^*}.$$

Proof: Since $\frac{d\widehat{V}^b}{dN}=\frac{dV^b}{dN}+\frac{dV^r}{dN}$ we need to calculate the derivative of V^b with respect to N. We start by rewriting the expression (6) as

$$V^b=\left(\frac{C}{r+\kappa}+\frac{C(1-L)}{r+\lambda N}\frac{\kappa}{r+\kappa}+\left(E[P^s]+V^b\frac{\eta}{r+\eta}-K\right)\frac{\kappa}{r+\kappa}\frac{\lambda N}{r+\lambda N}-P^b-K\right)\frac{\lambda N}{r+\lambda N}, \tag{41}$$

and solve it for V^b to obtain

$$V^b=\frac{\frac{\lambda N}{r+\lambda N}}{1-\frac{\kappa}{r+\kappa}\frac{\eta}{r+\eta}(\frac{\lambda N}{r+\lambda N})^2}\left[\frac{C}{r+\kappa}+\frac{C(1-L)}{r+\lambda N}\frac{\kappa}{r+\kappa}+(P^s-K)\frac{\kappa}{r+\kappa}\frac{\lambda N}{r+\lambda N}-(P^b+K)\right]. \tag{42}$$

Taking into account that

$$\frac{d\left(\frac{\lambda N}{r+\lambda N}\right)}{d\lambda N}=\frac{r}{(r+\lambda N)^2}, \tag{43}$$

we obtain the following expression for the derivative in question

$$\frac{dV^b}{d\lambda N}=\left(\frac{r+\lambda N}{\lambda N}-\frac{\eta}{r+\eta}\frac{\kappa}{r+\kappa}\frac{\lambda N}{r+\lambda N}\right)^{-1}\left[\frac{rV^b}{(\lambda N)^2}\left(1+\frac{\kappa}{r+\kappa}\frac{\eta}{r+\eta}\left(\frac{\lambda N}{r+\lambda N}\right)^2\right)-\right. \tag{44}$$
$$\left.-\frac{\kappa}{r+\kappa}\frac{r}{(r+\lambda N)^2}\left(\frac{C(1-L)}{r}-P^s+K\right)+\frac{dP^s}{d\lambda N}\frac{\kappa}{r+\kappa}\frac{\lambda N}{r+\lambda N}-\frac{dP^b}{d\lambda N}\right], \tag{45}$$

which after setting $\frac{d\widehat{V}^b}{dN}=0$ leads to 40 after some algebra.

References

Afonso, Gara, Anna Kovner, and Antoinette Schoar, 2013, "Trading partners in the interbank lending market," Federal Reserve Bank of New York Staff Report.

Atkeson, Andrew, Andrea Eisfeld, and Pierre-Olivier Weill, 2015, "Entry and exit in OTC Derivatives Markets," *Econometrica* 83, 2231-2292.

Babus, Ana and Tai-Wei Hu, 2016, "Endogenous Intermediation in Over-the-Counter Markets," *Journal of Financial Economics*, Forthcoming.

Babus, Ana, 2016, "The Formation of Financial Networks," *RAND Journal of Economics* 47, 239-272.

Babus, Ana and Peter Kondor, 2013, "Trading and Information Diffusion in Over-The-Counter Markets," Working Paper, Federal Reserve Bank of Chicago.

Bernhardt, Dan, Vladimir Dvoracek, Eric Hughson, and Ingrid Werner, 2005, "Why do large orders receive discounts on the London Stock Exchange?" *Review of Financial Studies* 18, 1343-1368.

Bessembinder, Hendrik, Stacey Jacobsen, William Maxwell, and Kumar Venkataraman, 2016, "Capital Commitment and Illiquidity in Corporate Bonds," working paper.

Bessembinder, Hendrik, William Maxwell, and Kumar Venkataraman, 2006, "Market transparency, liquidity externalities, and institutional trading costs in corporate bonds," *Journal of Financial Economics* 82, 251-288.

Board, Simon, 2011, "Relational Contracts and the Value of Loyalty," *American Economic Review* 101, 3349-3367.

Chang, Briana and Shengxing Zhang, 2015, "Endogenous market making and network formation," Mimeo.

Choo, Eugene, 2015, "Dynamic marriage matching: An empirical framework,", *Econometrica* 83, 1373-1423.

Colliard, Jean-Edouard and Gabrielle Demange, 2014, "Cash Providers: Asset Dissemination Over Intermediation Chains," Paris School of Economics Working Paper No. 2014-8.

Condorelli, Daniele and Andrea Galeotti, 2012, "Endogenous Trading Networks," Working Paper, University of Essex.

Diamond, Peter, 1971, "A model of price adjustment," *Journal of Economic Theory* 3, 156-168.

DiMaggio, Marco, Amir Kermani, and Zhaogang Song, 2015, "The Value of Trading Relationships in Turbulent Times ," Mimeo.

Duffie, Darrell, Garleanu, Nicolae, and Lasse Pedersen, 2005, "Over the counter markets," *Econometrica* 73, 1815-1847.

