Equilibrium Models for Derivatives Markets with Frictions

THÈSE N^{O} 8703 (2018)

PRÉSENTÉE LE 10 AOÛT 2018

AU COLLÈGE DU MANAGEMENT DE LA TECHNOLOGIE
CHAIRE DU PROF. ASSOCIÉ MALAMUD
PROGRAMME DOCTORAL EN FINANCE

ÉCOLE POLYTECHNIQUE FÉDÉRALE DE LAUSANNE

POUR L'OBTENTION DU GRADE DE DOCTEUR ÈS SCIENCES

PAR

Yuan ZHANG

acceptée sur proposition du jury:

Prof. P. Collin Dufresne, président du jury
Prof. S. Malamud, directeur de thèse
Prof. N. Shuerhoff, rapporteur
Prof. A. Trolle, rapporteur
Prof. J. Hugonnier, rapporteur

ÉCOLE POLYTECHNIQUE
FÉDÉRALE DE LAUSANNE

Suisse
2018

Acknowledgements

I thank my supervisor, Semyon Malamud, for his advice and continuous support. His guidance and curiosity on research has greatly helped me during the years of research and writing of this thesis. I also thank the members of my thesis committee: Pierre Collin Dufresne, Julien Hugonnier, Norman Schuerhoff, Anders Trolle, for agreeing to sit on my thesis jury, and for their insightful discussions and encouragement. I would also like to thank Ruediger Fahlenbrach and Erwan Morellec for their kind advices and encouragement.
I thank Hui Chen for sponsoring my visits at MIT, and for many useful comments and suggestions on my research.
I thank my fellow officemates Damien, Thomas, and Vincent for the stimulating discussions, and for all the fun we have had in the last few years. A special thanks goes to Maxime for helping me translate the abstract of the thesis to French.
I thank my friends in Lausanne and Boston for taking me out to social events, climbing and running sessions.
Last, I would like to thank my family: my parents and my fiancée, Lingyu, for supporting me unconditionally throughout writing this thesis and for encouraging me to chase my dreams.

Lausanne, May 4, 2018 Y. Z.

Abstract

This thesis develops equilibrium models, and studies the effects of market frictions on risk-sharing, derivatives pricing, and trading patterns.

In the chapter titled "Imbalance-Based Option Pricing", I develop an equilibrium model of fragmented options markets in which option prices and bid-ask spreads are determined by the nonlinear risk imbalance between dealers and customers. In my model, dealers optimally exploit their market power and charge higher spreads for deep out-of-the-money (OTM) options, leading to an endogenous skew in both prices and spreads. In stark contrast to theories of price pressure in option markets, I show how wealth effects can make customers' net demand for options be negatively correlated with option prices. Under natural conditions, the skewness risk premium is positively correlated with the variance risk premium, consistent with the data.

In the chapter titled "The Demand for Commodity Options", we develop a simple equilibrium model in which commercial hedgers, i.e., producers and consumers, use commodity options and futures to hedge price and quantity risk. We derive an explicit relationship between expected futures returns and the hedgers' demand for out-of-the-money options, and show that the demand for both calls and puts are positively related to expected returns, and the relationship is asymmetric, tilted towards puts. We test and confirm the model predictions empirically using the commitment of traders report from CFTC.

In the chapter titled "Electronic Trading in OTC Markets vs. Centralized Exchange", we model a two-tiered market structure in which an investor can trade an asset on a trading platform with a set of dealers who in turn have access to an interdealer market. The investor's order is informative about the asset's payoff and dealers who were contacted by the investor use this information in the interdealer market. Increasing the number of contacted dealers lowers markups through competition but increases the dealers' costs of providing the asset through information leakage. We then compare a centralized market in which investors can trade among themselves in a central limit order book to a market in which investors have to use the electronic platform to trade the asset. With imperfect competition among dealers, investor welfare is higher in the centralized market if private values are strongly dispersed or if the mass of investors is large.

Key words: Market Structure; Market Power; Risk Imbalance; Hedging Pressure; Option Liquidity; Risk Premia; Information Leakage

Résumé

Cette thèse développe des modèles d'équilibre et étudie les effets des frictions de marché sur les structures de prix et d'échange des produits dérivés.

Dans le chapitre intitulé «Valorisation d'options basée sur les déséquilibres», je développe un modèle d'équilibre des marchés d'options fragmentées dans lequel les prix des options et les écarts entre les cours acheteur et vendeur sont déterminés par le déséquilibre de risque non linéaire entre les négociants et les clients. Dans mon modèle, les négociants exploitent de manière optimale leur pouvoir de marché et facturent des spreads plus élevés pour des options hors la monnaie (HLM), ce qui entraîne une distorsion endogène des prix et des spreads. À l'opposé des théories de la pression sur les prix sur les marchés d'options, je montre comment les effets de richesse peuvent faire en sorte que la demande nette d'options des consommateurs soit corrélée négativement avec les prix des options. Dans des conditions naturelles, la prime de risque d'asymétrie est positivement corrélée avec la prime de risque de variance, en adéquation avec les données.

Dans le chapitre intitulé «La demande d'options sur matières premières», nous développons un modèle d'équilibre simple dans lequel les sociétés de couverture commerciales, c'est-à-dire les producteurs et les consommateurs, utilisent des options sur matières premières et des contrats à terme afin de s'assurer contre les risques de prix et de quantité. Nous établissons une relation explicite entre les rendements futurs attendus et la demande des sociétés de couverture pour les options hors la monnaie. Nous montrons que la demande pour les options d'achat et les options de vente est positivement liée aux rendements attendus, et la relation est asymétrique, orientée vers les options de vente. Nous testons et confirmons les prédictions du modèle de manière empirique en utilisant une importante base de données d'opérations et de cours d'options sur des matières premières.

Dans le chapitre intitulé «La négociation électronique dans les marchés de gré à gré vs. les marchés centralisés», nous modélisons une structure de marché à deux niveaux dans laquelle un investisseur peut négocier un actif sur une plateforme de négociation avec un ensemble de courtiers qui ont accès à un marché intercourtiers. L'ordre de l'investisseur est instructif sur le gain de l'actif et les courtiers contactés par l'investisseur utilisent cette information sur le marché intercourtiers. Augmenter le nombre de courtiers contactés réduit les majorations de prix en raison de la concurrence accrue, mais augmente les coûts des courtiers pour fournir l'actif en raison de fuites d'informations. Nous comparons ensuite un marché centralisé dans lequel les investisseurs peuvent négocier entre eux au moyen d'un carnet d'ordres de bourse à cours limités à un marché dans lequel les investisseurs doivent utiliser la plate-

forme électronique pour négocier l'actif. En présence d'une concurrence imparfaite entre les courtiers, le bien-être des investisseurs est plus important sur le marché centralisé si les valeurs privées sont fortement dispersées ou si la masse des investisseurs est importante.
Mots clés : Structure de Marché ; Pouvoir de Marché ; Déséquilibre des Risques ; Pression de Couverture ; Option de Liquidité ; Primes de Risque ; Fuite d'Informations

Contents

List of Figures

List of Tables

Introduction

Options play a fundamental role in the functioning of modern financial markets. Jumps, trading costs, and stochastic volatility are among many factors that make options non-redundant and attractive vehicles for spanning risks. Despite many options are traded on exchange, the market structure is highly fragmented and has a pronounced two-tiered structure, whereby dealers trade with customers in the dealer-to-customer (D2C) segment and then rebalance their inventories with each other in the dealer-to-dealer (D2D) segment.

In light of these facts, in the first chapter I develop an equilibrium model of fragmented options markets in which option prices and bid-ask spreads are determined by the nonlinear risk imbalance between dealers and customers. Consequently, option prices in my model consists of three parts: compensation for the fundamental risk, compensation for dealers' inventory risk which arises endogenously due to distorted risk sharing, and markups customers pay to dealers. The latter two are specific to my model and can play a role in resolving the main empirical puzzles in option pricing and trading patterns.

In the second chapter, we try to understand the origin of non-linear endowment risks in commodity markets. To this end, we develop a simple, two period general equilibrium model populated by three types of agents: commodity producers, commodity consumers, and speculators. We assume that, in addition to price risk, producers face quantity risk: In this case, options become necessary to hedge the endowment risk, and producers may in fact find it optimal to take a long position in the futures contract, and expected futures returns are positive if and only if price risk is larger than quantity risk.

Many derivatives contracts are traded in a two-tiered market structure, despite the fact that dealers can have market power, and that state prices and risk-sharing in the economy are distorted. In the third chapter, we build an equilibrium model to show that investors prefer to trade in such a two-tiered market structure when they are concerned about information leakage.

1 Imbalance-Based Option Pricing

Yuan Zhang[1]

1 – EPFL and Swiss Finance Institute

I develop an equilibrium model of fragmented options markets in which option prices and bid-ask spreads are determined by the nonlinear risk imbalance between dealers and customers. In my model, dealers optimally exploit their market power and charge higher spreads for deep out-of-the-money (OTM) options, leading to an endogenous skew in both prices and spreads. In stark contrast to theories of price pressure in option markets, I show how wealth effects can make customers' net demand for options be negatively correlated with option prices. Under natural conditions, the skewness risk premium is positively correlated with the variance risk premium, consistent with the data.

1.1 Introduction

Options play a fundamental role in the functioning of modern financial markets. Jumps, trading costs, and stochastic volatility are among many factors [1] that make options non-redundant and attractive vehicles for spanning risks. In addition to the fundamental risk factors, agents' exposure on each possible future state may as well be *nonlinear*, resulting in another source of option demand. [2] This extra demand has no effect on equilibrium option prices in the absence of trading frictions.

However, despite many options are traded on exchange, the market structure is highly fragmented and has a pronounced two-tiered structure, whereby dealers trade with customers in

[1] Jumps refer to discontinuous price movements; trading costs refer to transaction fees and financing/short-selling constraints; stochastic volatility refers to the randomness in the range of price movements.

[2] For instance, the advancement of new technology may on average improve the performance of the stock market but can have adverse effects on those industries being replaced (think of the idiosyncratic income shock in Constantinides and Duffie [1996]).

the dealer-to-customer (D2C) segment and then rebalance their inventories with each other in the dealer-to-dealer (D2D) segment. [3]

In light of these facts, I develop an equilibrium model of fragmented options markets in which option prices and bid-ask spreads are determined by the nonlinear risk imbalance between dealers and customers. I show that dealers optimally exploit their market power and charge higher spreads for deep out-of-the-money (OTM) options, leading to an endogenous skew in both prices and spreads. Consequently, option prices in my model consists of three parts: compensation for the fundamental risk, compensation for dealers' inventory risk which arises endogenously due to distorted risk sharing, and markups customers pay to dealers. The latter two are specific to my model and can play a role in resolving the main empirical puzzles in option pricing and trading patterns.

My model works as follows. There are two trading rounds: A round of D2C trade is followed by a round of D2D trade. The D2D trade happens in a centralized exchange, while the D2C trade is an outcome of bilateral bargaining. In the D2C round, each of the dealers is randomly assigned a customer and the two share nonlinear endowment risk by bargaining on option prices across all strikes. The bargaining outcome depends on the rational expectations of both dealers and customers regarding the future equilibrium prices in the D2D trade. Because markets in the D2D round are complete, its prices are determined by the total inventories that the dealers accumulate from trading with customers. This trading process determines a fixed point system for equilibrium prices in both trading rounds. I show explicitly how the distribution of the nonlinear risk enters into the pricing kernels for all exchanges.

The Black-Scholes formula is acknowledged to be consistent with equilibrium in a frictionless market if all agents have the same constant relative risk aversion (CRRA) preferences and the aggregate endowment is log-normal. I show that this result still holds when dealers' bargaining power is zero: In fact, in this case, equilibrium in my model always coincides with that in the frictionless model. However, this result breaks down as soon as dealers have some market power. Thus, customers are not able to trade at the D2D option prices and, hence, efficient risk sharing between dealers and customers is not feasible. This market power effect then leads to a pecuniary externality: Dealers do not internalize the impact that their market power has on the total inventories of the dealers' population; the latter determines the total risk to be shared in the D2D trade and hence its pricing kernel.

[3] I use the word "dealers" to denote option-trading specialists, designated market makers, members of a multi-dealer platform, or any entity that has direct access to option markets; "customers" refers to anyone who uses options but cannot access the markets directly and has to trade with dealers. For customers who trade options on exchange, (i) large orders (e.g., block trade), complex orders (e.g., trade involving multiple strikes), and orders with non-standard strike/maturity are often negotiated privately with dealers before execution on the exchange; (ii) in the US, 15 options exchanges at the moment make sourcing and providing liquidity extremely difficult, and (iii) retail orders are usually aggregated and internalized by brokers. Moreover, many options are also traded over-the-counter, for example, Back for International Settlements (BIS) reports that the notional amount outstanding for equity-linked options is \$ 3,987 billion and for commodity options is \$ 378 billion for the first-half of 2017 (Semiannual OTC derivatives statistics).

My model can generate the skew in both the percentage bid-ask spreads (measured in $ terms), and the implied volatility (IV) curve. Specifically, the spreads for out-of-the-money options are larger than those for at-the-money options, and the implied volatilities for OTM puts are higher than those for OTM calls (e.g., equity index options). To understand this, I consider a customer endowed with short positions in options, trades in an almost competitive D2C exchange. With little market power, his dealer, in response to the buying demand, optimally quotes a D2C pricing kernel that is the sum of a mean-preserving spread and the D2D pricing kernel. This D2C trade results in the out-of-the-money option prices to increase more than the prices of at-the-money options, as the former loads more on the tails of the pricing kernel. For customers endowed with long positions in options, the opposite happens (i.e., out-of-the-money option prices decrease more than the prices of at-the-money options). Hence, dealers' optimal quoting strategies on the D2C exchanges generate the price wedge between option buyers and option sellers, resulting in higher spreads for out-of-the-money options than at-the-money options. Further, the IV smile derived from the average option prices across D2C exchanges ('mid' prices) is skewed to the left and the variance risk premium [4] is positive if customers' net buy of options is positive and skewed to the left (i.e., buying more OTM puts than calls).

A well-known puzzle in the literature on demand-based option pricing [5] is that in recent years customers' net buy of options has become negatively correlated with the variance risk premium (Chen et al., forthcoming;Constantinides and Lian, 2015), which is in stark contrast to the earlier observations in Gârleanu et al. [2009]. In this paper, I use the Open/Close dataset from the largest three US options exchanges [6] to construct customers' net option demand for liquid exchange-traded fund (ETF) options and show that this demand is often negatively correlated with the corresponding variance risk premium, confirming the puzzle. In addition, I document another puzzling observation: The relation between the downside risk [7] and the variance risk implied from options data is often positive, despite a negative relation implied from historical underlying returns.

My model can address both puzzles. For the first puzzle, because of the wealth effect, when dealers' net worth drops, their effective risk aversion rises, increasing their incentive to smooth consumption. As a result, dealers find it optimal to give price concessions to customers, inducing the latter to take more risk off dealers' balance sheets. This active hedging activity by dealers explains the first puzzle. For the second puzzle, in my model, in addition to the fundamental risk factors, the option prices [8] are affected by the distribution of the nonlinear risk

[4] The variance risk premium is defined as the difference between the risk-neutral variance and the physical variance.

[5] Bollen and Whaley [2004] find that changes in implied volatility are correlated with option order-flow imbalance. Gârleanu et al. [2009] provide a theoretical model and empirical evidence to demonstrate the importance of customers' option demand in determining option prices (the level and the skew). Fournier and Jacobs [2016] show that dealers' inventory and wealth matter.

[6] Specifically, the Chicago Board of Option Exchange (CBOE), the NASDAQ Philadelphia Option Exchange (PHLX), and the International Stock Exchange (ISE).

[7] Downside risk is measured as the risk-neutral skewness.

[8] The volume-weighted average equilibrium D2C option prices, to be more precise.

between dealers and customers. Intuitively, as dealers are risk-averse, a disaster risk generates upward price pressure on OTM puts relative to OTM calls. At the same time, customers with short positions on OTM calls create buying pressure, resulting in downward price pressure on OTM puts relative to OTM calls. This extra source of option premium due to dealers' market power can explain the second puzzle, as long as the occurrence of endowment shocks on the upside are more frequent than that of the fundamental shock.

On the other hand, the demand pressure theory in Gârleanu et al. [2009] predicts customers' demand pressure causes option prices to move up, not down. Nevertheless, their model can potentially explain the second puzzle, as exogenous option demands from customers may allow changes in option prices due to demand pressure (if the source of the market incompleteness is chosen carefully), which is different from price changes due to a shift in physical density. Meanwhile, to explain the first puzzle, Chen et al. (forthcoming) argue that dealers become more risk-averse [9] when the perceived intensity of the disaster risk is high, and hence, they cannot accommodate the option demand from customers, in turn causing option prices to increase and option demand to decrease. However, their model cannot explain the second puzzle as customers' option demand is tightly linked to the fundamental risk: Indeed, in their model, in light of a disaster risk, customers buy protection from dealers, pushing up OTM put prices more relative to OTM call prices while simultaneously increasing the variance risk premium.

To test the prediction of my model on the second puzzle, I use liquid ETF options. Specifically, I show that the cross-sectional difference in the correlation between the risk-neutral variance and the risk-neutral skewness is explained by the proxy for the shape of the customers' net option demand. Moreover, in the time series, the panel regression with controls for the physical skewness shows that the risk-neutral skewness increases with the risk-neutral variance when the customers' demand is skewed towards the OTM call options.

1.1.1 Related Literature

My paper is related to several strands of literature.

First, the literature on equilibrium option pricing (cited above) explicitly models changes in option prices due to supply and/or demand shocks. I contribute to the literature by modeling a realistic two-tier market structure and showing that my model can explain several puzzles on option pricing and trading patterns. To this end, I use the approach of Malamud and Schrimpf [2017] and extend their model to allow for an intermediate bargaining power of the dealers.[10]

Second, my assumption that customers can only trade options with dealers is closely related to the proliferating literature on intermediary-based asset pricing. Bernanke and Gertler

[9] To also capture the notion of constrained dealers, Constantinides and Lian [2015] specify a Value-at-Risk constraint for risk-neutral dealers.

[10] In contrast to my paper, Malamud and Schrimpf [2017] have a dynamic model, but they assume that dealers have monopoly power and use their model to study monetary policy pass-through.

[1989] and Moore and Kiyotaki [1997] highlight the importance of intermediation frictions in determining equilibrium prices. Subsequently, the financial frictions are micro-founded as limits-to-arbitrage [Shleifer and Vishny, 1997, Gromb and Vayanos, 2002], collateral constraints [Geanakoplos, 2010], Value-at-Risk constraints [Adrian and Shin, 2010, Adrian and Boyarchenko, 2012, Danielsson et al., 2012, Etula, 2013, Constantinides and Lian, 2015], equity financing constraints [Brunnermeier and Sannikov, 2014, He and Krishnamurthy, 2013, He et al., 2017], and margin constraints [Brunnermeier and Pedersen, 2009], among others. I recognize the importance of the limited risk-bearing capacity of dealers and model risk-averse dealers with market power. My assumption on nonlinear endowments is in the spirit of Constantinides and Duffie [1996] and Franke et al. [1998]; that is, the nonlinear risks render options non-redundant [11]. My assumption on fragmented markets [12] is borrowed from the literature on over-the-counter markets [e.g., Duffie et al., 2005, 2015, Atkeson et al., 2015, Malamud and Schrimpf, 2017]; that is, markets are fragmented and local prices are determined by bilateral bargaining. To the best of my knowledge, my paper is the first to incorporate these assumptions into an equilibrium option pricing model and show that they are indeed necessary to explain the main empirical puzzles in option pricing and trading patterns.[13]

Third, the literature on market microstructure identifies the following determinants of bid-ask spreads: dealers' inventory [Amihud and Mendelson, 1980, Ho and Stoll, 1983], asymmetric information [Copeland and Galai, 1983, Kyle, 1985, Glosten and Milgrom, 1985, Easley and O'Hara, 1987], and operation costs, among others. I further complement the literature by showing options spreads across strikes are non-trivially determined by customers' option demand and dealers' market power. The predictions are consistent with the empirical evidence, including George and Longstaff [1993], Cho and Engle [1999], and De Fontnouvelle et al. [2003].

Fourth, option prices imply a skewed and fat-tailed risk-neutral distribution for the underlying returns [Buraschi and Jackwerth, 2001, Bakshi et al., 2003]. According to Bates [2003], the literature on no-arbitrage option pricing models [e.g., Merton, 1976, Heston, 1993, Bates, 1996] does not fully capture the empirical properties of option prices. Another strand of literature specifies unspanned risk factors in representative agent models to generate fat-tailed risk-neutral distribution [e.g., Bollerslev et al., 2009, Drechsler and Yaron, 2011, Drechsler, 2013, Bekaert and Engstrom, 2017]. I take a different approach by showing how an implied volatility skew emerges naturally when customers with nonlinear endowments trade with non-competitive dealers.

Fifth, the literature on information content of options prices and trades contains documented

[11] Options are also non-redundant if agents have heterogeneous utilities [Bates, 2008, Baker and Routledge, 2016] or heterogeneous beliefs [Liu et al., 2005, Buraschi and Jiltsov, 2006].

[12] Papers that have also assumed market fragmentation but with different trading protocols include Basak and Cuoco [1998], Edmond and Weill [2012], and Goldstein et al. [2014]. For endogenous market fragmentation, see Alvarez et al. [2002] and Babus and Parlatore [2017].

[13] While Malamud and Schrimpf [2017] also have fragmented markets and state-contingent claims traded in their model, they do not study option pricing and do not have non-linear endowments' imbalance.

empirical evidence. On the information content of option prices, Bollerslev et al. [2009] show that variance risk premium (VRP) derived from S&P 500 index options predicts equity market returns, Trolle and Schwartz [2010] show that crude oil and natural gas returns are correlated with the contemporaneous VRP computed from their respective option prices, Trolle and Schwartz [2014] show that VRP and skewness risk premium are correlated with changes in the yield curve. On the information content of option trades, Chen et al. (forthcoming) show that customers' net buy of put options is negatively correlated with next-period S&P 500 returns as well as returns on other asset classes, Bollen and Whaley [2004], Cremers et al. [2015], Hu [2014], Muravyev [2016], and Malamud et al. [2017] show that the option order-imbalance measure is correlated with option premium and/or underlying returns, Pan and Poteshman [2006a] show that the put-call ratio is correlated with next-period single equity returns, Roll et al. [2010a, 2014] and Ge et al. [2015] show that the ratio between option volume and stock volume (O/S) is correlated with future equity volatility and returns. I further contribute to the literature by building a model and providing an explicit formula that extracts physical density of the underlying asset from the customers' net option demand and the option bid-ask quotes.

1.2 A Model of Fragmented Options Markets

I consider an economy with two rounds of trading and three time periods $t = 0^-, 0^+, 1$. At time $t = 1$, the state of the world, $X \sim P(X)$, is realized, and consumption takes place.

1.2.1 Market Structure

Markets are fragmented. Time 0^- is the D2C exchanges trading round: At this time, each dealer is randomly matched with a customer[14] and they trade contingent claims following a bargaining protocol described below. Time 0^+ is the D2D trading round: In this round, dealers trade with each other in a competitive, centralized inter-dealer market. In both rounds, agents trade derivatives, contingent on the realization of X. In addition, I assume that customers have access to the centralized market for trading the security with payoff X at time $t = 1$. I use s to denote the price of this security, and I normalize its supply to 1. In addition, all agents can trade a risk-free bond maturing at $t = 1$. The bond has an exogenous interest rate r and is in zero net supply.

Formally, my assumptions imply that there is a continuum of fragmented markets: a continuum of bilateral D2C exchanges [15], indexed by a pair (i, j) and a single D2D exchange. Throughout this paper, the following assumption is always present:

[14] Such market structures are frequently used in modeling decentralized trade [e.g., Duffie et al., 2005, 2015, Atkeson et al., 2015, Malamud and Schrimpf, 2017]. My model can be extended to allow for simultaneous trading with multiple customers (e.g., an over-the-counter (OTC) trading hub) or for trading with multiple dealers (order splitting). The bilateral trade assumption can be viewed as a reduced form of modeling aggregate customer orders.

[15] In practice, the continuum of D2C exchanges are exemplified by the large amount of order-flows from the customers if the options are exchange-traded, or by the over-the-counter trading desks of dealers if the options are OTC-traded.

Assumption 1. *In each exchange, markets are complete: Agents have access to a set of securities (e.g., options with a continuum of strikes) that spans the entire range of X, and there is no arbitrage in either trading rounds.*

From the fundamental theorem of asset pricing [see, e.g., Dybvig and Ross, 2003], no arbitrage in either exchange implies the existence of exchange-specific, positive state prices; that is, prices of Arrow Debreu contingent claims, $M^{\text{D2C}}_{(\text{i,j})}$ or M^{D2D} paying one unit of consumption good in state X and nothing in any other states. Since, by assumption, markets are complete, these state prices are unique. The exchange-specific price of an asset paying $W(X)$ at $t = 1$ is then given by the following:

$$E[M^{\text{D2C}}_{(\text{i,j})}(X)W(X)] \text{ or } E[M^{\text{D2D}}(X)W(X)].$$

Given the risk-free rate r, equilibrium state prices must also satisfy the no-arbitrage condition:

$$E[M^{\text{D2C}}_{(\text{i,j})}(X)] = E[M^{\text{D2D}}(X)] = \mathrm{e}^{-r}.$$

Similarly, since all agents can trade the underlying, the following no-arbitrage condition should hold in all exchanges:

$$E[M^{\text{D2C}}_{(\text{i,j})}(X)X] = E[M^{\text{D2D}}(X)X] = s.$$

1.2.2 Agents' Preferences and Endowments

The economy is populated by a continuum of dealers (indexed by $j \in [0,1]$) and a continuum of customers (indexed by $i \in [0,1]$).

For simplicity [16], I assume that all agents in the economy share the same utility function U defined on an interval $(\underline{X}, +\infty)$ with some $\underline{X} \geq -\infty$, satisfying the standard Inada conditions $U'(\underline{X}) = +\infty$, $U'(+\infty) = 0$. Each agent $a = i, j$ is initially endowed with a portfolio of options, represented by a nonlinear function $F_a(X)$, $a = i, j \in [0,1]$. I assume that the agents' option endowments net out, so that X represents the payoff of the "market portfolio". Formally, I make the following assumption.

Assumption 2.

$$X = \int_0^1 F^{\text{D}}_j(X)\,dj + \int_0^1 F^{\text{C}}_i(X)\,di.$$

[16] This is to remove the effects of heterogeneous risk-aversion on the pricing kernel.

1.2.3 Agents' Outside Options

Absent the D2C trading, customers can only trade X and the risk-free bond, and their indirect utility is given by

$$\bar{v}_i \equiv \max_{\beta_i} E[U(F_i^{\mathrm{C}}(X)+\beta_i(X-se^r))]. \tag{1.1}$$

In contrast to customers, a dealer j has access to complete markets and hence he can attain an arbitrary consumption profile $C_j(X)$ satisfying the budget constraint

$$E[M^{\mathrm{D2D}}(X)C_j(X)] = E[M^{\mathrm{D2D}}(X)F_j^{\mathrm{D}}(X)].$$

Denoting $G_j(X) = C_j(X) - F_j^{\mathrm{D}}(X)$, we can rewrite dealer j's indirect utility as

$$\bar{v}_j \equiv \max_{G_j} \{E[U(F_j^{\mathrm{D}}(X)+G_j(X))] : E[M^{\mathrm{D2D}}(X)G_j(X)] = 0\}. \tag{1.2}$$

These indirect utilities will play an important role in the subsequent analysis because they define agents' outside options in the D2C trading round. Hereafter, when no confusion arises, I will omit X for brevity and use a capital letter to denote any state-dependent function; for example, $M^{\mathrm{D2D}}(X)$ becomes M^{D2D}.

1.2.4 Trading Protocols

In the D2C trading round, agents i and j bargain over prices of all state-contingent claims written on X. As an outcome of this bargaining, dealer j quotes a kernel $M_{(\mathrm{i,j})}^{\mathrm{D2C}}$ that encodes the prices of all possible state-contingent claims. The quote is binding: The dealer commits to buy/sell contingent claims at the quoted prices in arbitrary quantities. Given such a kernel, customer i submits his or her demand schedule $G_{(\mathrm{i,j})}(M_{(\mathrm{i,j})}^{\mathrm{D2C}})$ to dealer j. Without loss of generality, I assume that $G_{(\mathrm{i,j})}$ satisfies $E[M_{(\mathrm{i,j})}^{\mathrm{D2C}}G_{(\mathrm{i,j})}] = 0$.[17]

Observing the kernel $M_{(\mathrm{i,j})}^{\mathrm{D2C}}$, customer i submits the optimal demand schedule $G^*_{(\mathrm{i,j})}(M_{(\mathrm{i,j})}^{\mathrm{D2C}})$ as the solution of

$$v_{(\mathrm{i,j})}[M_{(\mathrm{i,j})}^{\mathrm{D2C}}] \equiv \max_{G_{(\mathrm{i,j})}} \{E[U(F_i^{\mathrm{C}}+G_{(\mathrm{i,j})})] : E[M_{(\mathrm{i,j})}^{\mathrm{D2C}}G_{(\mathrm{i,j})}] = 0\}. \tag{1.3}$$

This in turn determines dealer j's indirect utility after optimally hedging the total exposure (i.e.,the D2C inventories and the endowments) in the D2D exchange:

$$v_{(\mathrm{j,i})}[M_{(\mathrm{i,j})}^{\mathrm{D2C}}] \equiv \max_{G_{(\mathrm{j,i})}} \{E[U(F_j^{\mathrm{D}}-G^*_{(\mathrm{i,j})}(M_{(\mathrm{i,j})}^{\mathrm{D2C}})+G_{(\mathrm{j,i})})] : E[M^{\mathrm{D2D}}G_{(\mathrm{j,i})}] = 0\}, \tag{1.4}$$

[17] Indeed, if an agent chooses a claim C and transfers $E[M_{(\mathrm{i,j})}^{\mathrm{D2C}}C]$ to the dealer, this is equivalent to buying $G = C - E[M_{(\mathrm{i,j})}^{\mathrm{D2C}}C]$ from the dealer, with $E[M_{(\mathrm{i,j})}^{\mathrm{D2C}}G] = 0$.

by choosing his optimal demand schedule $G^*_{(j,i)}(M^{\text{D2C}}_{(i,j)})$.

Given dealer j's bargaining power θ, the pair bargain and choose the pricing kernel that solves a version of the Nash bargaining problem, that is, maximizing the weighted surplus from trade, [18]

$$\max_{M^{\text{D2C}}_{(i,j)}} (1-\theta)\log(v_{(i,j)}[M^{\text{D2C}}_{(i,j)}] - \bar{v}_i) + \theta \log(v_{(j,i)}[M^{\text{D2C}}_{(i,j)}] - \bar{v}_j)\,, \tag{1.5}$$

subject to the no-arbitrage constraints

$$0 = E[M^{\text{D2C}}_{(i,j)} X] - E[M^{\text{D2D}} X]\,, \tag{1.6}$$

$$0 = E[M^{\text{D2C}}_{(i,j)}] - E[M^{\text{D2D}}]\,. \tag{1.7}$$

Note that when dealers' bargaining power $\theta \neq 0$, the participation constraints $v_{(i,j)} \geq \bar{v}_i$ and $v_{(j,i)} \geq \bar{v}_j$ never binds: Indeed, by the no-arbitrage condition, at any prices offered by the dealer, the customer can still decide to just trade the risky asset and the risk-free bond and, hence, reach his autarky utility. Formally,

Lemma 1. *In a fragmented equilibrium, customers' participation constraints never bind, while dealers' participation constraints bind only for those D2C exchanges that are competitive.*

1.2.5 Equilibrium

I denote by $\mathscr{E}(P, r, \{F_a\}_{a=i,j}, U, \theta)$ the primitives of the economy. A *fragmented equilibrium* of the economy $\mathscr{E}$ is a pricing kernel M^{D2D} and a continuum of D2C pricing kernels $M^{\text{D2C}}_{(i,j)}$, as well as a set of trading strategies $\{G^*_{(i,j)}, G^*_{(j,i)}, \beta^*_i, \bar{G}_{(j,i)}\}$ such that given $M^{\text{D2C}}_{(i,j)}$ and M^{D2D},

- $G^*_{(i,j)}$ maximizes customer i's utility in (1.3),
- $G^*_{(j,i)}$ maximizes dealer j's utility in (1.4) conditional on the customer's demand schedule $G^*_{(i,j)}(M^{\text{D2C}}_{(i,j)})$ as a function of the quoted pricing kernel $M^{\text{D2C}}_{(i,j)}$,
- β^*_i maximizes customer i's autarky utility in (1.1),
- $\bar{G}_{(j,i)}$ maximizes dealer j's autarky utility in (1.2),
- $M^{\text{D2C}}_{(i,j)}$ maximizes the Nash bargaining protocol (1.5) given constraints (1.6), (1.7),
- the D2D market clears, $0 = \int_{[0,1]^2} G^*_{(j,i)}\, di\, dj$.

[18] The trading protocol is standard in the literature on OTC markets [see e.g., Duffie et al., 2005, Malamud and Schrimpf, 2017]. Other trading protocol will deliver qualitatively similar results, for example, demand schedule game as in Kyle [1989]. In Appendix A.1, I use the uncertainty of getting a competitive execution in a two-stage trading game to micro-found the trading protocol.

1.3 Equilibrium Characterization

This section characterizes the fragmented equilibrium. I first establish a benchmark equilibrium that features a centralized exchange for all agents to trade contingent claims. Then I compare the fragmented equilibrium to the centralized competitive equilibrium.

1.3.1 Economy without Frictions: Centralized Exchange

Suppose there are no D2C exchanges and all dealers and customers can trade in a centralized exchange; then there exists a unique pricing kernel M that prices all contingent claims. Under this kernel, any agent, $a = i, j$, chooses a demand schedule G_a that solves

$$\max_{G_a} \{E[U(G_a + F_a)] : 0 = E[MG_a]\}.$$

The Lagrangian for this problem is

$$E[U(G_a + F_a)] - \lambda_a E[MG_a],$$

where λ_a is the Lagrange multiplier of the budget constraint. The first-order condition with respect to G_a yields

$$U'(G_a^* + F_a) = \lambda_a M.$$

For ease of representation, I denote the inverse of function $U'(\cdot)$ as $J(\cdot)$ and solve for G_a^*.

Lemma 2. *In a centralized exchange, the optimal demand schedule G_a^* for each agent $a = i, j$, satisfies*

$$G_a^*(M) = J(\lambda_a M) - F_a,$$

and the Lagrange multiplier is given by the budget constraint $E[MJ(\lambda_a M)] = E[MF_a]$.

It is then obvious that agent a's optimal consumption plan $J(\lambda_a M)$ depends only on the pricing kernel M and his own Lagrange multiplier. As options are in zero-net supply, all agents' demand schedules sum up to zero for any realized state of the world,

$$0 = \int_0^1 G_i^*(M)\,di + \int_0^1 G_j^*(M)\,dj.$$

Then according to assumption 2, I obtain the market clearing condition

$$X = \int_0^1 J(\lambda_i M)\,di + \int_0^1 J(\lambda_j M)\,dj,$$

after substituting in agents' optimal demand schedules.

Throughout the paper, I use the Black-Scholes formula as a benchmark to evaluate the effects of nonlinear risk imbalance and dealers' market power on option prices. To this end, I always use the following assumption in the comparative statics analysis as well as in simulations.

Assumption 3. *I assume all agents have the same CRRA utility function,*

$$U(X) = \frac{X^{1-\gamma}}{1-\gamma}.$$

The state of the world X is log-normally distributed with density $P(X) \sim \mathrm{lognormal}(\mu, \sigma^2)$.

I use $\mathbb{Q}$ to denote the risk neutral measure with the density $\mathrm{e}^r MP$; and $E^{\mathbb{Q}}$ to denote the corresponding expectation. I use $m_1^{\mathbb{P}} = E[\log X]$, $m_1^{\mathbb{Q}} = E^{\mathbb{Q}}[\log X]$ and

$$m_i^{\mathbb{P}} \equiv E[(\log X - m_1^{\mathbb{P}})^i], \; m_i^{\mathbb{Q}} \equiv E^{\mathbb{Q}}[(\log X - m_1^{\mathbb{Q}})^i], \; i > 1$$

to denote the moments of $\log X$ under the two measures. The following lemma is well known [see e.g., Rubinstein, 1976] and shows that, under log normality and CRRA preferences, option prices are given by the Black-Scholes formula.

Lemma 3. *Under assumption 3, a competitive equilibrium features a unique pricing kernel*

$$M = \mathrm{e}^{-r} \frac{X^{-1/\gamma}}{E[X^{-1/\gamma}]},$$

and all options are priced by the Black-Scholes formula. In particular, the implied volatility curve is flat across strikes, there is no variance risk premium as $m_2^{\mathbb{Q}} = m_2^{\mathbb{P}} = \sigma^2$, and there is no skewness risk premium as $m_3^{\mathbb{Q}} = m_3^{\mathbb{P}} = 0$.

1.3.2 Economy with Frictions

At the D2C trading round, the Lagrangian for customer i's optimization problem (1.3) is

$$E[U(G_{(\mathrm{i,j})} + F_i^{\mathrm{C}})] - \lambda_{(\mathrm{i,j})} E[M_{(\mathrm{i,j})}^{\mathrm{D2C}} G_{(\mathrm{i,j})}],$$

where $\lambda_{(\mathrm{i,j})}$ is the Lagrange multiplier for the budget constraint. This is the same problem as in Lemma 2, except the pricing kernel is now exchange-specific. Therefore, customer i's optimal demand schedule is $G^*_{(\mathrm{i,j})} = J(\lambda_{(\mathrm{i,j})} M_{(\mathrm{i,j})}^{\mathrm{D2C}}) - F_i^{\mathrm{C}}$, a function of the pricing kernel $M_{(\mathrm{i,j})}^{\mathrm{D2C}}$. In addition, the Lagrange multiplier is determined by the budget constraint

$$0 = E[M_{(\mathrm{i,j})}^{\mathrm{D2C}} G^*_{(\mathrm{i,j})}]. \tag{1.8}$$

After trading, customer i's optimal consumption plan is $J(\lambda_{(i,j)} M^{D2C}_{(i,j)}) = G^*_{(i,j)} + F^{C}_{i}$, which then determines the indirect utility,

$$v_{(i,j)}[M^{D2C}_{(i,j)}] \;=\; E[U(J(\lambda_{(i,j)} M^{D2C}_{(i,j)}))]. \tag{1.9}$$

At the D2D trading round, the Lagrangian for dealer j's optimization problem (1.4) is

$$E[U(G_{(j,i)} + F^{D}_{j} - G^*_{(i,j)}(M^{D2C}_{(i,j)}))] - \lambda_{(j,i)} E[M^{D2D} G_{(j,i)}],$$

where again $\lambda_{(j,i)}$ is the Lagrange multiplier. This is similar to Lemma 2, except that dealer j faces the D2D pricing kernel and his or her total exposure consists of two parts, the endowments and the inventories from D2C trading round. The optimal demand schedule is $G^*_{(j,i)} = J(\lambda_{(j,i)} M^{D2D}) - F^{D}_{j} + G^*_{(i,j)}(M^{D2C}_{(i,j)})$, where $\lambda_{(j,i)}$ is determined by

$$0 \;=\; E[M^{D2D} G^*_{(j,i)}]\,. \tag{1.10}$$

Given the optimal consumption plan $J(\lambda_{(j,i)} M^{D2D})$, I can write dealer j's indirect utility as

$$v_{(j,i)}[M^{D2C}_{(i,j)}] \;=\; E[U(J(\lambda_{(j,i)} M^{D2D}))]\,. \tag{1.11}$$

Proposition 1. *The fragmented equilibrium is unique and coincides with the centralized competitive equilibrium if all D2C exchanges are competitive; that is, the bargaining power* $\theta = 0$.

In a competitive D2C exchange, the dealer cannot charge any markup on the D2C pricing kernel and hence earns zero profit. Therefore, when all D2C exchanges are competitive, dealers essentially become agency brokers, and the allocation of risk is efficient. As long as the outside options are well defined, such equilibrium exists. From now on, the competitive equilibrium refers to either the centralized equilibrium or the fragmented equilibrium with $\theta = 0$.

Having established the benchmark, I now solve the generic D2C Nash bargaining problem (1.5). Its Lagrangian is

$$\begin{aligned}&(1-\theta)\log(v_{(i,j)}[M^{D2C}_{(i,j)}] - \bar{v}_i) + \theta \log(v_{(j,i)}[M^{D2C}_{(i,j)}] - \bar{v}_j)\\ &- \mu_{(i,j),s}(E[M^{D2C}_{(i,j)} X] - s) - \mu_{(i,j),r}\left(E[M^{D2C}_{(i,j)}] - e^{-r}\right).\end{aligned}$$

The second line consists of the no-arbitrage constraint for the risky asset (1.6) and that for the risk-free bond (1.7), where $\mu_{(i,j),s}$ and $\mu_{(i,j),r}$ are the corresponding Lagrange multipliers. I differentiate the Lagrangian function with respect to the D2C pricing kernel to get

$$0 \;=\; (v_{(i,j)} - \bar{v}_i)^{-1}(1-\theta)\frac{\delta v_{(i,j)}[M^{D2C}_{(i,j)}]}{\delta M^{D2C}_{(i,j)}} + (v_{(j,i)} - \bar{v}_j)^{-1}\theta \frac{\delta v_{(j,i)}[M^{D2C}_{(i,j)}]}{\delta M^{D2C}_{(i,j)}} - \mu_{(i,j),s} P X - \mu_{(i,j),r} P\,.$$

Then for $\theta \in (0,1]$, I define

$$\pi_{(i,j)} \equiv \frac{\lambda_{(i,j)}}{\lambda_{(j,i)}} \frac{1-\theta}{\theta} \frac{v_{(j,i)} - \bar{v}_j}{v_{(i,j)} - \bar{v}_i}.$$

The first-order condition can be rewritten as [19]

$$0 = \pi_{(i,j)} \frac{\delta v_{(i,j)}[M^{\text{D2C}}_{(i,j)}]}{\delta M^{\text{D2C}}_{(i,j)}} + \frac{\delta v_{(j,i)}[M^{\text{D2C}}_{(i,j)}]}{\delta M^{\text{D2C}}_{(i,j)}} - \mu_{(i,j),s} PX - \mu_{(i,j),r} P. \tag{1.12}$$

$\pi_{(i,j)}$ is an endogenous variable measuring the competitiveness of the D2C exchange ($1-\pi_{(i,j)}$ measures the dealer's market power). Indeed, when θ goes to zero, $\pi_{(i,j)}$ converges to one and the D2C exchange becomes fully competitive. On the other hand, when $\theta = 1$, $\pi_{(i,j)}$ becomes zero, and the D2C exchange becomes monopolistic. Formally,

Lemma 4. *The endogenous variable $\pi_{(i,j)}$ lies in the unit interval $[0,1]$ and corresponds one-to-one to the exchange-specific bargaining parameter $\theta_{(i,j)}$.*

In the above Lemma, I have relaxed the D2C Nash bargaining problem by introducing an exchange-specific bargaining power $\theta_{(i,j)}$. Due to the one-to-one mapping between $\theta_{(i,j)}$ and $\pi_{(i,j)}$, I can treat $\pi_{(i,j)}$ as exogenous and infer the bargaining parameter $\theta_{(i,j)}$, as well as other indirect utilities (1.1), (1.2), (1.9), and (1.11) after computing the equilibrium.

Given the relaxed and simplified D2C problem, I next compute the functional derivatives for the pricing kernel $M^{\text{D2C}}_{(i,j)}$ in Appendix A.3 and plug them into the relaxed first-order condition (1.12) to get

$$0 = (\kappa_{(i,j)} - \pi_{(i,j)})(J(\lambda_{(i,j)} M^{\text{D2C}}_{(i,j)}) - F^{\text{C}}_i) + \lambda_{(i,j)} J'(\lambda_{(i,j)} M^{\text{D2C}}_{(i,j)})(\kappa_{(i,j)} M^{\text{D2C}}_{(i,j)} - M^{\text{D2D}}) - \mu_{(i,j),s} X - \mu_{(i,j),r},$$

where $\kappa_{(i,j)}$ measures the price difference between the D2C and the D2D exchange; it is denoted as

$$\kappa_{(i,j)} \equiv \frac{E\left[J'(\lambda_{(i,j)} M^{\text{D2C}}_{(i,j)}) M^{\text{D2C}}_{(i,j)} M^{\text{D2D}}\right]}{E\left[J'(\lambda_{(i,j)} M^{\text{D2C}}_{(i,j)})(M^{\text{D2C}}_{(i,j)})^2\right]}. \tag{1.13}$$

Indeed, when the pricing kernel in the D2C exchange equals that in the D2D exchange, $\kappa_{(i,j)} = 1$.

Next, to have an explicit expression of the D2C pricing kernel in terms of the D2D pricing kernel and other endogenous parameters, I assume all the agents have log utility. Then, substituting D2C pricing kernels into the D2D market clearing condition,

$$0 = \iint_{[0,1]^2} G^*_{(j,i)}\, di\, dj; \tag{1.14}$$

I arrive at the following theorem.

[19] In appendix A.1, I show that this first-order condition endogenously arises in a two-stage D2C trading game.