Duffie, Darrell, Garleanu, Nicolae, and Lasse Pedersen, 2007, "Valuation in over the counter markets," *Review of Financial Studies* 20, 1865-1900.

Edwards, Amy, Lawrence Harris, and Michael Piwowar, 2005, "Corporate bond markets transparency and transaction costs," *Journal of Finance* 62, 1421-1451.

Eckstein, Zvi, and Gerard J. Van Den Berg, 2007, "Empirical Labor Search: A Survey," *Journal of Econometrics*, 136, 531564.

Ellul, Andrew, 2011, "Regulatory pressure and fire sales in the corporate bond market," *Journal of Financial Economics* 101, 596-620.

Feldhutter, Peter, 2012, "The same bond at different prices: Identifying search frictions and selling pressures," *Review of Financial Studies* 25, 1155-1206.

Gale, Douglas M., and Shachar Kariv, 2007, "Financial Networks," *American Economic Review* 97, 99-103.

Gavazza, Alessandro, 2016, "An Empirical Equilibrium Model of a Decentralized Asset Market," *Econometrica* 84, 1755-1798.

Glode, Vincent and Christian C. Opp, 2014, "Adverse Selection and Intermediation Chains," Working Paper, University of Pennsylvania.

Gofman, Michael, 2011, "A Network-Based Analysis of Over-the-Counter Markets," Working Paper.

Green, Richard C., Burton Hollifield, and Norman Schürhoff, 2007, "Financial Intermediation and the Costs of Trading in an Opaque Market," *Review of Financial Studies* 20(2), 275-314.

Green, Richard C., Dan Li, and Norman Schürhoff, 2010, "Price Discovery in Illiquid Markets: Do Financial Asset Prices Rise Faster Than They Fall?," *The Journal of Finance* 65(5), 1669-1702.

Guerrieri, Veronica, Robert Shimer, and Randall Wright, 2010, "Adverse selection in competitive search equilibrium," *Econometrica* 78, 1823-1862.

Guerrieri, Veronica and Robert Shimer, 2014, "Dynamic adverse selection: A theory of illiquidity, fire sales, and flight to quality," *American Economic Review* 104, 1875-1908.

Harris, Christopher and John Vickers, 1985, "Perfect Equilibrium in a Model of a Race," *Review of Economic Studies* 52, 193-209.

Harris, Larry, 2015, "Transaction Costs, Trade Throughs, and Riskless Principal Trading in Corporate Bond Markets," Mimeo.

Hitsch, Gunter J., Ali Hortaçsu, and Dan Ariely, 2010, "Matching and Sorting in Online Dating Markets," *American Economic Review* 100, 130-163.

Hollifield, Burton, Artem Neklyudov, and Chester Spatt, 2015, "Bid-Ask Spreads and the Pricing of Securitizations: 144a vs. Registered Securitizations," Working Paper, Carnegie Mellon University.

Hugonnier, Julien, Benjamin Lester, and Pierre-Olivier Weill, 2015, "Heterogeneity in Decentralized Asset Markets," Working Paper, Swiss Finance Institute.

Jackson, Matthew O. and Asher Wolinsky, 1996, "A Strategic Model of Social and Economic Networks," *Journal of Economic Theory* 71, 44-74.

Koijen, Ralph S.J. and Motohiro Yogo, 2015, "The Cost of Financial Frictions for Life Insurers," *American Economic Review* 105, 445-475.

Lagos, Ricardo and Guillaume Rocheteau, 2007, "Search in asset markets: Market structure, liquidity, and welfare," *American Economic Review* 97, 198-202.

Lagos, Ricardo and Guillaume Rocheteau, 2009, "Liquidity in asset markets with search frictions," *Econometrica* 77, 403-426.

Lester, Benjamin, Guillaume Rocheteau, and Pierre-Olivier Weill, 2015, "Competing for Order Flow in OTC Markets," *Journal of Money, Credit and Banking* 47, 77-126.

Moen, Espen R., 1997, “Competitive Search Equilibrium,” *Journal of Political Economy* 105, 385-411.

Neklyudov, Artem, 2014, “Bid-Ask Spreads and the Over-the-Counter Interdealer Markets: Core and Peripheral Dealers,” Working Paper, Swiss Finance Institute.

Neklyudov, Artem and Batchimeg Sambalaibat, 2016, “Endogenous Specialization and Dealer Networks,” Working Paper, Swiss Finance Institute.

O’Hara, Maureen, Yihui Wang, and Xing Zhou, 2015, “The Best Execution of Corporate Bonds,” Working paper, Cornell University.

Postel-Vinay, Fabien, and Jean-Marc Robin, 2002, “Equilibrium wage dispersion with worker and employer heterogeneity,” *Econometrica* 70, 2295-2350.

Vayanos, Dimitri and Tan Wang, 2007, “Search and endogenous concentration of liquidity in asset markets,” *Journal of Economic Theory* 136, 66-104.

Üslü, Semih, 2015, “Pricing and Liquidity in Decentralized Asset Markets,” Working Paper, UCLA.

Weill, Pierre-Olivier, 2007, “Leaning against the wind,” *Review of Economic Studies* 74, 1329-1354.