Theorem 1. *Suppose all agents have $U(X) = \log(X + c)$ [20]. Suppose also that an equilibrium exists. Then, for each pair (i, j), the state-by-state D2C pricing kernel is given by*

$$M^{\text{D2C}}_{(i,j)}[M^{\text{D2D}}] = 2M^{\text{D2D}}\left(\pi_{(i,j)} + \sqrt{\pi^2_{(i,j)} - 4M^{\text{D2D}}\lambda_{(i,j)}Z_{(i,j)}}\right)^{-1},$$

where I have defined $Z_{(i,j)} \equiv -(F^{\text{C}}_i + c)(\kappa_{(i,j)} - \pi_{(i,j)}) - X\mu_{(i,j),s} - \mu_{(i,j),r}$.

The state-by-state D2D pricing kernel is characterized by a positive real root of a polynomial equation with order $2^{N_{I\times J}}$, where $N_{I\times J}$ is the number of different dealer-customer pairs (D2C exchanges).

Taking $\pi_{(i,j)}$ as given, all other parameters are determined by equations (1.8) (1.13), (1.6), (1.7), *and* (1.10), *respectively, for each D2C exchange.*

Once the relaxed system is solved , I can then determine the indirect utilities and bargaining powers for all D2C exchanges. Due partly to the nonlinearity of the system that characterizes the fragmented equilibrium, it is not trivial to provide a general condition such that the equilibrium exists. However, as long as the endowment function F^{C}_i is such that customer i' outside option (1.1) has an interior solution in the competitive equilibrium, then a unique fragmented equilibrium exists for a certain range of market competitiveness $\pi_{(\text{i,j})}$ and its local uniqueness is given by the implicit function theorem.[21]

1.4 Examples

In this section, I provide numerical examples [22] and use them to show how my model can generate empirically observed patterns in option prices and trading volume.

1.4.1 Primitives

Table 1.1 reports parameter specifications used throughout this section. The comparative statics are derived locally by asymptotic expansion. Specifically, I consider two cases: (i) when the D2C exchanges are 'almost' competitive and (ii) when the D2C exchanges are monopolistic, and the size of the nonlinear risks is 'small'. The local properties for both cases are qualitatively quite similar. Furthermore, the numerical example shows that the results of local comparative statics hold well globally.

[20] The parameter c is the subsistence of the agent and assures that the agents' outside options are well-defined and have an interior solution.

[21] My extensive numerical results suggest that equilibrium is in fact always unique. One can show that the equilibrium is indeed unique, if dealers are monopolist and either the risky asset or the risk free asset is centrally traded.

[22] For details of computation, see appendix A.3.

Table 1.1 – Primitives of the numerical example.

Variables	**Values**
Agents' utility	$U(X) = \log(X + c)$
Risky asset payoff	Lognormal(μ, σ^2)
Interest rate	1%
Supply of the risky asset	1
Supply of the risk free asset	0
Supply of nonlinear risk	0
Population of customer S	measure 0.5
Population of customer B	measure 0.5
Population of dealer	measure 1
Endowment of customer S	$F_S^C(0) = 0.6X + 0.8F_J$
Endowment of customer B	$F_B^C(0) = 0.8X - 2.0F_J$
Endowment of dealer j	$F_j^D(0) = 0.3X + 0.6F_J$
Market Power	$\theta_i(0) = 1$ for $i = S, B$

Risks The payoff of the risky asset follows a log-normal distribution with mean $\mu = 0.05$ and volatility $\sigma = 0.4$. Recall the results in Lemma 3: The Black-Scholes formula holds, and the variance and the skewness risk premia are zero in a centralized competitive environment. [23] I use this competitive equilibrium as the benchmark.

For ease of illustration, the nonlinear risk is specified as, [24]

$$F_J = \left(e^r s - X\right)^2, \; X \le e^r s.$$

Agents differ only in their respective loadings on this function. Clearly, this function is everywhere convex in the domain of the asset payoff $X \in (0, \infty)$. According to Carr and Madan (2002) [25], a continuous twice differentiable function can be replicated by a portfolio of a risky asset, a risk-free bond and a continuum of options. Formally,

Lemma 5. *For a continuous twice differentiable function $F(X)$ defined on $X \in [\underline{X}, \infty)$, the following representation holds,* [26]

$$\begin{aligned} F(X) &= F(\xi) + F'(\xi)(X - \xi) \\ &\quad + \int_{\underline{X}}^{\xi} F''(K)(K - X)^+ \, dK + \int_{\xi}^{\infty} F''(K)(X - K)^+ \, dK. \end{aligned}$$

[23] This result does not hold perfectly in my numerical example, as the utility function has a subsistence parameter $c \ge 0$. Nevertheless, the implied volatility curve in the competitive benchmark is 'almost' flat and close to σ.

[24] For $X < X^*$, I set $F_J = (e^r s + X^* - 2X)(e^r s - X^*)$ with $0 < X^* < e^r s$. This assures that the customers' outside options have interior solutions.

[25] Chapter 29 of "Volatility: New estimation techniques for pricing derivatives". Edited by R. Jarrow. The same result is also in Bakshi and Madan [2000].

[26] As a convention, I use $F'(\cdot)$ to represent the first-order derivative, $F''(\cdot)$ for the second-order derivative, and $(K - X)^+$ for the maximum between $K - X$ and 0.

where ξ is an arbitrary constant in the domain of X, $F'(\xi)$ is the number of shares held in the risky asset, $F''(K)$ is the number of options with strike K, and $F(\xi) - \xi F'(\xi)$ is the amount invested in the risk-free bond.

The choice of the cut-off ξ is arbitrary. In this section I set ξ equal to the future price of the risky asset, $e^r s$. Effectively, contingent claims in all exchanges are implemented by a continuum of out-of-the-money put and call options.

According to the lemma, F_J represents a long portfolio in options. Hence, option sellers will hold positive F_J and vice versa. Moreover, F_J is non-symmetric around the future price, $e^r s$, that is, more convex for the low state of X than for the high state. Hence, hedging demand for out-of-the-money (OTM) put options is higher than the demand for OTM call options. Formally, the skewness of F_J is defined as follows.

Definition 1. *From the dealers' perspective, any convex nonlinear risk F_J is said to be skewed to the left if*

$$E\left[\mathscr{M}[F_J]\left(\left(\log\frac{X}{s} - m_1^{\mathbb{Q}}\right)^3 - 3m_2^{\mathbb{Q}}\log\frac{X}{s}\right)\right] < 0.$$

with the linear operator $\mathscr{M}[\cdot]$ defined in 4, and $m_i^{\mathbb{Q}}$ such that $i = 1,2$ the risk-neutral moments for log returns.

This definition is closely linked to the third centered risk-neutral moments of log returns (see Proposition 3 below). In fact, it measures the first-order effect of the nonlinear risk F_J on the risk-neutral skewness. Intuitively, we may think that the risk-neutral skewness increases with customers' buying pressure on OTM call options (i.e., $F_J''(X) > 0$ for $X \geq e^r s$). This is mostly the case if the physical distribution P is log-normal. However, when $\log X$ follows a left-skewed distribution, customers' buying pressure on OTM call options for certain range of strikes may push down the risk-neutral skewness.

When P is log-normal, this definition covers broadly four trading activities. Specifically, when F_J is convex and skewed to the left, the dealers have the incentive to sell OTM put options; when F_J is convex and skewed to the right, the dealers have the incentive to sell OTM call options. It is also possible that F_J is concave and skewed to the left [27]; then the dealers have the incentive to buy OTM put options. Similarly, for F_J concave and skewed to the right, dealers have the incentive to buy OTM call options.

Agents There are two classes of customers, labeled B (buyers) and S (sellers), each class accounting for half of the customer population. I specify the endowments to ensure that customer S is the option seller and customer B is the option buyer. Specifically, customer S holds 0.8 units of F_J and 0.6 units of the risky asset, while customer B holds -2.0 units of F_J

[27] Here, the 'left' refers to the region that the second-order derivative of F_J is non-zero.

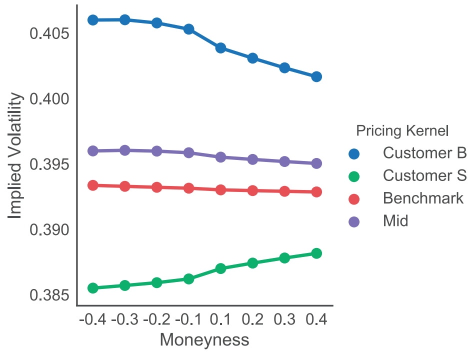


Figure 1.1 – Implied volatility for D2C exchanges.

Moneyness is defined as the log $\frac{K}{e^r s}$. This example uses parameters in Table 1.1 and does not include any shocks.

and 0.8 units of the risky asset. The difference in the holdings of the risky asset assures that both customers have a comparable size of wealth in the benchmark equilibrium. Formally, I denote customer's endowment without shocks as $F_i^{\mathrm{C}}(0)$ with $i = B, S$.

Dealers are homogeneous. According to assumption 2 (i.e., the aggregate nonlinear risks are zero), dealers' total nonlinear endowments are given by $X - \sum_{i=S,B} F_i^{\mathrm{C}}$. Each dealer therefore starts with endowment $F_j^{\mathrm{D}}(0) = 0.3X + 0.6F_J$. This endowment in effect makes dealers option sellers. In addition, F_J measures the nonlinear risk imbalance between dealers and customers.

To show the effects of nonlinear risks and market power, all dealers are initially monopolists in their respective D2C exchanges. Given the parametrization, I then solve the model numerically.

Figure 1.1 shows the implied volatility [28] computed using four different pricing kernels, namely, the two D2C pricing kernels, the 'mid' pricing kernel and the centralized benchmark pricing kernel. The 'mid' pricing kernel is defined below.

Definition 2. *The 'mid' pricing kernel is the wealth-weighted average pricing kernel among all D2C exchanges,* $\bar{M}^{\mathrm{D2C}} = \iint_{[0,1]^2} \lambda_{(i,j)}^{-1} M_{(i,j)}^{\mathrm{D2C}}\, di\,dj / \iint_{[0,1]^2} \lambda_{(i,j)}^{-1}\, di\,dj$.

In my model, customers' demand for options is proportional to their wealth, $\lambda_{(i,j)}^{-1}$. Empirically, we can think of the 'mid' pricing kernel as the volume-weighted average transaction (or quoted) price.

Clearly, the implied volatility (hereafter, IV) for customer B is the highest among all the four IVs. Meanwhile, the IV for customer S is the lowest. Hence, I refer to the price paid by customer B

[28] To generate a more pronounced implied volatility skew, I would need to assume that the endowment risks satisfy one of the following conditions: (i) customers on average buying out-of-the-money put options, and selling out-of-the-money call options; (ii) the endowment risks are reasonably large, and tilt towards down-side risks (in the numerical example, the endowment risk is flat in the sense that customers buy the same amount of options across strikes for put options); (iii) the dealer to customer pricing kernel has a corner solution; (iv) either the risky asset or the risk free asset is centrally traded, but not both.

as the ask, while the price paid by customer S as the bid. Note that the IV for the 'mid' is above the benchmark IV. Not surprisingly, as the net option demand from customers is positive (i.e., more buy orders than sell orders), dealers raise the 'mid' price to charge a high markup for customer B.

Another interesting aspect to note is that the IV for the bid price is below the benchmark IV, suggesting a negative variance risk premium. This contradicts the findings in Carr and Wu [2009], who report that the variance risk premium for SPX options measured from bid prices is also positive. One possible explanation is that the physical distribution is not log-normal in reality, or agents have different risk aversion. Both channels are shut down here in the example.

To measure the overall effects of nonlinear risk imbalance on option prices, I consider two option premia, namely, the variance risk premium and the skewness risk premium. In particular, the variance (skewness) risk premium is defined as the difference between the risk-neutral variance (skewness) and the physical variance (skewness) of the risky asset return (i.e., of $\log(X/s)$):

$$\mathrm{RP}_i \equiv m_i^{\mathbb{Q}} - m_i^{\mathbb{P}}, \; i = 2, 3 .$$

I compute both premia using the 'mid' pricing kernel. Intuitively, the 'mid' pricing kernel measures the total compensation (markup plus the risk premium) customers pay to dealers for bearing the endogenous nonlinear risk.

Proposition 2 (Variance Risk Premium). *Suppose the nonlinear risk imbalance F_J is convex (concave) in the domain of the random payoff X, then the variance risk premium computed from the 'mid' pricing kernel, $\bar{M}^{\text{D2C}}$, is larger (smaller) than the premium computed from the benchmark pricing kernel $M^{(0)}$.*

Intuitively, by Lemma 5, the variance risk premium can be replicated by long positions in a portfolio of puts and calls. In the example, the 'mid' IV is uniformly above the benchmark IV, suggesting a positive markup charged by dealers in response to customers' net buying pressure. Hence, the variance risk premium becomes positive.

Regarding the skewness risk premium, first note that the IV for the 'mid' price is skewed to the left, suggesting a negative risk-neutral skewness. Meanwhile, the physical skewness is 0 for log-normal distribution. Hence, the skewness risk premium is negative. This is the result of customers' excessive demand on OTM put options rather than call options. If the physical distribution is log-normally distributed, the skewness of the nonlinear risk imbalance F_J determines the direction of the skewness risk premium. However, more generally, unlike the variance risk premium, the skewness risk premium depends on both the shape of F_J and the physical distribution.

Proposition 3 (Skewness Risk Premium). *Suppose that P is log-normally distributed, then the* *skewness risk premium*

- *is positive if the nonlinear risk imbalance F_J is convex and right-skewed;*
- *is negative if the nonlinear risk imbalance F_J is convex and left-skewed.*

Intuitively, if F_J is a long skewness exposure (i.e., short OTM call options), then to hedge their short skew risk the customers need to buy a portfolio of options that resembles F_J. This demand allows dealers to charge a premium on the skew, leading to a negative skewness risk premium.

1.4.2 Macro Shocks

I consider three 'macro' shocks, the imbalance shock (ϵ^{IMB}), the market power shock (ϵ^{MP}), and the wealth shock (ϵ^{W}). They are called macro shocks precisely because of their effects on a sub population of agents rather than atom-less individual. For each of the shocks, I consider three levels, labeled *Small*, *Medium* and *Large*.

Imbalance Shock This shock captures the distribution of nonlinear risks between customers and dealers. It is a shock on the size of the nonlinear risk among dealers. For simplicity, the shock affects dealers' endowments uniformly,

$$F_j^{\text{D}}(\epsilon^{\text{IMB}}) \;=\; F_j^{\text{D}}(0) + \epsilon^{\text{IMB}} F_J \,,\; \epsilon^{\text{IMB}} \in \{0, -0.2, -0.4\} \,.$$

Here, the imbalance shock reduces the dealers' long position in the nonlinear risk F_J, making them sell fewer options. Specifically, for a small shock, dealers and customer S hold the same amount of long positions in F_J (sellers); for a medium shock, dealers do not hold any F_J (neutral), and for a large shock, dealers hold negative positions in F_J (buyers). Consistent with Assumption 2, customers are assumed to hold an off-setting position in F_J. The off-setting shock affects customers uniformly (see Table 1.1), so that their respective endowments become

$$F_i^{\text{C}}(\epsilon^{\text{IMB}}) \;=\; F_i^{\text{C}}(0) - \epsilon^{\text{IMB}} F_J \,,\; i = B, S.$$

Notably, both customers receive additional long option positions after the shock. Therefore, customer B wants to buy fewer options, while customer S wants to sell more options.

Market Power Shock The market power shock is uniformly distributed among the D2C exchanges,

$$\theta_i(\epsilon^{\text{MP}}) \;=\; \theta_i(0) + \epsilon^{\text{MP}} \,,\; \epsilon^{\text{MP}} \in \{0, -0.5, -1\} \,.$$

When the shock is 0, dealers have full market power and can charge the highest markups in D2C exchanges. When the shock is -1, all D2C exchanges become competitive, and the option prices and trading patterns coincide with the competitive benchmark.

Wealth Shock The wealth shock also affects all dealers uniformly,

$$F_j^{\mathrm{D}}(\epsilon^{\mathrm{W}}) = F_j^{\mathrm{D}}(0) + \epsilon^{\mathrm{W}} X, \; \epsilon^{\mathrm{W}} \in \{0.0, -0.2, -0.4\}.$$

Recall the total supply of the risky asset is normalized to one; a unit increase in the dealers' wealth thus implies a unit reduction in the customers' wealth. The wealth shock also affects customers' endowment uniformly,

$$F_i^{\mathrm{C}}(\epsilon^{\mathrm{W}}) = F_i^{\mathrm{C}}(0) - \epsilon^{\mathrm{W}} X, \; i = B, S.$$

As the risky asset is centrally traded, the number of shares held does not affect directly the option trading. However, indirectly, due to wealth effect, for dealers, a negative wealth shock effectively reduces their risk aversion and, hence, their risk bearing capacity.

1.4.3 Option Premia

Now we look at the correlation between customers' option demand and option risk premia. Figure 1.2, column one, shows that with a reduction in the size of the nonlinear risk imbalance F_J, customers on average buy fewer options from dealers. Consequently, the 'mid' option prices become cheaper (see Figure 1.3, column one). Meanwhile, the shock also reduces the inventory in the D2D exchange, hence alleviating the distortion on the D2D prices.

Proposition 4 (Imbalance Shock)**.** *Customers' net buy of options is positively (negatively) correlated with the variance risk premium measured from the 'mid' (D2D) pricing kernel if dealers experience an imbalance shock.*

This result relates directly to the findings in Gârleanu et al. [2009]. The authors show both theoretically and empirically that customers' net buy of options is positively correlated with the variance risk premium, primarily due to the premium paid to dealers for bearing the non-hedgeable risks (e.g., jumps). Here, the economic reasoning is different: Dealers are able to hedge perfectly; however, due to the market fragmentation, each dealer charges a markup to customers, resulting in endogenous inventories to be shared in the D2D exchange. When the risk imbalance is reduced, customers buy fewer options from dealers and, consequently, the prices they pay become cheaper. On the other hand, due to the reduction in the aggregate inventory in the D2D exchange, the D2D prices become less distorted and hence increase towards the centralized benchmark. Empirically, we can think of the imbalance shock as demand shocks.

Next, Figure 1.2, column two, shows that, after the reduction of dealers' market power, cus-

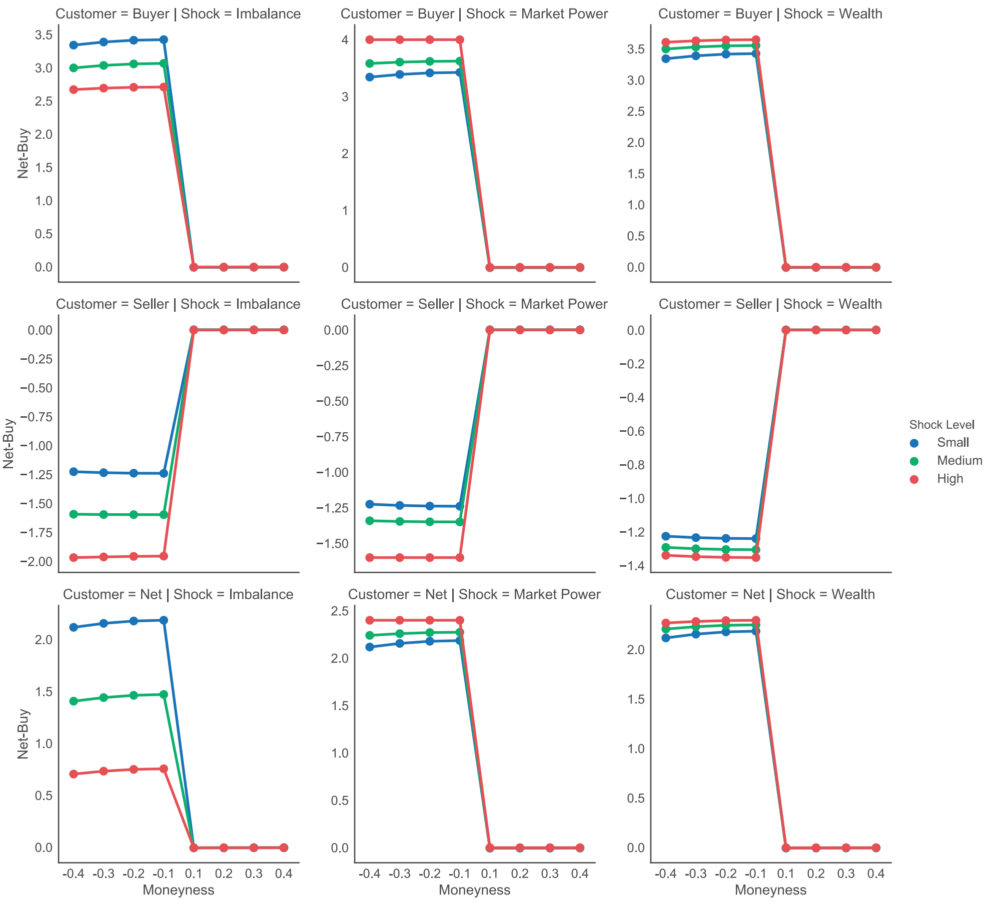


Figure 1.2 – Effects of 'macro' shocks on option demand.

tomers buy more options. At the same time, the variance risk premium becomes smaller (see Figure 1.3, column one). This result is in stark contrast to the imbalance shock.

Proposition 5 (Market Power Shock)**.** *Customers' net buy of options is negatively (positively) correlated with the variance risk premium measured from the 'mid' (D2D) pricing kernel if dealers experience a market power shock.*

The intuition is as follows. When D2C exchanges become more competitive, customers are able to trade at more favorable prices, resulting in better risk sharing. This in turn helps to reduce the size of the inventories in the D2D exchange. Hence, the price distortions on both the D2D exchange and the D2C exchanges are reduced.

Figure 1.2, column three, shows that customers buy more options from dealers after the decrease in dealers' wealth. Meanwhile, the option prices on average become cheaper for customers to trade (see Figure 1.3, column one). This is consistent with the intuition that dealers become more risk-averse after the negative wealth shock. Hence, they provide price concessions to customers to off-load inventories.

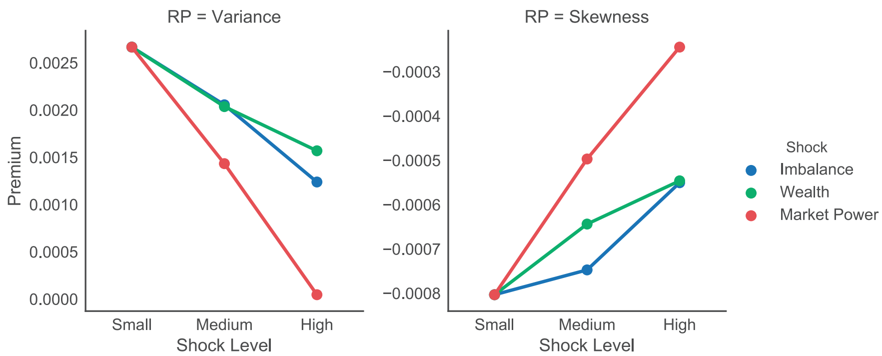


Figure 1.3 – Effects of 'macro' shocks on option risk premium.

Proposition 6 (Wealth Shock)**.** *Customers' net buy of options is negatively (negatively) correlated with the variance risk premium measured from the 'mid' (D2D) pricing kernel if dealers a experience wealth shock.*

Interestingly enough, contrary to the conventional wisdom, the impact of a wealth shock or market power shock is very different from that of an imbalance shock (see Proposition 4). Indeed, while an imbalance shock mostly leads to a positive correlation between customers' price pressure and option prices, this is not the case for the wealth shock. Specifically, the wealth shock always induces a negative correlation between customers' total net buy of options and the variance (skewness) risk premium at the 'mid' price, consistent with the findings in Chen et al. (forthcoming). The underlying mechanism is based on the dealers' effective risk aversion: When dealers' net worth drops, their risk aversion rises, increasing their incentive to smooth consumption. In this case, dealers find it optimal to give large price concessions to customers, forcing the latter to take more risk off dealers' balance sheets.

Furthermore, there is also a fundamental difference between the wealth shock and the market power shock: Although customers can trade more at more favorable prices under both shocks, the prices on the D2D exchange change differently. A market power shock allows for better risk-sharing; hence, the inventory reduction is an efficient outcome on the D2D exchange. On the other hand, the wealth shock reduces dealers' risk bearing capacity, forcing them to provide price concessions to customers and also increasing the price they require to bear risks in the D2D exchange.

Notably, Figure 1.3 also shows that, for all the macro shocks, the variance risk premium is always negatively correlated with the skewness risk premium. The next proposition shows that in fact the two risk premia are closely linked through the nonlinear risk imbalance F_J.

Proposition 7. *When any of the three macro shocks hits and P is log-normally distributed, the correlation between the variance risk premium and the skewness risk premium*

- *is negative if dealers' aggregate nonlinear risk F_J is left-skewed;*

- *is positive if dealers' aggregate nonlinear risk F_J is right-skewed.*

Note that F_J can be either concave or convex.

Four possible trading activities are covered in Proposition 7, customers buy (sell) OTM put (call) options, and buy (sell) OTM call (put) options. However, regardless of whether dealers hold long or short options, the sign of the correlation between the two option risk premia is always determined by whether the trading activities are concentrated on the OTM calls or puts.

Since the physical distribution is fixed, the negative correlation between the skewness and the variance risk premia immediately implies that the correlation between the risk-neutral skewness and the risk-neutral variance is also negative. This prediction is consistent with the empirical evidence in Constantinides and Lian [2015]. In particular, the authors find that in SPX options markets, customers usually long OTM puts and sell OTM calls, causing a decrease in the risk-neutral skewness (i.e., implied volatility skews to the left). Meanwhile, the number of puts being bought exceeds the number of calls being sold, resulting in an increase in risk-neutral variance. Both price effects arise in my model due to dealers' market power and nonlinear risk imbalance.

1.4.4 Cost of Trading

To begin with, I define the effective spreads as follows.

Definition 3. *The cost of trading is defined as the effective percentage bid-ask spreads,*

$$\frac{\left|E\left[(M^{\mathrm{D2C}}_{(i,j)} - \bar{M}^{\mathrm{D2C}})O(K)\right]\right|}{E\left[\bar{M}^{\mathrm{D2C}}O(K)\right]}.$$

Note that $O(K)$ denotes call/put option payoff with strike price K.

Figure 1.4 and 1.5 show that the effective spreads are higher for OTM options than for at-the-money (ATM) and in-the-money (ITM) options for both calls and puts. This pattern is consistent with the findings in George and Longstaff [1993] and Cho and Engle [1999]. In my model, as the spreads for each strike are normalized by their respective 'mid' prices, the spreads for OTM options become large. In addition, when the nonlinear risk is 'small', dealers' optimal quoting strategy tends to have a larger impact on the tails of the pricing kernel. This effect results in higher spreads for OTM options than ATM options.

Figure 1.2, column two, shows that both customers' option demand decreases with dealers' market power. Consequently, the aggregate trading volume in the D2C segment decreases. Meanwhile, Figure 1.4 and 1.5, column two, show that the effective spreads for both customers increase. Intuitively, the market power allows dealers to charge a higher markup on the D2C

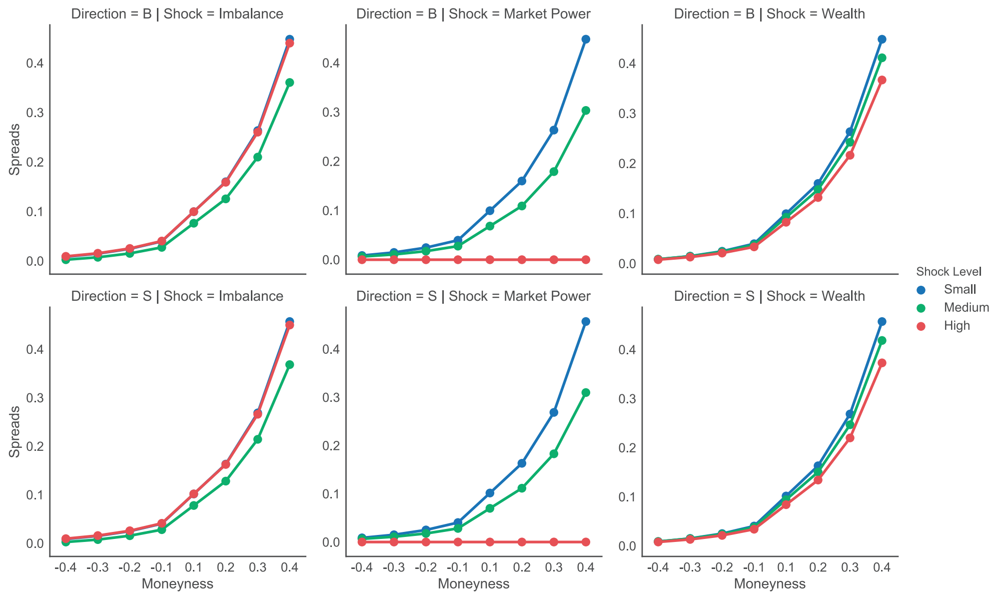


Figure 1.4 – Effects of ‘macro’ shocks on effective percentage bid-ask spreads for call options.

exchanges, and customers respond by trading fewer options. This market power effect limits the risk sharing between dealers and customers.

Figure 1.2, column three, shows that when increasing dealers’ wealth, customer S sells more options while customer B buys fewer options. Interestingly, the aggregate trading volume in the D2C segment decreases. Meanwhile, Figure 1.4 and 1.5, column three, show that the effective spreads increase with dealers’ wealth for both customers. However, the spreads for option sellers increase more than the spreads for option buyers. This is not surprising because dealers are also option sellers, hence pushing ‘mid’ prices further away from bid prices.

I now summarize the results formally in the next proposition.

Proposition 8. *The aggregate trading volume in the D2C segment decreases with dealers’ market power or wealth, while the effective spreads*

- *increase with the dealers’ market power;*
- *increase with the dealers’ wealth for customers trading in the same direction of the dealers, and can increase or decrease for customers trading in the opposite direction of the dealers.*

However, the average of the effective spreads across all D2C exchanges increases with dealers’ market power or wealth.

The results of Proposition 8 are consistent with the existing empirical evidence. For example, De Fontnouvelle et al. [2003] find that options bid-ask spreads decreased after the introduction

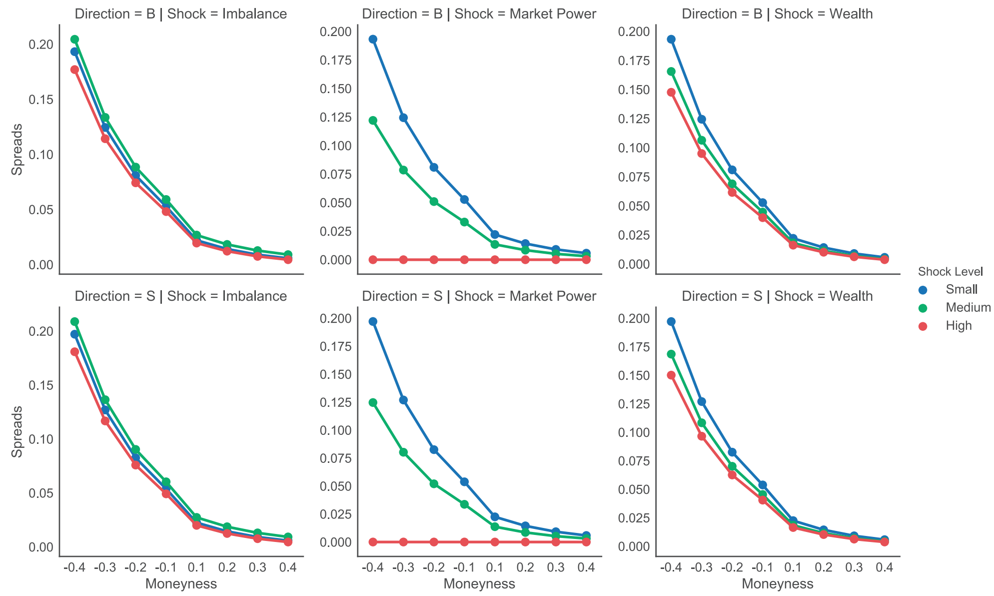


Figure 1.5 – Effects of 'macro' shocks on effective percentage bid-ask spreads for put options.

of multi-listed options, likely due to improved competition. Similarly, in a recent study, Christoffersen et al. show that in recent years (2004 to 2012), the market-wide option bid-ask spreads decreased while the trading volume increased, consistent with the reduction of dealers' market power.

1.5 Empirics

In this section, for 34 liquid ETF options, I show empirically that the customers' net buy of options is sometimes negatively correlated with the variance risk premium. This result confirms the result for SPX options documented in Chen et al. (forthcoming) and Constantinides and Lian [2015] but is in contrast to the evidence in the literature on demand based option pricing [Gârleanu et al., 2009, Bollen and Whaley, 2004, Fournier and Jacobs, 2016]. My model provides a rational explanation for such result based on the dealers' wealth effect.

Second, I show empirically that for the same sample of options the correlation between the risk-neutral variance and the risk-neutral skewness is often positive despite the negative correlation between the realized variance and realized skewness. This result is puzzling, as in an equilibrium model in which only the fundamental risk is priced, such a result will not arise. For example, models with disaster risks predict that the risk-neutral variance increases with the intensity of the disaster risk, while the risk-neutral skewness decreases. This prediction emerges because the marginal investor requires extra compensation for bearing the disaster risk. This intensity shock also raises the physical variance and decreases the physical skewness. Hence, the correlation of the two physical moments and the correlation of the two risk-neutral moments should go hand in hand.

Third, I use the result in Proposition 7 to build a measure for the shape of customers' net option demand and show that this measure can explain the cross-section variation in correlations between the risk-neutral variance and skewness. In addition, this measure can also explain a small amount of the time series variation in my ETF panel.

1.5.1 Data Description

I use four database in my empirical study: Open/Close (CBOE, ISE and NASDAQ PHLX), OPRA, OptionMetrics, commodity options TAQ. The OPRA and the commodity options TAQ data are provided by Nanex.

The Open/Close data allow me to construct order-flow imbalance measures and have been used in several empirical studies, including Pan and Poteshman [2006a], Gârleanu et al. [2009], Chen et al. (forthcoming) and Fournier and Jacobs [2016]. For each option contract (ticker, put or call, strike, maturity), the data report separately the trading volume for several trader types [29]: Market-Maker, Firm (proprietary firms and broker/dealers), Customer (small, medium, large), Professional Customer (small, medium, large). Furthermore, the trading volume is separated into four types: Open Buy, Close Buy, Open Sell and Close Sell. In particular, Open Buy means the trader has bought the contract to open a new long option position, while Close Buy means the trader has bought the contract to close an existing short position.

The OPRA data run from January 2010 through December 2015. This allows me to observe trades and quotes for index, equity and ETF options traded in all the US options exchanges. I use the trades and the corresponding quotes data from this database to construct the bid-ask spreads measure as well as the order-flow imbalance measure using the Lee and Ready (1991) algorithm.

The OptionMetrics data provide option Greeks and prices for index and ETF options.

To test my model, I select a sample [30] of actively traded index and ETF options, as well as commodity options. These are options based on macro risks and, hence, are subject less to the concern regarding asymmetric information.[31]

[29] NASDAQ PHLX directly reports the buy and sell volume for market-makers. For ISE and CBOE, the market-maker's position can be deduced from the difference between volume of the other traders and the volume of the exchange.

[30] For a full list of option tickers, see Appendix A.4. The selection criterion is based on the ranking of daily average trading volume for ETF options on ISE.

[31] I expect that these market power effects are stronger for over-the-counter order-flows [see, for example, Harald et al., 2017]. Using my model to understand pricing and trading of OTC derivatives is an interesting direction for future research.

1.5.2 Variable Definitions

Moneyness Bins In practice, multiple options across strikes and maturities are listed for one underlying asset. To compare prices and trading activities over time, I group options into bins according to their moneyness and maturities. Assuming zero interest rates, I use the Black-Scholes delta to proxy option moneyness of a European call,

$$\Delta(C,K,T) = \Phi\left[\frac{\log(K/S)+0.5\sigma^2 T}{\sigma\sqrt{T}}\right],$$

in which $\Phi(\cdot)$ is the standard Normal cumulative distribution function, σ is the realized volatility of the underlying asset over the most recent 60 trading days, K is the strike price, T is the time-to-maturity, and S is the underlying price. For a European put, I take the $1+\Delta(P,K,T)$ as its moneyness. Hence, OTM put options have the same moneyness as OTM call options. I then group options into five moneyness bins (Table 1.2) as in Bollen and Whaley [2004]. Formally, I denote the moneyness bin as $\mathscr{B} = \mathrm{OTM}, \mathrm{DOTM}, \mathrm{ATM}, \mathrm{ITM}, \mathrm{DITM}$.

Table 1.2 – Moneyness Bins Definitions

Bins	**Range**
Deep in-the-money (DITM)	[0.875,1.000]
In-the-money (ITM)	[0.625,0.875]
At-the-money (ATM)	[0.375,0.625]
Out-of-the-money (OTM)	[0.125,0.375]
Deep out-of-the-money (DOTM)	[0.000,0.125]

Variance Risk Premium The variance risk premium is defined as the ratio between the risk-neutral variance and the physical variance. I proxy the physical variance by the realized variance computed from a 30-day rolling-window. For the risk-neutral variance, I use the model-free formula in Bakshi et al. [2003]. [32]

$$\mathrm{VRP}_t = \mathrm{Variance}_t^{\mathbb{Q}} - \mathrm{Variance}_{t,t+30}^{\mathbb{P}}.$$

Skewness Risk Premium Similarly, the skewness risk premium is the ratio between the risk-neutral skewness and the physical skewness. The physical skewness is estimated based on the formulas in Neuberger [2012]. The risk-neutral skewness is again from the formula in Bakshi et al. [2003]. Then, the skewness risk premium is

$$\mathrm{SRP}_t = \mathrm{Skew}_t^{\mathbb{Q}} - \mathrm{Skew}_{t,t+30}^{\mathbb{P}}.$$

[32] For the details of the formula, refer to the appendix.

Order-Flow Imbalance Use the Open/Close data (CBOE, ISE and NASDAQ), I compute the customers' aggregate net buy of options as [33]

$$\mathrm{IMB}_t(\mathscr{B}, i) \;=\; \sum_{K\in\mathscr{B}} \mathrm{OB}_t(i,K,T) + \mathrm{CB}_t(i,K,T) - \mathrm{OS}_t(i,K,T) - \mathrm{CS}_t(i,K,T)\,,\; i = C, P\,,$$

in which OB (OS) stands for open buy (sell) orders, and CB (CS) stands for close buy (sell) orders. I also construct the order-flow imbalance measure based on options TAQ data.

$$\mathrm{IMB}_t(\mathscr{B}, i) \;=\; \sum_{\tau\in[t-h,t)} \sum_{K\in\mathscr{B}} \mathrm{OF}^{\text{BUY}}(\tau,i,K,T) - \mathrm{OF}^{\text{SELL}}(\tau,i,K,T)\,.$$

The sign of the order-flow OF is determined using the Lee and Ready (1991) algorithm. Specifically, the order is defined as an aggressive buy if it is executed above the 'mid' quote and vice versa.

Demand Pressure After calculating the order-flow imbalance, I can define the demand pressure measure. The first demand pressure is on the level of the option prices,

$$\mathrm{IMB}_t^{\text{LEVEL}} \;=\; \sum_{\mathscr{B}} \sum_{i=C,P} \mathrm{IMB}_t(\mathscr{B}, i)\,.$$

If the measure is positive, customers on average buy options from dealers. The next measure is the demand pressure on the skewness of the option prices,

$$\mathrm{IMB}_t^{\text{SKEW}} \;=\; \mathrm{IMB}_t(\mathrm{OTM}, C) - \mathrm{IMB}_t(\mathrm{OTM}, P)\,.$$

Intuitively, customers' net buy of OTM call options or net sell of OTM put options drives up the skew. [34]

Shape of Imbalance Shock To capture the shape of the imbalance shock, I use the following measure,

$$\mathrm{IMB}_t^{\text{SHAPE}} \;=\; \frac{\left|\sum_{\mathscr{B}=\mathrm{OTM,ATM}} \mathrm{IMB}_t(\mathscr{B}, C)\right| - \left|\sum_{\mathscr{B}=\mathrm{OTM,ATM}} \mathrm{IMB}_t(\mathscr{B}, P)\right|}{\left|\sum_{\mathscr{B}=\mathrm{OTM,ATM}} \mathrm{IMB}_t(\mathscr{B}, C)\right| + \left|\sum_{\mathscr{B}=\mathrm{OTM,ATM}} \mathrm{IMB}_t(\mathscr{B}, P)\right|}\,.$$

Motivated by the model, the larger the shape measure, the more positive the correlation between the risk-neutral variance and the risk-neutral skewness.

[33] Filters employed: i) remove expired options; ii) remove day of trade/expiration pairs not found in OptionMetrics database; iii) remove day of trade and option strike not found in OptionMetrics database; iv) remove options on the expiration day.

[34] To be more precise, if the physical density is highly left-skewed, the call buying pressure needs to be reasonably high in order to move the skew to the right.

1.5.3 Demand Pressure?

The first test is to show that, customers' net option demand may drive down option prices instead of driving them up. Formally, I run the following regression for each ETF in my sample,

$$\text{VRP}_t \;=\; \beta_0 + \beta_1 \text{IMB}_t^{\text{LEVEL}} + \varepsilon_t\,.$$

Table 1.3 shows that for XLI, SPY, EWJ and XOP, the correlation between the customers' option net demand and the VRP is negative. Hence, for certain ETF options, the demand pressure theory is inconsistent with the empirically observed patterns.

In addition to the test on the variance risk premium, I run another test,

$$\text{Variance}_t^{\mathbb{Q}} \;=\; \beta_0 + \beta_1 \text{IMB}_t^{\text{LEVEL}} + \gamma_1 \text{Variance}_{t,t+30}^{\mathbb{P}} + \varepsilon_t\,.$$

Table 1.4 shows that, indeed, after controlling for the physical variance, the daily variation in the risk-neutral variance is often negatively associated with the contemporaneous customers' option demand.

For the demand pressure on the skewness risk premium, I run the following regression and control for the realized skewness,

$$\text{Skew}_t^{\mathbb{Q}} \;=\; \beta_0 + \beta_1 \text{IMB}_t^{\text{SKEW}} + \gamma_1 \text{Skew}_{t,t+30}^{\mathbb{P}} + \varepsilon_t\,.$$

Table 1.5 shows that, compared to the risk-neutral variance, risk-neutral skewness is much harder to explain. Indeed, even after including the controls, the adjusted R^2 is not very large. Importantly, we note that if the demand pressure theory holds, then the coefficient β_1 should be significantly positive. Clearly, for some of the options (two out of seven), this is not the case.

In light of the demand pressure puzzle, Chen et al. (forthcoming) propose that the dealers' limited risk bearing capacity may be the cause of the negative correlation. In particular, they consider an environment with negative jump risks in the asset returns. Dealers' risk aversion rises with the intensity of the disaster risk, inducing them to offer less risk-sharing to customers at higher prices. However, in their model, the correlation between the risk-neutral variance and the skewness is closely linked to the disaster risk. Precisely, when the intensity of the disaster rises, the physical variance increases and the physical skewness decreases. In turn, dealers require higher risk premium for bearing risks; therefore, the risk-neutral variance increases and the risk-neutral skewness decreases. This suggests that the correlation in the physical variance and skewness should go hand in hand with the correlation in the risk-neutral variance and skewness.

1.5.4 Correlation Puzzle

Table 1.6 shows that, the correlation between the realized skewness and the realized variance is negative and statistically significant for most of the ETF options except for the long-term bond (TLT), the commodity ETFs (UNG: natural gas; GDX: gold miner; USO: crude oil), and the US dollar (UUP). This is broadly consistent with the fact that equity ETFs are subject to negative jumps that occur simultaneously with high volatility.

In contrast, the correlation between the risk-neutral skewness and the risk-neutral variance paints a rather different picture: The correlation is mostly positive and statistically significant (19 out of 34), suggesting that the state with a high level of option prices is associated with expensive OTM call options. Hence, together with the negative correlation between the realized variance and skewness, the data suggest the short-term variation in the correlation between the variance and skewness premia cannot be purely driven by fundamentals.

My model provides an explanation for this puzzle. The main intuition is that customers' nonlinear risk endowments may not be linearly aligned with the physical states of the world. Specifically, some customers may want to buy OTM put options due to receiving nonlinear shocks that resemble short OTM put positions, without any actual changes in the intensity of disaster risk. Hence, empirically, we can look at the particular shape of customers' option demand. For example, if customers demand pressure (in absolute terms) is concentrated on OTM calls rather than OTM puts, then we are likely to observe a positive correlation between the variance and skewness risk premia. In addition, if the physical distribution has not moved, then the correlation between the risk-neutral variance and skewness will also be positive. Having said that, does the shape of the customers' option demand actually affect the correlation between the risk-neutral variance and the skewness?

Based on the empirical measure for the shape of customers' net demand, $\text{IMB}_t^{\text{SHAPE}}$, I proceed as follows. First, for each ETF option, I divide the time series into quintiles based on the value of the shape measure. In particular, the fifth quintile corresponds to the largest excessive call trading activities by customers. Within each quintile, I compute the following correlations: [35] the correlation between the risk-neutral variance and skewness,

$$\text{Corr}\left[\text{Variance}_t^{\mathbb{Q}}, \text{Skew}_t^{\mathbb{Q}}\right];$$

the correlation between the realized variance and skewness,

$$\text{Corr}\left[\text{Variance}_{t,t+30}^{\mathbb{P}}, \text{Skew}_{t,t+30}^{\mathbb{P}}\right];$$

the correlation between the two risk premia,

$$\text{Corr}\left[\text{SRP}_t, \text{VRP}_t\right].$$

[35] The correlation is computed based on daily observations.

According to the prediction of my model, the correlation between the risk-neutral variance and skewness should increase with the shape parameter. My model does not restrict the correlation between the realized variance and skewness; the correlation between the variance risk premium and the skewness risk premium decreases in the shape measure.

Table 1.7 shows that for certain ETF options the results seem to align with my model's prediction. In particular, for equity sector ETFs (XLV, XLU, IYR, XLF), for index ETFs (SPY), for international equity ETFs (ASHR, EWJ, EFA, EWZ, FXI), for commodity ETFs (UNG, OIH, GDX, GLD, USO, XOP), and for currency options (FXE, UUP), there appears to be an uptrend while increasing the shape measure.

Table 1.8 shows no particular relationship between the shape measure and the correlation between the realized variance and skewness.

Table 1.9 shows that, except for index ETF options, most of other ETF options do not have a strong correlation between the variance risk premium and the skewness risk premium. This is likely because various shocks may work in the opposite direction, or the correlation varies dramatically over time, leading to insignificant whole sample correlation. The index options also may differ from other option categories in terms of the underlying risk dynamics. I leave this question for future research.

Cross-Section Admittedly, a correlation measure requires a large volume of data. To circumvent this problem, I explore the cross-sectional properties of my data. Specifically, I compute the correlation for the risk-neutral variance and risk-neutral skewness for each of the ETF options in my sample. Then I test whether the shape measure of customers' option demand can capture the variation across ETF options. Specifically, I run the following univariate regression,

$$\text{Corr}_o\left[\text{Variance}_t^{\mathbb{Q}}, \text{Skew}_t^{\mathbb{Q}}\right] = \beta_0 + \beta_1 \text{IMB}_o^{\text{SHAPE}} + \epsilon_o\,,\ o = 34 \text{ ETF options}\,.$$

Consistent with the prediction of my model, β_1 is positive (= 0.12) and has a t statistic of 1.99. The adjusted R^2 for this regression is 0.13. This result suggests that the cross-sectional difference in the correlation between the two risk-neutral moments can be partially captured by the difference in the trading activities across those ETF options markets.

Time Series Next, I run the following time series regression,

$$\text{Skew}_t^{\mathbb{Q}} = \beta_0 + \beta_1 \text{Variance}_t^{\mathbb{Q}} + \beta_2 \text{Variance}_t^{\mathbb{Q}} \times \text{IMB}_t^{\text{SHAPE}} + \gamma_1 \text{Skew}_{t,t+30}^{\mathbb{P}} + \varepsilon_t\,.$$

In particular, I expect β_2 to be positive and significant, as the joint correlation between the risk-neutral variance and skewness depends on $\beta_1 + \beta_2 \times \text{IMB}_t^{\text{SHAPE}}$. When customers' demand is concentrated at the OTM call options, the model predicts that the correlation between the risk-neutral variance and skewness will increase.

Table 1.10 shows that β_2 seems to be positive for most of the ETF options. However, only 6 out of 34 ETF options have statistically significant β_2. In addition, most of these significant results come from the commodity-linked ETF options. It is thus definitely worth exploring further the commodity options.

The insignificant results for other options may be due to the following fact: Certain customers may trade competitively and, hence, their order-flows do not create price pressure. Another measure for the shape of the customers' net demand is the ratio between the bid-ask spreads for OTM call options vs. put options, as the model predicts that the large trading cost for certain options is likely associated with a large imbalance in risk distribution and dealers' market power. Hence, the bid-ask spreads measured using intra-day data, separately for buy and sell orders, are valuable sources for explaining the patterns. I leave this for future research.

After the individual time series regression, I run the following panel regression with time and ETF fixed effects to estimate the coefficient β_2.

$$\text{Skew}^{\mathbb{Q}}_{o,t} = \beta_i + \beta_1 \text{Variance}^{\mathbb{Q}}_{o,t} + \beta_2 \text{Variance}^{\mathbb{Q}}_{o,t} \times \text{IMB}^{\text{SHAPE}}_{o,t} + \gamma_1 \text{Skew}^{\mathbb{P}}_{o,(t,t+30)} + \gamma_t + \varepsilon_{o,t}\,.$$

Table 1.11 summarizes the panel regression results. β_2 is positive and significant. Hence, the model seems to explain a fraction of the within ETF variations for the correlation between the two risk-neutral moments.

1.6 Conclusion

The real world option markets have a pronounced two-tier structure, whereby dealers trade with customers in the D2C market segment, and then use the D2D market to rebalance their inventories. For the first time in the literature, I develop a model of option markets that accounts for this two-tier structure. In my model, an endogenous structure of option implied volatilities and bid-ask spreads arises because of dealers' market power. This active role of dealers and their price shading behavior allows me to generate patterns of trade that are very different from other existing micro-structure models of option markets, including the demand-based option pricing theory of Gârleanu et al. [2009]. In particular, my model can explain a wide range of stylized facts about demand imbalance in option markets and its link to skewness and variance risk premia.

Given my model's ability to generate realistic option price behavior, it would be interesting to see whether the model can be used to extract physical probabilities from option prices, extending the ideas of Ross [2015]. Furthermore, while my model is static, it can easily be extended to dynamic settings, in which case I can study the joint endogenous nonlinear dynamics of imbalance and its impact on risk premia and the dynamics of the implied volatility surface. I leave these important questions for future research.

Table 1.3 – Regression of the variance risk premium on customers' total option demand.

Data is from January 2010 to April 2016. The bold numbers are significant at 5%. Daily frequency. I report the t-statistic for regressors based on White heteroskedasticity robust standard errors.

Ticker	$\text{IMB}_t^{\text{LEVEL}}$	adj. R^2	Obs
XRT	0.67	-0.00	1553
SMH	0.65	-0.00	611
XBI	**2.85**	0.01	552
XLY	-1.07	0.00	1530
IBB	1.89	0.00	1249
XLV	-0.10	-0.00	1520
XLI	**-2.80**	0.00	1537
XLU	-0.81	-0.00	1550
XLE	**2.22**	0.00	1560
IYR	-1.57	0.00	1558
XLF	-1.64	0.00	1558
DIA	0.27	-0.00	1561
IWM	-0.66	-0.00	1561
SPY	**-3.04**	0.00	1561
QQQ	-1.71	0.00	1254
ASHR	-0.69	-0.00	197
RSX	**2.29**	0.00	585
EWJ	**-2.40**	0.00	1298
DXJ	0.08	-0.00	672
EFA	0.11	-0.00	1561
EWZ	-1.24	-0.00	1560
FXI	-0.45	-0.00	1560
EEM	1.68	0.00	1561
HYG	-1.54	0.00	1270
TLT	0.10	-0.00	1560
UNG	0.01	-0.00	1558
OIH	1.61	0.00	1523
SLV	**2.36**	0.00	1560
GDX	0.35	-0.00	1560
GLD	-0.07	-0.00	1561
USO	-1.57	0.00	1561
XOP	**-2.03**	0.00	1502
FXE	0.76	-0.00	1556
UUP	0.35	-0.00	1285

Table 1.4 – Regression of risk-neutral variance on the demand pressure (level).

Control for the physical variance. The data is from January 2010 to April 2016. The bold numbers are significant at 5%. Daily frequency. I report the t-statistic for regressors based on White heteroskedasticity robust standard errors.

Ticker	$\text{IMB}_t^{\text{LEVEL}}$	$\text{Variance}_{t,t+30}^{\mathbb{P}}$	adj. R^2	Obs
XRT	-1.44	11.20	0.368	1553
SMH	-0.47	7.06	0.122	611
XBI	**2.50**	14.25	0.310	552
XLY	**-3.52**	11.55	0.336	1530
IBB	-0.63	13.42	0.214	1249
XLV	-0.66	11.33	0.226	1520
XLI	-1.90	11.63	0.375	1537
XLU	0.02	12.11	0.191	1550
XLE	**-3.13**	18.81	0.440	1560
IYR	**-4.08**	14.79	0.415	1558
XLF	-1.65	13.75	0.379	1558
DIA	0.01	12.29	0.303	1561
IWM	**-2.27**	13.50	0.421	1561
SPY	**-2.58**	12.09	0.330	1561
QQQ	-0.77	10.80	0.311	1254
ASHR	**-2.13**	-1.42	0.010	197
RSX	-0.99	7.48	0.340	585
EWJ	**-2.76**	5.25	0.020	1298
DXJ	0.34	8.05	0.198	672
EFA	0.26	12.25	0.386	1561
EWZ	-0.40	23.89	0.453	1560
FXI	-0.37	17.40	0.385	1560
EEM	**-1.97**	16.68	0.423	1561
HYG	-1.28	9.90	0.267	1270
TLT	-1.83	13.41	0.365	1560
UNG	-0.70	12.83	0.201	1558
OIH	-1.64	21.50	0.493	1523
SLV	**-2.52**	15.44	0.351	1560
GDX	**-2.47**	21.19	0.376	1560
GLD	-1.63	15.36	0.301	1561
USO	**-3.83**	23.56	0.558	1561
XOP	-1.03	27.91	0.512	1502
FXE	-0.57	21.75	0.407	1556
UUP	-1.21	19.03	0.346	1285

Table 1.5 – Regression of risk-neutral skewness on the demand pressure (skew).

Control for the realized skewness. The bold numbers are significant at 5%. Daily frequency. I report the t-statistic for regressors based on White heteroskedasticity robust standard errors.

Ticker	$\text{IMB}_t^{\text{SKEW}}$	$\text{Skew}_{t,t+30}^{\mathbb{P}}$	adj. R^2	Obs
XRT	1.46	-0.17	0.000	1553
SMH	1.89	3.30	0.016	611
XBI	-0.06	2.84	0.027	552
XLY	0.71	1.73	0.001	1530
IBB	1.05	2.33	0.007	1249
XLV	-1.22	0.58	-0.000	1520
XLI	1.91	0.57	0.001	1537
XLU	-1.29	4.15	0.011	1550
XLE	1.08	-0.04	-0.001	1560
IYR	-0.07	-1.37	-0.000	1558
XLF	1.76	1.66	0.003	1558
DIA	0.76	0.21	-0.001	1561
IWM	0.94	1.04	-0.000	1561
SPY	**-5.84**	2.12	0.020	1561
QQQ	0.53	5.01	0.015	1254
ASHR	-0.11	-2.75	0.006	197
RSX	0.33	1.56	-0.001	585
EWJ	-0.22	-0.47	-0.001	1298
DXJ	0.09	0.89	-0.002	672
EFA	-0.05	1.60	0.001	1561
EWZ	**2.08**	3.19	0.010	1560
FXI	**-2.28**	2.81	0.009	1560
EEM	0.72	1.04	-0.000	1561
HYG	0.55	0.14	-0.001	1270
TLT	**2.80**	1.02	0.005	1560
UNG	-1.30	6.76	0.017	1558
OIH	**3.10**	1.35	0.006	1523
SLV	1.47	9.39	0.017	1560
GDX	**2.48**	4.50	0.012	1560
GLD	-1.29	6.00	0.019	1561
USO	-0.80	5.30	0.016	1561
XOP	**3.35**	3.09	0.012	1502
FXE	1.44	6.25	0.022	1556
UUP	1.21	0.88	0.001	1285

Table 1.6 – Correlation for the skewness and variance risk premia, the realized variance and skewness, the risk-neutral variance and skewness.

RP stands for the correlation of the risk premium. P stands for the correlation of the realized moments. Q stands for the correlation of the risk-neutral moments. The bold numbers are significant at 5%. Daily frequency.

Ticker	Corr (RP)	Corr (P)	Corr (Q)
XRT	**0.17**	**-0.61**	**-0.23**
SMH	-0.07	**-0.31**	0.13
XBI	0.06	**-0.43**	**0.48**
XLY	**0.26**	**-0.58**	0.06
IBB	**0.11**	**-0.57**	**0.17**
XLV	0.09	**-0.59**	0.09
XLI	0.07	**-0.58**	**0.11**
XLU	**-0.11**	**-0.49**	-0.01
XLE	**0.20**	**-0.57**	0.06
IYR	0.06	**-0.44**	0.10
XLF	**0.16**	**-0.44**	0.07
DIA	**0.45**	**-0.63**	**0.13**
IWM	**0.39**	**-0.56**	**0.14**
SPY	**0.54**	**-0.53**	**0.18**
QQQ	**0.23**	**-0.58**	-0.05
ASHR	-0.11	-0.08	**0.67**
RSX	-0.02	**-0.24**	**0.17**
EWJ	-0.09	**-0.28**	**0.14**
DXJ	**0.24**	**-0.42**	-0.06
EFA	0.03	**-0.55**	**0.12**
EWZ	0.05	**-0.18**	**0.37**
FXI	-0.06	**-0.49**	**-0.43**
EEM	0.07	**-0.53**	0.02
HYG	**0.20**	**-0.60**	**-0.16**
TLT	-0.06	**0.36**	**0.38**
UNG	-0.01	**0.37**	**0.36**
OIH	0.08	**-0.50**	**0.23**
SLV	0.01	**-0.19**	0.04
GDX	-0.01	-0.01	**0.37**
GLD	-0.03	**-0.23**	0.09
USO	**-0.11**	-0.03	**0.33**
XOP	0.06	**-0.40**	**0.29**
FXE	-0.05	**-0.35**	**-0.46**
UUP	-0.05	**0.24**	**0.21**

Table 1.7 – Correlation between the risk-neutral variance and risk-neutral skewness.

The quantile is based on the shape measure, IMB_t^{SHAPE}. The bold numbers are significant at 5%. Daily frequency.

Quantile Ticker	0.2	0.4	0.6	0.8	1.0
XRT	-0.05	**-0.13**	**-0.17**	**-0.22**	-0.12
SMH	0.14	-0.09	0.06	0.07	-0.02
XBI	**0.39**	**0.42**	**0.43**	**0.38**	**0.22**
XLY	**0.16**	-0.09	-0.05	0.02	0.05
IBB	-0.01	**0.15**	0.06	0.10	-0.09
XLV	0.03	0.07	**0.19**	**0.14**	**0.14**
XLI	**0.12**	-0.01	0.05	0.08	0.05
XLU	-0.01	0.07	0.01	**0.13**	0.03
XLE	-0.01	**0.15**	-0.00	-0.02	0.09
IYR	0.02	0.06	0.08	0.04	**0.14**
XLF	0.09	**0.19**	**0.19**	**0.19**	**0.18**
DIA	**0.15**	0.06	**0.13**	0.11	0.08
IWM	0.09	**0.14**	**0.14**	0.11	**0.11**
SPY	**0.20**	0.09	**0.11**	**0.15**	**0.19**
QQQ	-0.07	-0.04	-0.05	0.08	0.07
ASHR	**0.55**	**0.51**	**0.46**	**0.35**	**0.61**
RSX	**0.21**	**0.27**	0.09	0.13	**0.27**
EWJ	**0.27**	**0.29**	**0.30**	**0.36**	**0.28**
DXJ	-0.09	-0.04	0.02	0.04	-0.12
EFA	0.07	0.08	**0.16**	-0.02	**0.13**
EWZ	**0.31**	**0.27**	**0.32**	**0.37**	**0.33**
FXI	**-0.36**	**-0.42**	**-0.26**	**-0.32**	**-0.32**
EEM	0.08	**0.14**	0.02	-0.02	-0.02
HYG	**-0.18**	-0.10	**-0.21**	-0.09	-0.06
TLT	**0.25**	**0.29**	**0.29**	**0.32**	**0.25**
UNG	**0.34**	**0.38**	**0.44**	**0.46**	**0.50**
OIH	0.11	**0.15**	**0.24**	**0.13**	**0.22**
SLV	**0.17**	0.06	**0.11**	0.08	-0.01
GDX	**0.22**	**0.25**	**0.36**	**0.29**	**0.35**
GLD	0.07	0.07	0.05	0.08	**0.13**
USO	**0.26**	**0.31**	**0.32**	**0.34**	**0.22**
XOP	0.04	**0.16**	**0.23**	**0.21**	**0.26**
FXE	**-0.25**	**-0.38**	**-0.19**	**-0.39**	**-0.27**
UUP	0.05	0.02	**0.17**	**0.22**	**0.17**

Table 1.8 – Correlation between the realized variance and realized skewness.

The quantile is based on the shape measure, IMB_t^{SHAPE}. The bold numbers are significant at 5%. Daily frequency.

Quantile Ticker	0.2	0.4	0.6	0.8	1.0
XRT	**-0.26**	**-0.23**	**-0.21**	**-0.18**	**-0.24**
SMH	0.08	**0.35**	**0.39**	**0.29**	0.15
XBI	-0.01	0.17	0.11	0.12	-0.09
XLY	**-0.18**	-0.11	**-0.14**	**-0.13**	**-0.23**
IBB	**-0.20**	-0.01	0.06	-0.02	**-0.22**
XLV	**-0.26**	**-0.18**	**-0.12**	**-0.19**	**-0.20**
XLI	**-0.11**	**-0.23**	**-0.21**	**-0.35**	**-0.21**
XLU	**-0.40**	**-0.43**	**-0.45**	**-0.45**	**-0.44**
XLE	0.09	-0.03	-0.02	**-0.17**	-0.01
IYR	**0.41**	**0.39**	**0.39**	**0.30**	**0.23**
XLF	0.02	**-0.16**	**-0.20**	**-0.18**	**-0.16**
DIA	**-0.17**	**-0.29**	**-0.27**	**-0.33**	**-0.37**
IWM	-0.02	0.06	**-0.15**	**-0.13**	-0.05
SPY	**-0.19**	**-0.25**	**-0.19**	-0.11	**-0.19**
QQQ	-0.01	-0.04	0.03	**-0.13**	-0.04
ASHR	**0.34**	0.28	**0.45**	**0.55**	**0.49**
RSX	-0.01	-0.16	-0.13	-0.04	**-0.20**
EWJ	**-0.30**	**-0.27**	**-0.33**	**-0.36**	**-0.28**
DXJ	0.11	**0.22**	**0.22**	0.16	**0.24**
EFA	-0.07	-0.05	-0.07	0.08	-0.10
EWZ	-0.04	**-0.14**	**-0.14**	**-0.22**	-0.10
FXI	**-0.18**	**-0.24**	**-0.24**	**-0.24**	**-0.34**
EEM	-0.09	**-0.19**	**-0.17**	**-0.27**	**-0.31**
HYG	**-0.29**	**-0.18**	**-0.15**	**-0.23**	**-0.16**
TLT	**0.24**	**0.13**	**0.16**	**0.24**	**0.33**
UNG	**0.44**	**0.38**	**0.41**	**0.42**	**0.42**
OIH	0.06	-0.08	**-0.26**	**-0.15**	-0.04
SLV	**-0.60**	**-0.58**	**-0.55**	**-0.55**	**-0.55**
GDX	-0.10	-0.02	-0.08	-0.07	-0.07
GLD	**-0.32**	**-0.34**	**-0.40**	**-0.40**	**-0.47**
USO	**0.24**	**0.24**	**0.23**	**0.26**	**0.29**
XOP	-0.06	**0.16**	**0.13**	**0.14**	0.09
FXE	-0.10	**-0.16**	**-0.28**	**-0.22**	**-0.11**
UUP	**0.22**	**0.45**	**0.34**	**0.31**	**0.34**

Table 1.9 – Correlation between the variance and skewness risk premia.

The quantile is based on the shape measure, IMB_t^{SHAPE}. The bold numbers are significant at 5%. Daily frequency.

Quantile Ticker	0.2	0.4	0.6	0.8	1.0
XRT	**0.27**	-0.01	0.12	0.03	0.09
SMH	0.18	0.11	0.00	0.06	-0.11
XBI	0.18	0.06	0.24	-0.08	0.03
XLY	**0.34**	**0.14**	**0.14**	**0.24**	**0.28**
IBB	**0.15**	0.09	0.07	**0.28**	-0.02
XLV	0.02	0.02	0.06	0.10	**0.03**
XLI	0.04	0.03	0.10	0.09	**0.16**
XLU	0.04	-0.03	-0.02	-0.08	**-0.21**
XLE	**0.19**	0.06	0.05	**0.12**	0.02
IYR	0.11	0.07	**0.13**	-0.02	0.09
XLF	0.02	0.10	0.08	0.10	0.11
DIA	**0.46**	**0.47**	**0.34**	**0.43**	**0.30**
IWM	**0.19**	**0.44**	**0.35**	**0.52**	**0.28**
SPY	**0.59**	**0.42**	**0.42**	**0.52**	**0.63**
QQQ	**0.33**	**0.28**	**0.56**	**0.44**	**0.48**
ASHR	-0.09	**-0.32**	-0.18	**-0.60**	0.22
RSX	0.02	-0.06	0.05	0.02	-0.02
EWJ	-0.03	-0.05	0.05	-0.10	0.00
DXJ	**0.26**	0.11	0.06	-0.13	0.00
EFA	**0.35**	**0.19**	**-0.16**	**0.19**	**0.24**
EWZ	0.11	**0.15**	0.09	0.04	0.06
FXI	0.11	0.10	0.08	-0.10	0.08
EEM	0.09	**0.20**	0.05	0.11	0.03
HYG	0.01	**0.27**	0.08	**0.22**	0.00
TLT	0.03	-0.06	-0.02	-0.05	-0.08
UNG	-0.03	-0.07	0.02	-0.05	-0.04
OIH	0.11	0.07	-0.01	0.02	-0.04
SLV	0.04	-0.04	-0.01	0.10	-0.07
GDX	0.01	-0.04	-0.05	0.07	-0.02
GLD	0.00	-0.02	0.06	-0.00	-0.04
USO	**0.15**	-0.07	-0.09	-0.02	-0.01
XOP	0.10	-0.06	**0.12**	-0.00	-0.07
FXE	-0.10	-0.04	0.01	0.04	0.07
UUP	0.07	-0.05	-0.04	**-0.13**	0.06

Table 1.10 – Regression of the risk-neutral skewness on the risk-neutral variance and the interaction between the risk-neutral variance and the shape measure.

Data is from January 2010 to April 2016. The bold numbers are significant at 5%. Daily frequency. I report the t-statistic for regressors based on White heteroskedasticity robust standard errors.

Ticker	$\text{Variance}_t^{\mathbb{Q}}$	$\text{Variance}_t^{\mathbb{Q}} \times \text{IMB}_t^{\text{SHAPE}}$	adj. $R2$	Obs
XRT	-4.28	1.66	0.018	1548
SMH	0.65	0.52	0.006	515
XBI	9.59	-1.56	0.145	540
XLY	0.15	-0.78	0.000	1514
IBB	1.02	-1.31	0.006	1203
XLV	3.80	1.03	0.012	1511
XLI	2.19	0.53	0.002	1531
XLU	1.73	0.16	0.011	1554
XLE	1.67	1.67	0.002	1571
IYR	3.14	0.54	0.004	1566
XLF	5.37	**2.25**	0.034	1569
DIA	4.46	0.48	0.010	1572
IWM	5.02	0.98	0.013	1572
SPY	6.44	-0.26	0.020	1572
QQQ	0.06	1.48	0.016	1265
ASHR	7.62	-1.83	0.237	205
RSX	3.81	-0.74	0.026	588
EWJ	10.26	1.81	0.087	1259
DXJ	-1.41	1.06	0.003	661
EFA	2.94	-0.05	0.007	1571
EWZ	12.05	1.32	0.094	1571
FXI	-15.61	0.38	0.109	1571
EEM	1.57	0.63	0.001	1572
HYG	-5.21	-0.44	0.015	1245
TLT	11.52	0.69	0.079	1570
UNG	16.51	**2.42**	0.203	1569
OIH	6.81	**3.15**	0.037	1527
SLV	3.04	1.18	0.021	1571
GDX	9.18	**2.78**	0.092	1571
GLD	2.60	1.64	0.026	1572
USO	10.05	1.19	0.092	1572
XOP	7.42	**3.03**	0.047	1507
FXE	-8.21	0.08	0.106	1561
UUP	1.70	**2.20**	0.019	1282

Table 1.11 – Panel regression with day and ETF fixed effects.

Daily frequency from January 2010 to April 2016. The total number of observation is 46,148.

Variables	**T-stat**
$\text{Variance}^{\mathbb{Q}}_{o,t}$	7.6
$\text{Variance}^{\mathbb{Q}}_{o,t} \times \text{IMB}^{\text{SHAPE}}_{o,t}$	3.9
$\text{Skew}^{\mathbb{P}}_{o,(t,t+30)}$	4.3
$R^2 = 0.04$ and within $R^2 = 0.0088$	

2 The Demand for Commodity Options

Semyon Malamud[1], Michael C. Tseng[1], Yuan Zhang[1]

1 – EPFL and Swiss Finance Institute

We develop a simple equilibrium model in which commercial hedgers, i.e., producers and consumers, use commodity options and futures to hedge price and quantity risk. We derive an explicit relationship between expected futures returns and the hedgers' demand for out-of-the-money options, and show that the demand for both calls and puts are positively related to expected returns, and the relationship is asymmetric, tilted towards puts. We test and confirm the model predictions empirically using the commitment of traders report from CFTC.

2.1 Introduction

In the original normal backwardation theory of Keynes and Hicks, producers take a short position in the futures market to hedge their exposure to price risk. Speculators require a positive risk premium for taking the other side of this trade, leading to positive expected futures returns. While this theory generates clear predictions about the linear hedging instruments, it is silent about the effects of hedging with non-linear instruments such as commodity options. The goal of this paper seeks to fill this gap and derive theoretical predictions about the interaction of risk premia and hedging demand in the options market.

To this end, we develop a simple, two period general equilibrium model populated by three types of agents: commodity producers, commodity consumers, and speculators. We assume that, in addition to price risk, producers face quantity risk: In this case, producers may in fact find it optimal to take a long position in the futures contract,[1] and expected futures returns are positive if and only if price risk is higher than quantity risk. We assume that, in addition to the futures contract, the agents have access to a full menu of out of the money (OTM) put and

[1] See, for example, Rolfo [1980].

call options rendering the market complete. The simple, complete market setup allows us to derive an explicit formula for the agents' option demand. We find that

(1) the hedgers' (producers plus consumers) demand for both OTM puts and OTM calls always has the same sign, and this demand is positively related to expected futures returns

(2) there is an asymmetry in the demand for OTM puts versus that for OTM calls: the demand for puts is significantly higher in absolute value even if expected returns are (moderately) positive

(3) the sensitivity of the log of the absolute size of the option demand to log option strike is positively related to expected returns

(4) the net hedgers' option demand in terms of OTM calls minus OTM puts is negatively related to expected futures returns

The intuition behind these findings is as follows. Since markets are complete and all agents have the same risk aversion, they linearly share the total revenue of producers and consumers. This total revenue is either a globally convex or a globally concave function of the spot price, and commercial hedgers (producers and consumers) jointly sell a fraction of these revenues to speculators. When these revenues are convex, it involves selling OTM puts and calls; when they are concave, it involves buying OTM options. Still, agents naturally worry more about the downside risk, hence the results of item (2). The result of item (3) is a direct implication of item (1) and (2): As puts dominate calls, the net option demand is driven by short OTM puts.

To test our model predictions empirically, we need to construct a measure of the net (signed) commercial hedgers' demand for options. We cannot directly use CFTC commitments of traders (COT) reports because these data report options demand that is aggregated across strikes and option types (for example, one cannot distinguish between a short position in the put and long position in the call). Hence, we cannot directly test the implications (1)-(3). However, CFTC data provides us a measure of OTM calls minus OTM puts, we then use this measure to test the prediction (4). We find strong support for the prediction on energy and precious metal sector.

The rest of this paper is organized as follows. Section 2.2 reviews relevant literature. Section 2.3 derives the equilibrium model. Section 2.4 contains empirical results. Section 2.5 concludes.

2.2 Literature Review

Our paper belongs to the literature that tries to understand the effects of hedging pressure in commodity markets. Starting with Keynes [1923] and Hicks [1939], many papers have argued that hedgers' supply of futures contracts (hedging pressure) drives down the futures price relative to the expected value of the later spot price, and this way it generates a downward bias (normal backwardation) in the futures price. See, for example, Stein [1961], Cootner [1960], Cootner [1967], and Stoll [1979]. Subsequent papers have argued that the sign of the hedging pressure can be ambiguous, for example, due to complementarities in consumer preferences (see, Hirshleifer [1990]) or due to quantity risk (see, for example, Rolfo [1980], Newberry and Stiglitz [1981], Newbery [1983], Anderson and Danthine [1983], Hirshleifer [1988a], and Hirshleifer [1988b].[2] The empirical evidence on the link between the commodity futures hedging pressure and futures returns is mixed. For example, Chang [1985] finds that futures prices for grains on average rise when hedgers are short, and fall when hedgers are long. Consistent with the Keynes-Higgs normal backwardation theory, Kang et al. [2017] find that hedgers are indeed on average net short in most commodity futures markets, while Moskowitz et al. [2012] and Cheng and Xiong [2014] provide evidence that hedger's price pressure is an important driver of the shape of the futures curve. However, Kang et al. [2017] find that hedgers follow short-term contrarian strategies and short-term fluctuations in hedging pressure are largely driven by the liquidity demands of speculators.[3] Taking a longer-term moving average of the hedging pressure eliminates these short term fluctuations and recovers the validity of Keynes' normal backwardation theory. All these papers consider exclusively futures hedging pressure. To the best our knowledge, our paper is the first one to study the effects of the hedging pressure in the commodity options market.[4]

In our model, hedging with options is optimal for producers and consumers because of the interaction between quantity and price risk. This links our paper to Brown and Toft [2002] and Gay et al. [2002, 2003] who show in a partial equilibrium setting that producers should use convex (concave) hedging strategies if price and quantity risks are negatively (positively) correlated.[5] By contrast, we show that in general equilibrium, when both commodity spot prices and option prices are determined by market clearing, the link between the hedging strategy and the quantity-price risk correlation is much more subtle.[6] The same interaction between quantity and price risk implies that the shape of the implied volatility smile in our model can be tilted both ways, depending on which risk dominates. In addition, we show that the shape of this smile is closely related to expected futures returns: when OTM puts (OTM calls) are expensive, expected futures returns are positive. These model predictions are consistent with the findings of Ellwanger [2015]. Furthermore, our model is also able to generate a negative commodity variance risk premium (that is, the difference between the risk

[2] In an influential paper, Litzenberger and Rabinowitz [1995], argued that backwardation can be attributed to the option value of commodity production.

[3] Such effects are partially consistent with the theory of Hong and Yogo [2012] who show that open interest growth rate is informative about futures returns due to under-reaction to news.

[4] The only exception is Woodard and Sproul [2016] who study the effects of hedging pressure in the options market createded by the Federal Crop Insurance Program. See also Gârleanu et al. [2009] who find evidence for the effects of hedging pressure using demand shocks for options of different strikes, as well as Hitzemann et al. [2016] who studies how hedging pressure impacts option returns in the presence of margin constraints.

[5] See also Lapan et al. [1991].

[6] Using a sample of US oil and gas producers, Mnasri et al. [2013] find evidence that the correlation between price and quantity risks is indeed important for hedging demand, but the relationship predicted by Brown and Toft [2002] does not hold in general.

neutral variance and the variance under the physical measure), consistent with the findings of Doran and Ronn [2008] and Trolle and Schwartz [2010].

Our paper also belongs to the literature on the informational content of options volume for expected returns. See, for example, Easley et al. [1998], Pan and Poteshman [2006b], Johnson and So [2012], and Roll et al. [2010b]. While we cannot exclude the effects of asymmetric information, our model implies that, for commodities, the net demand for calls and puts may contain information about future returns even when information is symmetric.

Our paper is also closely related to Moskowitz et al. [2012] show that the profitability of time series momentum strategies arises because speculators profit from time series momentum at the expense of hedgers, in agreement with Keynes [1923] theory. Importantly, consistent with Moskowitz et al. [2012], we find that the hedging demand in the option markets follows a one-year cycle, similar to that found in Moskowitz et al. [2012] for the futures market.

2.3 A Model for Commodity Options

There are two time periods, $t = 0, 1$, and two perishable goods, a consumption good and an investment good. The consumption good also serves as the numeraire and the prices of the investment good are quoted in the units of the consumption good. There are three types of agents in the model: producers, consumers, and speculators. All agents have the same CRRA utility function $u_p(c) = u_c(c) = u_I(c) = (1-\gamma)^{-1}c^{1-\gamma}$ with relative risk aversion γ. Consumption takes place at period-1. For simplicity, we assume the exogenous interest rate is 0.

Commodity Market We model the spot commodity market in period-1 as follows: (i) Both producers and consumers observe the spot price P for the commodity; (ii) Producers investing κ at period-0, face a productivity shock ε_q and price uncertainty[7] $P = \varepsilon_p q^{\delta-1}$ with some $\delta \in (0,1)$. Their production function is $q = \kappa\varepsilon_q^{1/\delta}$. Thus, producers' revenue is given by $Pq = \varepsilon_p q^\delta = \varepsilon_p\varepsilon_q\kappa^\delta$. (iii) The consumers have a technology to transform Q units of the investment good into $\delta^{-1}Q^\delta\varepsilon_p$ units of consumption good, where ε_p is their productivity shock. Their problem is to maximize the total revenue $-PQ + \delta^{-1}Q^\delta\varepsilon_p$, which gives the demand curve

$$P = \varepsilon_p Q^{\delta-1}.$$

Hence, their total revenue is $(\delta^{-1}-1)PQ = (\delta^{-1}-1)\varepsilon_p\varepsilon_q\kappa^\delta$.

We summarize the result from the commodity market in the following lemma.

Lemma 1. *The producers' total revenue in the commodity market is*

$$Pq = \varepsilon_p\varepsilon_q\kappa^\delta,$$

[7] This is the demand curve of the consumers.

while the consumers' total revenue is

$$(\delta^{-1}-1)PQ = (\delta^{-1}-1)\varepsilon_p\varepsilon_q\kappa^{\delta}.$$

Commodity Options Market We assume that there is a continuum of commodity options with strike price in the support of the spot commodity price P. The option market opens at period-0, and as options complete the market, there is a unique pricing kernel M.

Producers In addition to the revenue from the commodity market, the producers receive period-1 consumption good w_p. At period-0, the producers observe the equilibrium pricing kernel M in the options market and maximize

$$\max_{X_p} E[u_p(w_p-\kappa+\varepsilon_p\varepsilon_q\kappa^{\delta}+X_p)] \text{ s.t. } E[MX_p]=0.$$

Consumers In addition to the revenue from the commodity market, the consumers receive period-1 consumption good w_c. Like producers, consumers take positions in the options market to smooth out consumption across shocks. Given producers' investment κ and pricing kernel M for the options market, the consumers' problem is

$$\max_{X_c} E[u_c(w_c+(\delta^{-1}-1)\varepsilon_p\varepsilon_q\kappa^{\delta}+X_c)] \text{ s.t. } E[MX_c]=0.$$

Intermediaries Finally, speculators take the other side in the options market and maximize

$$\max_{X_I} E[u_I(w_I+X_I)] \text{ s.t. } E[MX_I]=0.$$

In equilibrium, market clearing implies $X_I=-X_p-X_c$.

Standard results for CRRA preferences imply that the agents will proportionally split the aggregate endowment in period-1, $w_p+w_c+w_I-\kappa_*+\varepsilon_p\varepsilon_q\delta^{-1}\kappa_*^{\delta}$, where κ_* is the equilibrium production rate. The following is true

Lemma 2. *Conditional on the production rate κ_*, the optimal demand for state contingent claims are given by*

$$\begin{aligned} X_p &= \lambda_p(w_I+w_p+w_c-\kappa_*+\varepsilon_p\varepsilon_q\delta^{-1}\kappa_*^{\delta})-(w_p-\kappa_*+\varepsilon_p\varepsilon_q\kappa_*^{\delta}) \\ X_c &= \lambda_c(w_I+w_p+w_c-\kappa_*+\varepsilon_p\varepsilon_q\delta^{-1}\kappa_*^{\delta})-(w_c+(\delta^{-1}-1)\varepsilon_p\varepsilon_q\kappa_*^{\delta}), \end{aligned}$$

whereas the constants λ_p and λ_c are consumption shares of producers and consumers, respectively. The pricing kernel is

$$M=\lambda_M(w_I+w_p+w_c-\kappa_*+\varepsilon_p\varepsilon_q\delta^{-1}\kappa_*^{\delta})^{-\gamma}, \lambda_M=E[(w_I+w_p+w_c-\kappa_*+\varepsilon_p\varepsilon_q\delta^{-1}\kappa_*^{\delta})^{-\gamma}]^{-1}.$$

The equilibrium futures price is

$$F = E[MP] = \lambda_M E\left[\varepsilon_p \varepsilon_q^{(\delta-1)/\delta} \kappa_*^{\delta-1}(w_I + w_p + w_c - \kappa_* + \varepsilon_p \varepsilon_q \delta^{-1} \kappa_*^{\delta})^{-\gamma}\right].$$

Since $E[M] = 1$, we have

$$F - E[P] = \mathrm{Cov}(M, P) = \lambda_M \mathrm{Cov}\left(\varepsilon_p \varepsilon_q^{(\delta-1)/\delta} \kappa_*^{\delta-1}, (w_I + w_p + w_c - \kappa_* + \varepsilon_p \varepsilon_q \delta^{-1} \kappa_*^{\delta})^{-\gamma}\right). \quad (2.1)$$

Thus, we arrive at the following result.

Proposition 1. *The following is true:*

(1) If quantity shocks are small (i.e., ε_q has a small variance), we have

$$F < E[P].$$

Thus, futures prices are expected to appreciate.

(2) If price shocks are small (i.e., ε_p has a small variance),

$$F > E[P].$$

Thus, futures price are expected to depreciate.

The intuition is as follows: When price shock dominates, producers want to short futures contract, creating a downward pressure in the futures market. Hence, we have the claim of item (1). On the other hand, when quantity shock dominates, producers want to long futures contract, creating an upward pressure in the futures market.

The scenario of item (1) is commonly viewed as the situation when futures prices are in normal backwardation: Assuming that the spot futures price follows a martingale, $E[P]$ coincides with the time zero spot price, P_0, in which case (1) implies that $P_0 > F$. Similarly, under the same martingale condition, item (2) implies that futures prices are in contango: $P_0 < F$. Note however that this link strongly relies on the assumption that spot prices follow a martingale. However, within our model, spot prices are almost never martingales because production decisions themselves depend on the equilibrium risk premium.

2.3.1 Futures Return and Options Demand

Our next goal is to understand the link between futures prices and the demand for options. To this end, we will assume that both ε_p, ε_q are powers of a common stochastic shock $Z : \varepsilon_p = Z^\alpha$ and $\varepsilon_q = Z^\beta$ for some positive random variable Z. This specification allows us to uniquely map the realizations for the shock Z to spot prices $P = \varepsilon_p \varepsilon_q^{(\delta-1)/\delta} \kappa_*^{\delta-1} = \kappa_*^{\delta-1} Z^{\alpha+\beta(\delta-1)/\delta}$.

Here, α and β have an intuitive interpretation: They measure the sensitivity of $\log \varepsilon_p$ and $\log \varepsilon_q$ to the common shock $\log Z$. Define

$$\psi \equiv \frac{\alpha+\beta}{\alpha+\beta(\delta-1)/\delta}. \tag{2.2}$$

The following result is a direct consequence of (2.1).[8]

Corollary 2. *Expected futures return, $E[P]/F-1$, is positive if and only if $\psi > 0$.*

The result of Corollary 2 is very intuitive: In agreement with Proposition 1, futures prices are expected to appreciate if and only if price shocks dominate quantity shocks. In particular, this is the case when $|\alpha| > |\beta|$. By Lemma 2, we have

$$\begin{aligned} X_p(P) &= (\delta^{-1}\lambda_p - 1)(P\kappa_*^{1-\delta})^{\psi}\kappa_*^{\delta} + const_c, \\ X_c(P) &= (\delta^{-1}\lambda_c - (\delta^{-1}-1))(P\kappa_*^{1-\delta})^{\psi}\kappa_*^{\delta} + const_p. \end{aligned} \tag{2.3}$$

Market completeness implies that producers can replicate the desired state-contingent contract (2.3) using options. We will assume that agents only trade simple, European calls and puts with maturity $t=1$. Given the time zero spot price, we refer call (put) options with strikes above (below) F out-of-the-money (OTM). Put-Call parity implies that in the money (ITM) options are redundant if the agents can trade the underlying futures contract as well as OTM options. Thus, the optimal trading strategy is not uniquely defined. Everywhere in the sequel, we will make the following assumption:

Assumption 4. *Agents only trade futures as well as OTM options.*[9]

Given an arbitrary twice continuously differentiable claim $W(x)$, integration by parts implies that the following is true:[10]

$$W(x) = W(F) + W'(F)(x-F) + \int_{-\infty}^{F} W''(K)(K-x)^+ dK + \int_{F}^{\infty} W''(K)(x-K)^+ dK,$$

and we arrive at the following result.

Lemma 3. *Suppose that P takes values in $(P_{\min}, P_{\max})$. Then, producers' optimal demand for*

[8] Corollary 2 can be shown by directly substituting P into (2.1):

$$\mathrm{Cov}[M,P] = \lambda_M \mathrm{Cov}[(w_I + w_p + w_c - \kappa_* + \delta^{-1}(P\kappa_*^{1-\delta})^{\psi}\kappa_*^{\delta})^{-\gamma}, P]$$

This is negative, as long as $\psi > 0$. Hence, the futures market is in backwardation.

[9] The open interest and volume in OTM options is several times higher than that in ITM options for all commodities in our sample.

[10] See, for exampple, Carr and Madan [2001] and Demeterfi and Zou [1999].

options of strike K is given by

$$\frac{\partial^2 X_p^*(K)}{\partial K^2} dK = \psi(\psi-1)(\delta^{-1}\lambda_p - 1)\kappa_*^{(1-\delta)\psi+\delta} K^{\psi-2} dK,$$
$$\frac{\partial^2 X_c^*(K)}{\partial K^2} dK = \psi(\psi-1)(\delta^{-1}\lambda_c - (\delta^{-1}-1))\kappa_*^{(1-\delta)\psi+\delta} K^{\psi-2} dK, \tag{2.4}$$

while, the joint demand of commercial hedgers (consumers plus producers) for OTM puts (respectively, OTM calls) is given by

$$D^{Put} = \delta^{-1}(\lambda_c + \lambda_p - 1)\psi\kappa_*^{\psi(1-\delta)+\delta}(F^{\psi-1} - P_{\min}^{\psi-1})$$
$$D^{Call} = \delta^{-1}(\lambda_c + \lambda_p - 1)\psi\kappa_*^{\psi(1-\delta)+\delta}(P_{\max}^{\psi-1} - F^{\psi-1}). \tag{2.5}$$

The consumers hold the most of the variable (risky) part of the endowment, hence, their option demand aligns perfectly with the hedgers' option demand. Meanwhile, the producers may trade differently comparing to the hedgers. In particular, if the producers' wealth w_p is large, then in equilibrium, they absorb a fraction of option orders from the consumers, leaving the rest being absorbed by the speculators.

Empirically, we do not observe the separate order-flows for producers and consumers. Instead, the information available is on the aggregate order-flows for the commercial hedgers (producers plus consumers). Hence, in the sequel, we focus on the hedgers' option demand.

Combining Lemma 3 with Corollary 2, we arrive at the following important result.

Proposition 3. *The following is true:*

- *when $\psi < 0$, both D^{Put}, D^{Call} (insurance, i.e., puts, seller) are positively related to expected futures returns (contango)*
- *when $0 < \psi < 1$, both D^{Put}, D^{Call} (insurance buyer) are positively related to expected futures returns (backwardation)*
- *when $1 < \psi < 2$, both D^{Put}, D^{Call} (insurance seller) are negatively related to expected futures returns (backwardation)*
- *when $\psi > 2$, both D^{Put}, D^{Call} (covered call seller) are negatively related to expected futures returns (backwardation)*

The intuition is as follows: First, ψ reflects the relative importance of price shock and quantity shock, when $\psi > 0$, the price shock dominates the quantity shock, and the futures market is in backwardation. Second, ψ also reflects the relationship between the aggregate endowment and commodity price P. In particular, when $0 < \psi < 1$, the aggregate endowment is a concave function of the commodity price (more so when the price realization is low), meaning that a drop in the price reduces disproportionately more aggregate consumption, hence hedgers buy options from speculators to hedge this risk, especially puts; when $1 < \psi < 2$, the aggregate endowment becomes a convex function of the commodity price, i.e., a drop in the price reduces disproportionately less aggregate consumption, hence hedgers supply put options to speculators, i.e., think of an insurance seller; in the extreme case when $\psi > 2$, an increase in the commodity price increases disproportionately more aggregate consumption, hedgers sell call options to speculators, i.e., think of a covered call selling strategy.

We can derive one more corollary from Proposition 3. We will see in the next section, that CFTC aggregates hedgers' options demand across puts and calls, and they assign a negative weight (the options' Black-Scholes delta) to long positions in put options. Hence, we need predictions on the relation between expected futures return, and hedgers' option demand in terms of $D^{Call} - D^{Put}$.

Corollary 4. *The following result is true:*

- *when $\psi < 0$, $D^{Call} - D^{Put}$ (puts seller dominates) are negatively related to expected futures returns (contango)*
- *when $0 < \psi < 1$, $D^{Call} - D^{Put}$ (puts buyer dominates) are negatively related to expected futures returns (backwardation)*
- *when $1 < \psi < 2$, $D^{Call} - D^{Put}$ (puts seller dominates) are positively related to expected futures returns (backwardation)*
- *when $\psi > 2$, both $D^{Call} - D^{Put}$ (covered call seller dominates) are negatively related to expected futures returns (backwardation)*

In fact, we argue that the first two cases correspond usually to commodity markets in practice. For example, for Crude Oil markets, we can find periods of contango (the price boom before the financial crisis due to positive demand shock), as well as periods of normal backwardation (commodity index funds earn positive returns for decades). In addition, the trading volume in OTM puts is almost always larger than that in OTM calls, suggesting the last item rarely happens. The third item is also rare for energy and metal market, as when the market is in backwardation, hedgers buy options instead of selling. Hence, in the empirical section, we mainly seek to test the first two items of Corollary 4. Before proceed to empirical tests, we want to know whether this result is robust when the production rate κ_* is endogenously determined.

2.3.2 Endogenous Production Rate

In this section, we determine the production rate κ in equilibrium, and examine the robustness of our results, Proposition 3. To do so, we assume that producers are small (continuum of measure 1), hence their individual production rate does not alter the aggregate production rate. Then from Lemma 2, we know the optimal demand of the contingent claims for the producer is

$$X_p(\kappa;\kappa_*) = \lambda_p(\kappa)(w_I + w_p + w_c - \kappa_* + \varepsilon_p\varepsilon_q\delta^{-1}\kappa_*^{\delta}) - (w_p - \kappa + \varepsilon_p\varepsilon_q\kappa^{\delta}),$$

given his own production rate κ and the equilibrium production rate κ_*. Specifically, the producer's decision κ only affects the consumption share he receives in equilibrium through $\lambda_p(\kappa)$ by increasing his individual wealth $w_p - \kappa + \varepsilon_p\varepsilon_q\kappa^{\delta}$. The aggregate endowment instead is unaffected by κ. From the budget constraint $E[MX_p] = 0$, we get that

$$\lambda_p(\kappa) = E[M(w_p - \kappa + \varepsilon_p\varepsilon_q\kappa^{\delta})]/E[M(w_I + w_p + w_c - \kappa_* + \varepsilon_p\varepsilon_q\delta^{-1}\kappa_*^{\delta})].$$

Hence, the producer maximizes his indirect utility function

$$\max_{\kappa}(1-\gamma)^{-1}\lambda_p(\kappa)^{1-\gamma}E\left[(w_I + w_p + w_c - \kappa_* + \varepsilon_p\varepsilon_q\delta^{-1}\kappa_*^{\delta})^{1-\gamma}\right].$$

As κ only shows up in the consumption shares $\lambda_p(\kappa)$, we get the first-order condition for κ_* as

$$\kappa_*^{1-\delta} = \delta E[M(\kappa_*)\varepsilon_p\varepsilon_q]. \tag{2.6}$$

This is a fixed point problem to determine κ_*. As we do not have an explicit formula for κ_*, this makes the comparative statics cumbersome. To have a cleaner prediction, we perturb the price shock α from a benchmark where the futures price does not exhibit contango nor normal backwardation, and the option demand is zero.

Benchmark: no aggregate endowment risk Suppose that $\alpha + \beta = 0$, and without loss of generality, we assume that $\alpha > 0$. In this case, the price risk ε_p and the quantity risk ε_q are perfectly negatively correlated, meaning a negative price shock is associated with a positive quantity shock with the same size. Hence, the aggregate endowment in the economy becomes risk-less, although the commodity price is still risky. Note that, the producers have no incentives to hedge the risks in commodity price, as the quantity risk is a nature hedge for price risk. In this particular example, the pricing kernel M becomes a constant

$$M = 1,$$

for any realization of the common shock Z. Hence, from equation (2.6), we have that

$$\kappa_* = \delta^{1/(1-\delta)}.$$

The consumption share of the producer is

$$\lambda_p = \frac{w_p - \kappa_* + \kappa_*^{\delta}}{w_I + w_c + w_p - \kappa_* + \delta^{-1}\kappa_*^{\delta}},$$

and that of the consumer is

$$\lambda_c = \frac{w_c + (\delta^{-1} - 1)\kappa_*^{\delta}}{w_I + w_c + w_p - \kappa_* + \delta^{-1}\kappa_*^{\delta}}.$$

For the speculators, we know that their share is $1 - \lambda_p - \lambda_c$. The futures price is

$$F = E[MP] = E[P].$$

Hence, we know that the expected futures return is

$$E[P]/F - 1 = 0.$$

The equilibrium future spot commodity price P is

$$P = \kappa_*^{\delta-1} Z^{\alpha/\delta}.$$

Interestingly, the commodity price is risky, but no agents would hedge it.

Perturbation on α We assume that the price risk changes by a small quantity $\epsilon\alpha^{(1)}$

$$\alpha_\epsilon = \alpha + \epsilon\alpha^{(1)}.$$

The optimal production rate now becomes $\kappa_\epsilon = \kappa_* + \epsilon\kappa^{(1)}$. The following is true.

Lemma 4. *The optimal production rate is*

$$\kappa_\epsilon = \kappa_* + \epsilon\frac{\delta}{1-\delta}\kappa_*^{\delta} E[\log Z]\alpha^{(1)}. \tag{2.7}$$

Hence, as long as the price risk increases, i.e., $\alpha^{(1)} > 0$, we have that the producer increases the production rate. Next, we use Lemma 3 to get

Proposition 5. *The optimal option demand per strike for producer is*

$$D_{p,\epsilon}(K) = \epsilon\frac{\delta - \lambda_p}{\alpha}\kappa_*^{\delta} K^{-2}\alpha^{(1)} dK, \tag{2.8}$$

and that for the consumer is

$$D_{c,\epsilon}(K) = \epsilon\frac{1 - \delta - \lambda_c}{\alpha}\kappa_*^{\delta} K^{-2}\alpha^{(1)} dK. \tag{2.9}$$

Corollary 6. *Hence, we know that hedgers' aggregate option demand per strike is*

$$D_\epsilon(K) = \epsilon\frac{1 - \lambda_p - \lambda_c}{\alpha}\kappa_*^{\delta} K^{-2}\alpha^{(1)} dk. \tag{2.10}$$

Meanwhile, the option demand for OTM puts is

$$D_{\epsilon}^{Put} = \epsilon \frac{1-\lambda_p-\lambda_c}{\alpha}\kappa_*^{\delta}(P_{\min}^{-1} - F^{-1})\alpha^{(1)}, \tag{2.11}$$

and the demand for OTM calls is

$$D_{\epsilon}^{Call} = \epsilon \frac{1-\lambda_p-\lambda_c}{\alpha}\kappa_*^{\delta}(F^{-1} - P_{\max}^{-1})\alpha^{(1)}. \tag{2.12}$$

Lemma 5. *The first-order expansion of the pricing kernel is*

$$M^{(1)} = \frac{\gamma\kappa_*^{\delta}}{-\delta\kappa + \delta w + \kappa_*^{\delta}}(E[\log Z] - \log Z)\alpha^{(1)}. \tag{2.13}$$

We define $w \equiv w_I + w_p + w_c$. Thus the difference between the futures price and futures spot price is

$$F_{\epsilon} - E[P_{\epsilon}] = -\epsilon \frac{\gamma\kappa_*^{\delta}}{-\delta\kappa_* + \delta w + \kappa_*^{\delta}}\operatorname{Cov}[\log Z, P]\alpha^{(1)}. \tag{2.14}$$

Note that $P = \kappa_*^{\delta-1}Z^{\alpha/\delta}$.

Hence, we learn that when $\alpha > 0$, a 'small' positive shock to the price risk ($\psi > 0$) immediately induce hedgers to buy put options from the speculators, and the expected futures return is positive (normal backwardation). This implies a positive correlation between expected futures return and aggregate option demand, consistent with Proposition 3. On the other hand, a positive wealth shock to the speculators capital w_I, assuming $\alpha^{(1)} > 0$, increases the expected return for the futures contract, and increases the demand for options for both calls and puts.

Hence, we confirm that our previous results still hold for endogenous production rate κ_*, when the aggregate endowment risk is 'small'. In the next section, we test these theoretical predictions empirically.

2.4 Empirical Results

In this section, we test the model's predictions on the relationship between the hedgers' option demand and the expected futures return.

Empirically, hedgers' option demand is hard to measure for the following reasons: (i) most of the commercial hedgers are small and hence they hedge their production risk via dealer banks (over-the-counter trading); (ii) for large hedgers, they might trade options for reasons other than hedging the risk of their business; (iii) options trading on exchange (such as Chicago Mercantile Exchange) is anonymous.

We use the CFTC commitment of traders report to build our proxy for hedgers' option demand. This data can partially resolve the issues being mentioned. However, it has its own shortcomings, which we will discuss now.

CFTC Hedgers' Option Demand The CFTC disaggregated commitment of traders' report provides information on open interest for US exchange-traded futures and options. The data is generated each Tuesday. The advantages of this data are: (i) the report classifies open interest by traders' types: Commercials, Swap Dealers, Money Managers, Other Reportable; (ii) the data dates back to June, 2006 for the disaggregated version, and to even earlier dates for the old reporting format; (iii) the report distinguishes long and short positions for each traders' category.

However, the disadvantages are: (i) the report does not separate options into puts and calls, nor strikes and maturities; (ii) the report only comes out once a week on Tuesday; (iii) although the disaggregated format corrects some problems in traders' classification, there are still some traders being assigned to the wrong group. Specifically, the puts and calls are transformed into the equivalent futures open interest by multiplying the corresponding Black-Scholes delta. Hence, when we see a report with open interest on the long side, it does not tell whether the group of traders is holding calls, or is selling puts. To circumvent this issue, as we have already mentioned in the previous section, we compute the net option positions for the group of traders, which is approximately calls minus puts. The last problem is hard to address, as we cannot see the identities of traders within each group besides the group name. However, we know at least that for energy and metal, Commercials include mainly producers and consumers of the commodity; Swap Dealers might represent the aggregate positions of small hedgers, or part of speculators' positions; Money Managers usually represent speculators; Other Reportable (e.g., pension funds, sovereign funds, high-frequency traders) contains everyone else.

We compute the option demand measure as follows: For each group of traders, we compute its option open interest of the long-side and the short-side, by removing the futures positions.

$$\mathrm{OOI}_t^{Long} = \mathrm{OI}_t^{Long,futures+options} - \mathrm{OI}_t^{Long,futures}.$$

Then we take the difference between the long side and the short side to get the net option position

$$\mathrm{NOOI}_t = \mathrm{OOI}_t^{Long} - \mathrm{OOI}_t^{Short}.$$

However, to measure the hedgers' option demand related to shocks, we compute the innovations in the net option positions

$$D_{t-k,t}^{Call-Put} = \mathrm{NOOI}_t - k^{-1}\sum_{\tau=1}^{k} \mathrm{NOOI}_{t-\tau},$$

for $k = 12, 24, 36, 50$ weeks. We compute this measure for Commercials, Swap Dealers, Money Managers, and Other Reportable separately. We expect that Commercials category captures the option demand of hedgers as defined in our model, while Swap Dealers might capture partially hedgers and partially speculators' option demand. For the other two categories, we sum up their positions, and call them Financials, which is a proxy for the speculators' option demand.

According to Corollary 4, we expect the Commercials' option demand being negatively correlated with the realized futures return, while the Financials' option demand being positively correlated with the realized futures return. Formally, we run the following test

$$R_{t,t+h} = \beta_0 + \beta_1 D^{Call-Put}_{t-k,t} + \beta_2 R_{t-h,t} + \beta_3 R_{t-k,t} + slope_t + \varepsilon_{t,t+1}, \tag{2.15}$$

with $h = 1, 2, 4, 12$ weeks.

We are mainly interested in the sign of β_1. We also control for the following variables to isolate the information content of our measure: (i) lagged returns, as commodity futures returns exhibit positive auto-covariance (see, e.g., Moskowitz et al. [2012]); (ii) lagged returns corresponding to the option demand measure, as option demand and underlying futures returns are correlated contemporaneously; (iii) slope of the futures term structure, i.e., the log difference between the front month futures contract and the next period futures contract.

Tables below confirm our main predictions: the hedgers' net option demand negatively predicts future returns for energy and metal. In particular, Table 2.1 shows that for energy sector, the Commercials' net option demand indeed negatively forecast front month futures return. Interestingly, in Table 2.2, we see that for metal sector, the Commercials' net option demand has the right sign, but does not show much statistical significance. Hence, we need to understand whether this is due to measurement error, or due to our model's lack of explanation. We proceed by looking at the Swap Dealers' option demand, which might capture partially the hedgers' demand. If a significant fraction of gold hedgers use swap dealers to facilitate hedging, we should expect that the Swap Dealers' option demand negatively predicts futures return. Table 2.3 confirms our conjecture. In addition, if we cannot accurately measure the hedgers' option demand, we may look directly at the speculators' option demand. Table 2.4 shows that indeed, the Financials' option demand positively predicts the futures return.

2.5 Conclusions

We develop an extension of Keynes-Higgs normal backwardation theory that incorporates hedging using commodity options. Our model predicts that hedgers' net demand for options is negatively related to expected futures returns. We use CFTC data on commodity options to construct net demand measures for energy and metal sector and confirm the model predictions empirically.

Table 2.1 – CFTC Commercials' (Energy) Option Demand v.s. Expected Futures Return.

The numbers shown are the t-statistics, and the data ranges from 2006.06-2018.04. The regressions errors are corrected according to Newey-West with lags equal to the forecasting horizon.

	Horizon (#Weeks) Lag (#Weeks)	1	2	4	12
Crude Oil (CL)	12	-2.26	-3.54	-4.78	-1.79
	24	-2.36	-3.08	-3.42	-1.57
	36	-2.32	-2.84	-3.03	-1.34
	50	-2.22	-2.62	-2.65	-1.20
Natural Gas (NG)	12	-1.81	-1.60	-1.91	-1.51
	24	-1.91	-1.60	-1.92	-2.27
	36	-2.03	-1.78	-1.95	-1.95
	50	-1.85	-1.56	-1.65	-1.83
Gasoil (RB)	12	-2.63	-1.63	-0.97	-2.03
	24	-2.91	-2.10	-1.48	-2.80
	36	-3.09	-2.25	-1.54	-1.97
	50	-2.99	-2.27	-1.53	-2.32
Heating Oil (HO)	12	-2.34	-2.35	-1.62	-1.44
	24	-2.36	-2.26	-1.51	-1.49
	36	-2.49	-2.42	-1.66	-1.64
	50	-2.63	-2.63	-1.86	-1.66

Table 2.2 – CFTC Commercials' (Metal) Option Demand v.s. Expected Futures Return.

The numbers shown are the t-statistics, and the data ranges from 2006.06-2018.04. The regressions errors are corrected according to Newey-West with lags equal to the forecasting horizon.

	Horizon (#Weeks) Lag (#Weeks)	1	2	4	12
Gold (GC)	12	-1.25	0.27	-1.47	-0.27
	24	-0.99	0.12	-1.80	-0.96
	36	-0.92	-0.02	-1.74	-0.31
	50	-0.93	-0.15	-1.84	-0.47
Silver (SI)	12	0.78	0.52	-0.78	-0.41
	24	-0.09	-0.42	-1.80	-1.06
	36	-0.56	-0.93	-2.18	-1.29
	50	-0.80	-1.14	-2.31	-1.55

Table 2.3 – CFTC Swap Dealers' (Metal) Option Demand v.s. Expected Futures Return.

The numbers shown are the t-statistics, and the data ranges from 2006.06-2018.04. The regressions errors are corrected according to Newey-West with lags equal to the forecasting horizon.

	Horizon (#Weeks) Lag (#Weeks)	1	2	4	12
Gold (GC)	12	-2.25	-1.96	-1.98	-1.62
	24	-1.79	-1.55	-1.96	-1.36
	36	-1.64	-1.45	-1.78	-1.22
	50	-1.49	-1.35	-1.79	-1.16
Silver (SI)	12	-1.52	-1.36	-1.40	0.27
	24	-1.42	-1.49	-1.39	0.36
	36	-1.43	-1.55	-1.37	0.37
	50	-1.43	-1.55	-1.32	0.14

Table 2.4 – CFTC Financials' (Metal) Option Demand v.s. Expected Futures Return.

The numbers shown are the t-statistics, and the data ranges from 2006.06-2018.04. The regressions errors are corrected according to Newey-West with lags equal to the forecasting horizon.

	Horizon (#Weeks) Lag (#Weeks)	1	2	4	12
Gold (GC)	12	2.50	1.65	2.43	1.58
	24	1.97	1.36	2.40	1.84
	36	1.80	1.31	2.15	1.47
	50	1.69	1.26	2.24	1.56
Silver (SI)	12	1.00	1.22	1.52	-0.27
	24	1.28	1.65	1.83	0.04
	36	1.46	1.85	1.92	0.19
	50	1.53	1.90	1.91	0.34

3 Electronic Trading in OTC Markets vs. Centralized Exchange

Ying Liu[1], Sebastian Vogel[2], Yuan Zhang[2]

1 – UNIL and Swiss Finance Institute

2 – EPFL and Swiss Finance Institute

We model a two-tiered market structure in which an investor can trade an asset on a trading platform with a set of dealers who in turn have access to an interdealer market. The investor's order is informative about the asset's payoff and dealers who were contacted by the investor use this information in the interdealer market. Increasing the number of contacted dealers lowers markups through competition but increases the dealers' costs of providing the asset through information leakage. We then compare a centralized market in which investors can trade among themselves in a central limit order book to a market in which investors have to use the electronic platform to trade the asset. With imperfect competition among dealers, investor welfare is higher in the centralized market if private values are strongly dispersed or if the mass of investors is large.

3.1 Introduction

Trading in over-the-counter (OTC) markets is traditionally done over the phone, i.e. an investor who wants to trade an asset has to call a dealer and negotiate the price bilaterally. A recent trend in OTC markets is the growing electronification. Instead of calling dealer by dealer separately, an investor can use electronic trading platforms to send a request-for-quote (RFQ) to many dealers at once to obtain quotes at which the dealers are willing to trade. Some estimates suggest that in 2015, more than 40% of OTC-traded credit default swaps and more than 60% of OTC-traded interest rate swaps were traded electronically.[1]

Electronic trading platforms can potentially increase the connectedness between market

[1] See for instance Stafford [2016] for a brief overview of recent developments in OTC markets.

participants and thereby make OTC markets more exchange-like. However, there remains a fundamental difference between centralized exchanges and electronic trading platforms in OTC markets. Whereas exchanges can be viewed as all-to-all platforms, electronic trading platforms in OTC markets are one-to-many platforms. On exchanges, each market participant can trade through a central limit order book with all other market participants. On electronic trading platforms, the RFQ trading protocol prescribes that only one investor can initiate a trade at a time and choose one dealer to trade with. Therefore, electronic trading still incorporates many of the features of traditional bilateral trading in OTC markets.

The contribution of this paper is twofold. First, we model the trading process on trading platforms via an RFQ protocol. In our model, an investor who has some information about the asset's payoff can choose a quantity to trade on the platform. In equilibrium, this quantity is informative about the asset's payoff. Our model therefore provides a theoretical foundation of information leakage on electronic trading platforms that is examined in empirical studies such as Hendershott and Madhavan [2015] or Hagströmer and Menkveld [2016]. Increasing the number of dealers who are contacted by an RFQ has three competing effects on trading costs: If an RFQ is sent to more dealers, (i) competition among dealers lowers the expected markup the investor has to pay, (ii) the investor is more likely to receive a quote in the first place, since each dealer's response is uncertain and (iii) information leakage about the asset's fundamental value increases the dealer's cost of providing the asset, which results in worse prices for the investor. If dealers respond very frequently to each RFQ, the cost of information leakage dominates the benefits of more competition and contacting only few dealers maximizes the investor's payoff. Only if the dealers' RFQ response rate is sufficiently low, an RFQ has to be sent to a certain minimum number of dealers in order for an equilibrium to exist in the first place. In an off-equilibrium analysis, we deal with the price impact an investor faces on the platform. The presence of adverse selection makes the permanent price impact on the trading platform larger than the permanent price impact in the interdealer market. This result is consistent with the findings of Collin-Dufresne et al. [2017].

Second, we determine conditions under which investors are better off trading on a centralized exchange among themselves and when they are better off in the two-tiered market structure with an electronic trading platform and an interdealer market. In our model, all investors are equally informed about the asset's fundamental value and benefits from trade in the centralized market only arise due to private values of obtaining the asset (e.g. hedging benefits). Since dealers are less informed about the asset's value, investors can also benefit from their information about the asset in the OTC market structure. The dealers are willing to trade with the more informed investor, because they expect to be able to partially offset the trade at a favorable price in the interdealer market. If private values of obtaining the asset are small, investors are better off in the OTC market structure where they can benefit from information asymmetries between them and the dealers. On the other hand, if the total mass of investors is large, information about the asset's fundamental value quickly leaks into the interdealer market. In this case, the price investors have to pay on the platform is approximately the sum of the fundamental value and a markup. Then, investors are better off in the centralized exchange where they can avoid the dealers' markups and uncertainty about transactions. Only if competition among dealers is very high, investors will prefer to trade in the OTC markets. In this case, markups are very low, a trade is very likely and dealers efficiently intermediate trades between their customers. Additionally, investors can benefit from their information advantage over dealers in the OTC market. These results extend previous research on the comparison between OTC markets and exchanges in terms of investor welfare [Babus and Parlatore, 2017, Glode and Opp, 2017]. In this strand of literature, our study is the first one to specifically look at electronic trading platforms.

The paper proceeds as follows. Section 3.2 relates our paper to previous research. In Section 3.3, we explain the basic setup that is studied in Section 3.4. In Section 3.5, we slightly modify this setup to accommodate a continuum of investors and compare the two-tiered market structure to a centralized market. Concluding remarks are presented in Section 3.6. All proofs are in Appendix A.

3.2 Related Literature

Collin-Dufresne et al. [2017] empirically study the two-tiered index CDS market in the US. In the market for the most liquid index CDSs, the Dodd-Frank Act required trading via swap execution facilities (SEFs). As a result, investors trade with dealers almost exclusively via RFQs on electronic trading platforms. Dealers, on the other hand, trade among themselves via a continuous limit order book.[2] This market structure very closely corresponds to the setup we assume in our paper. The results of Collin-Dufresne et al. [2017] suggest that the permanent price impact in the D2C segment, i.e. when the investor trades on the platform, is higher than the permanent price impact in the interdealer market. These results justify our assumption that investors have some information about the asset that dealers do not have and are consistent with our result that there is information leakage from the trading platform to the dealers. Hendershott and Madhavan [2015] empirically study what kind of bonds are traded over the phone and which bonds are traded on an electronic trading platform. Controlling for endogenous venue selection, they examine the trading costs on these two trading venues. Hagströmer and Menkveld [2016] estimate information flows between dealers and provide further empirical evidence for information leakage on trading platforms in OTC markets. Bjønnes et al. [2008] and Bjønnes et al. [2016] argue that dealers in the foreign exchange market learn from their clients' order flow and exploit this information in the interdealer market.

Babus and Parlatore [2017] and Glode and Opp [2017] theoretically study investor welfare in OTC markets and centralized markets. Our model is different from those studies, since we specifically assume an RFQ trading protocol in the OTC market. Moreover, the information structure in our model differs from that in Babus and Parlatore [2017], since we have a common

[2] Block trades are exempt from the requirement to be traded on SEFs. However, most trades in the interdealer market are executed in the continuous limit order book, which also allows for mid-market matching and workup.

value of the asset for both investors and dealers. Compared to Glode and Opp [2017] we allow the investor to trade continuous quantities of the asset in the OTC market. Malamud and Rostek [2017] show that decentralized exchange markets may be more efficient than centralized ones. Lester et al. [2017] show in a search-theoretic model that competition in fragmented markets may decrease welfare.

In modeling the information leakage on trading platforms, our paper relates to a large strand of literature that models how information is shared between economic agents. Notable papers in this strand of literature include Duffie and Manso [2007], Duffie et al. [2009], Andrei and Cujean [2017] and Babus and Kondor [2016]. Traditionally, OTC markets are modeled as pure search markets as for instance in Duffie et al. [2005], Weill [2007], Lagos and Rocheteau [2009], Gârleanu [2009], Lagos et al. [2011], Feldhütter [2005], Pagnotta and Philippon [2011] or Lester et al. [2015]. Zhu [2012] and Duffie et al. [2017] explicitly model dealer markets. Our paper differs from all of those those papers since we consider an electronic trading platform.

Our assumption that dealers' responses on trading platforms are uncertain has been used by Jovanovic and Menkveld [2015] and Yueshen [2017] to model the behavior of market makers in central limit order books to derive similar random-pricing strategies.[3]

We also draw on the techniques of noisy rational-expectations models of Grossman and Stiglitz [1980], Hellwig [1980] and Diamond and Verrecchia [1981]. These models assume that agents behave competitively. Kyle [1989] showed that those models can be extended to allow for strategic traders that take their price impact into account. However, few closed-form solutions are available in this case. Since the competitive case is generally viewed as a reasonable approximation to the strategic case in large markets [Vives, 2010], we will model a competitive dealer market.

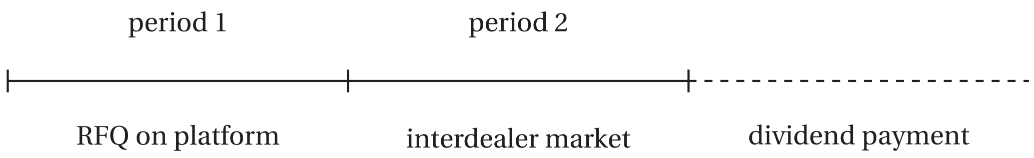
As Pagano and Röell [1996] argue, auction markets are in many ways more transparent than bilateral dealer markets. Naik et al. [1999] show that increased post-trade transparency has an ambiguous effect on dealers risk-sharing ability in two-tiered markets. Other papers who study the effects of transparency include De Frutos and Manzano [2002] and Yin [2005]. In this paper, however, we do not consider any specific disclosure policies that are enforced by regulators. In our model, information is disseminated through the different trading mechanisms.

3.3 Model

There are two periods and two types of agents. In the first period, an investor can contact a number of dealers via an RFQ trading protocol on an electronic trading platform to buy or sell a quantity of an asset. In the second period, dealers trade with each other in a central limit order book. After period 2, the dividend is paid. This is illustrated in Figure 3.1.

[3] Random-pricing strategies in turn have their origin in the consumer search literature. See for instance Varian [1980], Burdett and Judd [1983], Stahl [1989] and Janssen et al. [2005, 2011].

Figure 3.1 – Timeline



The asset pays an uncertain dividend $D = \theta + \varepsilon$ after the second period, where θ and ε are both independent and normally distributed random variables with zero mean and variances $\sigma_\varepsilon^2 > 0$ and $\sigma_\theta^2 > 0$, respectively. The informed investor knows the realization of θ already in the beginning of period 1. The investor also receives a private benefit $\delta \sim \mathcal{N}(0, \sigma_\delta^2)$, with $\sigma_\delta^2 \geq 0$ for holding one unit of the asset. This private benefit is realized in the beginning of period 1 and is independent from all other random variables. Only the investor can observe δ.

There are N, $\mathbb{N} \ni N \geq 2$ dealers. On the trading platform, the investor can specify the quantity x of the asset he wants to trade. The investor also selects M dealers, with $\mathbb{N} \ni M \leq N$ from which he wants to obtain prices at which they are willing to offer quantity x of the asset. The dealers respond independently with probability $q \in (0, 1]$ to the RFQ. That dealers do not necessarily respond may reflect the cost of paying attention. We will throughout this paper assume that the number of contacted dealers M is exogenously given, i.e. the trading protocol specifies that the investor has to contact exactly M dealers. This is a slight simplification of RFQ protocols in real-world markets where investors can often freely choose a number of dealers to contact.

The dealers are ex ante identical and hold zero initial inventory in the beginning of period 1. In period 2, the aggregate supply of the asset in the interdealer market is noisy. We denote the aggregate supply of the asset in the interdealer market by W. This aggregate supply is normally distributed: $W \sim \mathcal{N}(0, \sigma_W^2)$ and $\sigma_W^2 > 0$. A noisy aggregate supply is necessary in order to prevent uninformed dealers from observing the information of informed dealers. One can interpret noise in the aggregate supply as demand from noise traders or inventory shocks to dealers' portfolios, even though there is a slight difference between inventory shocks and noisy aggregate supply. Both dealers and the investor have mean-variance preferences. There is no discounting and each agent's utility is linear in the payments made when trading the asset. Let $\overline{\omega}_k$ denote dealer k's final inventory in the end of period 2 and let Z_k denote the sum of all payments made or received by dealer k from trading the asset. Then dealer k's utility in the end of period 2 with final inventory $\overline{\omega}_k$ is given by

$$U_d(\overline{\omega}_k, Z) = \overline{\omega}_k \cdot \mathbb{E}(D|\mathscr{I}_k) - \frac{\gamma_d}{2} \cdot \overline{\omega}_k^2 \cdot \mathbb{V}(D|\mathscr{I}_k) - Z_k, \tag{3.1}$$

where $\gamma_d > 0$ is the dealers' risk-aversion parameter. The expectation and the variance in equation (3.1) are taken with respect to each dealer k's specific information set $\mathscr{I}_k$, which will be determined later. Equation (3.1) says that dealers care linearly about the mean of their expected dividend payment in the end of period 1 and sum they have to pay in both period 1 and 2. They also have to pay an inventory cost which is increasing in the expected variance of the dividend payment. This inventory cost depends on the risk-aversion parameter $\gamma_d > 0$. It is clear that equation (3.1) can be derived from a first-order condition of an exponential utility function. We specifically do not assume exponential utility because an exponential utility function and the functional form specified in equation (3.1) have different implications for the equilibrium on the platform. On the platform, a dealer has to take into account the possibility of being undercut by another dealer when giving quotes to the investor. The model becomes more tractable, if the dealers' utility is linear in the payments made when trading. We will assume that the dealers follow symmetric strategies on the platform and symmetric and linear strategies in the interdealer market.

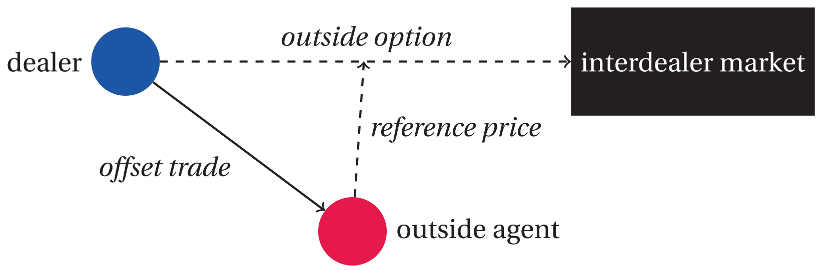
Similar to the dealers, the investor has mean-variance preferences. The investor, however, also receives the private benefit δ per unit of the asset held. If the investor buys a quantity $x_1 \in \mathbb{R}$ on the platform at price p_1, the investor's utility is given by

$$U_I(x_1, p_1) = x_1 \cdot (\theta + \delta) - \frac{\gamma_I}{2} \cdot x_1^2 \cdot \sigma_\varepsilon^2 - p_1 x_1, \tag{3.2}$$

where $\gamma_d > 0$ is the investor's risk-aversion parameter. Comparing (3.1) and (3.2), note that the investor's expectation of the dividend payment and its variance is given by θ and σ_ε, respectively. On the other hand, dealers potentially learn about the dividend from the other agents and thus have a less trivial information set $\mathscr{I}_k$ for each dealer k. Also, dealers trade with each other, which results in a more complex final inventory $\overline{\omega}_k$ and more complex total payments Z_k for each dealer k.

When dealing with the case of one investor in Section 3.4, we need to make a technical assumption in order to keep the model tractable. In Section 3.4, we will assume the presence of an "outside agent". If the investor contacts $M < N, M > 0$ dealers on the platform, these M dealers will learn from the investors about the the realization of θ. Thus, there will be informed and uninformed dealers in the interdealer market. In the interdealer market, the uninformed dealers may then make inferences about the dividend level from the observed market price. It will turn out that this price is affected by both the dealers' inventories and the informed dealers' expectation of the dividend payment. In order to keep this inference problem tractable, we make the dealers' inventory independent of the expected dividend level. To this end, we assume that a dealer who traded on the platform with the investor offsets this trade with the outside agent, who does not participate in the interdealer market. The outside agent does not behave strategically. The price at which the dealer offsets his trade with the investor is such that the dealer is indifferent between trading with the outside agent and going directly to the interdealer market. This way, we keep the dealers' inventories independent of the dividend level and still keep the key economic trade-offs that the dealers and the investor face in our model. This setup is summarized in Figure 3.2. After the trader has offset his trade with the outside agent, all dealers start to trade in the interdealer market. A version of our model without the outside agent will be studied in Section 3.5.

Figure 3.2 – The outside agent



3.4 Equilibrium with one investor

The equilibrium is determined by backward induction. The first step is to establish the equilibrium in the interdealer market. We will assume and later verify that the investor reveals a noisy signal about the dividend level θ to the dealers he contacts. After an equilibrium in the interdealer market has been established, dealers on the platform can anticipate their expected final payoff conditional on the quantity they trade on the platform. This payoff will ultimately be a key determinant of the expected price for the asset on the platform which is derived by standard auction-theoretic arguments. Using the derived quoting strategies of the dealers and assuming that the quantity the investor wants to trade is linear in $\theta + \delta$, an equilibrium on the trading platform can be constructed.

3.4.1 The equilibrium in the interdealer market

The equilibrium in the interdealer market considered in this paper is a rational expectations equilibrium in linear demand schedules as first studied by Grossman and Stiglitz [1980]. This means that dealers behave competitively. Even though not completely realistic, this assumption can be viewed as a rather good approximation in the case of large interdealer markets.

Let x_k denote the quantity of the asset that dealer k buys in the interdealer market. Since we assume that a dealer who trades on the platform offsets his trade with a outside agent, the final inventory $\overline{\omega}_k$ of dealer k is equal to the traded quantity in the interdealer market: $\overline{\omega}_k = q_k$ for all $k \in \{1, ..., N\}$.

Since $M \leq N$ dealers have been contacted on the platform, there will be M informed dealers, who observe $\theta + \delta$ from the investor's demand. The other $N - M$ dealers are uninformed and will use the market price to make inferences about the dividend level. In the following, we will represent all dealers by the set $\{1, ..., N\}$ and say that dealer k is informed if $k \leq M$. Conversely, we say that dealer k is uninformed if $k > M$.

Let p_2 denote the price for the asset in the interdealer market. Differentiating the dealer's utility (3.1) with respect to q_k and using $\frac{\partial Z_k}{\partial q_k} = p_2$ gives the first-order condition

$$\mathbb{E}(D|\mathscr{I}_k) - \gamma_d \overline{\omega}_k \mathbb{V}(D|\mathscr{I}_k) - p_2 = 0. \tag{3.3}$$

Since $\overline{\omega}_k = q_k$, the second order condition is $-\gamma_d \mathbb{V}(D|\mathscr{I}_k) < 0$. The second order condition always holds, since $\gamma_d > 0$ and $\mathbb{V}(D|\mathscr{I}_k) \geq \sigma_\varepsilon^2$. If dealer k receives the signal $s_d := \theta + \delta$, one obtains by standard Bayesian updating that

$$\xi := \mathbb{E}(D|s_d) = \frac{\sigma_\theta^2 s_d}{\sigma_\theta^2 + \sigma_\delta^2}. \tag{3.4}$$

Similarly, one obtains

$$\tau_\xi := \frac{1}{\sigma_\xi^2} := \frac{1}{\mathbb{V}(D|s_d)} = \frac{1}{\frac{\sigma_\theta^2 \sigma_\delta^2}{\sigma_\theta^2 + \sigma_\delta^2} + \sigma_\varepsilon^2}, \tag{3.5}$$

where we defined τ_ξ and σ_ξ^2 as the precision and the variance of the dividend payment based on the informed dealers' information that includes the signal s_d.

The first order condition (3.3) now implies the following demand schedule:

$$q_k = \frac{\tau_\xi(\xi - p_2)}{\gamma_d} \qquad \text{for } k \leq M. \tag{3.6}$$

If dealer k is uninformed, his demand is assumed to be of the form

$$q_k = \frac{\mathbb{E}(D|p_2) - p_2}{\gamma_d \mathbb{V}(D|p_2)} \qquad \text{for } k > M. \tag{3.7}$$

Equation (3.7) takes into account that uninformed dealers can only learn about the conditional distribution of D by observing the market price p_2. We will use the standard approach to conjecture a price that is linear in ξ and the aggregate supply of the asset W:

$$p_2 = a\xi + bW, \tag{3.8}$$

with $a, b \in \mathbb{R}$. Then, uninformed dealers can use the normal projection theorem to calculate $\mathbb{E}(D|p_2)$ and $\mathbb{V}(D|p_2)$.

In equilibrium, also the market clearing condition

$$\sum_{k=1}^{N} q_k = W \tag{3.9}$$

has to be satisfied. Using (3.6) and (3.7) in (3.9) determines the market clearing price. Matching of coefficients in the obtained expression for the market clearing price with the coefficients in the conjectured expression (3.8) then gives the rational expectations equilibrium price function. This price function in turn determines the uninformed dealers' equilibrium demand schedules.

The following Proposition confirms the existence of an equilibrium in the interdealer market and states the corresponding expressions for equilibrium price.

Proposition 1. *There is always a rational expectations equilibrium such that the market clearing price is given by (3.8). Define*

$$\begin{aligned}
\rho : &= \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\delta^2}, && (3.10)\\
\tau_u : &= \frac{1}{Var(D|p_2)} = \frac{1}{\sigma_\theta^2 + \sigma_\epsilon^2 - \psi\rho\sigma_\theta^2}, && (3.11)\\
\psi : &= \frac{a^2\rho\sigma_\theta^2}{a^2\rho\sigma_\theta^2 + b^2\sigma_W^2} = \frac{\rho}{\rho + \frac{\gamma_d^2\sigma_W^2}{M^2\tau_\xi^2\sigma_\theta^2}} && (3.12)
\end{aligned}$$

Then a and b are given by

$$a = \frac{M\tau_\xi + (N-M)\psi\tau_u}{M\tau_\xi + (N-M)\tau_u} \tag{3.13}$$

and

$$b = -\frac{\gamma_d}{M\tau_\xi} a, \tag{3.14}$$

One has $a > 0$ *if* $M > 0$. *One also has* $a \leq 1$ *with strict inequality if* $M < N$.

The fact that $a < 1$ for $M < N$ means tha the price in the interdealer market is inefficient in the sense that the price does not fully reflect the informed dealers' information. In the absence of private benefits for the investor ($\sigma_\delta^2 = 0$), dealers are only willing to trade with the investor because of this informational inefficiency in the interdealer market.

3.4.2 The equilibrium on the trading platform

The equilibrium on the trading platform is derived as follows. We will assume that dealers who are contacted by the investor can observe $s_d = \theta + \delta$ and therefore form a conditional expectation of θ given by ξ as defined in (3.4). We will then use Proposition 1 and the optimal demand schedules (3.6) to determine the lowest price at which a dealer is willing to sell (or the highest price at which he is willing to buy) a given quantity of the asset. The dealers then infer from the investor's utility function the maximum markup they can charge. In equilibrium, dealers will charge a random markup on the platform. The expectation of this price can be used to determine the investor's equilibrium strategy that reveals s_d.

Assume an investor submitted an RFQ to M dealers on the platform to buy x units of the asset (if $x < 0$, the investor wants to sell). If a dealer is contacted on the platform, but does not trade the asset, he will observe s_d and will therefore be informed in the interdealer market, expecting a dividend level of ξ. Let $V_{d,1} : \mathbb{R}^2 \to \mathbb{R}$ denote the function that maps the expectation ξ and dealer k's traded quantity to dealer k's expected utility that he will get after period 2. The dealer will anticipate that the price p_2 is a linear function of ξ and W as stated in Proposition 1. Now, the optimal demand schedule (3.6) and the dealers utility function (3.1) imply the following payoff from not trading (i.e. from trading quantity 0):

$$V_{k,1}(\theta, 0) := \mathbb{E}_k \left[q_k(D - p_2) - \frac{\gamma_d}{2} \sigma_\xi^2 q_k^2 \right] = \frac{\xi^2(1-a)^2 + b^2\sigma_W^2}{2\gamma_d \sigma_\xi^2}.$$

We now consider the case in which dealer k sells quantity x to the investor[4] and goes directly to the interdealer market, while other dealers think that dealer k already offset his trade with the outside agent. Now, dealer k has the initial inventory $-x$ in the beginning of period 2. However, only dealer k knows that.

[4] If $x < 0$ the dealer is buying from the investor.

The following result states the expected price in the interdealer market for dealer k and dealer k's optimal demand.

Lemma 1. *Assume dealer k traded quantity $x \neq 0$ with the investor on the platform and directly goes to the interdealer market. Let the other dealers believe, dealer k offset his trade before going to the interdealer market. Then, according to dealer k's information, the price in the interdealer market is given by*

$$p_2 = a\xi - bx + bW$$

and his optimal demand schedule is given by

$$q_k = \frac{\xi - p_2}{\gamma_d \sigma_\xi^2} + x,$$

where a and b are defined as in Proposition 1.

Using Lemma 1, one can calculate dealer k's expected utility if he goes directly to the interdealer market holding a quantity $-x \neq 0$ and having expectation about the dividend payment ξ.

$$\begin{aligned} V_{k,1}(\xi, x) &:= \mathbb{E}_k \left[D(q_k - x) - p_2 q_k - \frac{\gamma_d}{2} \sigma_\xi^2 (q_k - x)^2 \right] \\ &= \frac{\xi^2 (1-a)^2 + 2(1-a) b \xi x + b^2 \sigma_W^2 + b^2 x^2}{2\gamma_d \sigma_\xi^2} - (a\xi - bx)x. \end{aligned}$$

Comparing $V_{k,1}(\xi, x)$ and $V_{k,1}(\xi, 0)$ one can observe that the dealer expects a different return from holding a final inventory due to a different expected price. The second term in $V_{k,1}(\xi, x)$ represents the additional payment a dealer has to make to offset his inventory x in the interdealer market. We define

$$p_c(x) := \frac{V_{k,1}(\xi, 0) - V_{k,1}(\xi, x)}{x} \qquad k \leq M \tag{3.15}$$

as the break-even price for any contacted dealer k. A dealer who charges $p_c(x)$ per quantity of the asset and sells x units to the investor, does not change his final utility. The payment from the investor exactly matches the difference in utility due to different inventory holdings.

Analogously, we define

$$p_v(x) := \theta + \delta - \frac{\gamma_I}{2} x\sigma_\varepsilon^2 = \frac{\xi(\sigma_\theta^2 + \sigma_\delta^2)}{\sigma_\theta^2} - \frac{\gamma_I}{2} x\sigma_\varepsilon^2 \tag{3.16}$$

as the price at which the investor is indifferent between trading and not trading the asset. As one can immediately verify, equation (3.2) implies $U_I(x, p_v(x)) = 0$. One can interpret $p_c(x)$ as the cost for each contacted dealer of supplying x units of the asset. Analogously, $p_v(x)$ is the investor's value of acquiring x units of the asset. The investor can only trade a certain quantity $x > 0$ with a dealer if $p_v(x) \geq p_c(x)$. Analogously, it has to hold that $p_c(x) \geq p_v(x)$ if $x < 0$.

In the following, we assume that dealers follow symmetric strategies when giving a quote to the investor. This approach is standard, since dealers are ex-ante identical. In the appendix we show that standard search-theoretic arguments imply that the price a dealer quotes on the platform for a certain quantity x has to be a continuous random variable if $p_v(x) \neq p_c(x)$. Let $F_x : \mathbb{R} \to [0, 1]$ denote the distribution of the price a dealer quotes on the platform conditional on the quantity x that the investor wants to trade. If $x > 0$ and $p_v(x) > p_c(x)$, then $p_v(x)$ will turn out to be the supremum of the support of F_x. That quoting a higher price than $p_v(x)$ cannot be optimal follows from $U_I(x, p) < 0$ for $p > p_v(x)$ and $x > 0$. The investor would not be willing to buy the asset at such a price since doing so would make him worse off. Analogously, $p_v(x)$ is the infimum of the support of the distribution of quoted prices if $x < 0$. The investor would not be willing to sell the asset at a lower price.

Dealers are only willing to quote random prices if the expected profit they make is the same for any price in the support of F_x. If $p_v(x)$ is in the support of F_x, this indifference condition means that

$$x(p - p_c(x)) \sum_{j=0}^{M-1} \binom{M-1}{j} (1-q)^{M-1-j} q^j (1 - F_x(p))^j = (1-q)^{M-1} (p_v(x) - p_c(x)) x \tag{3.17}$$

has to hold for all $p \in \text{supp}(F_x)$. The left-hand side of equation (3.17) describes the expected profit a dealer makes by quoting any $p \in \text{supp}(F_x)$. The payment $x(p - p_c(x))$ in excess of the indifference level $xp_c(x)$ is weighted by the probability that the dealer has the best quote among all dealers that respond to the RFQ. Since the response of a dealer is uncertain and occurs with probability $q < 1$, one has to consider the cases in which $j = 0, ..., M-1$ other dealers respond. The right hand side describes the expected profit for a dealer that quotes $p_v(x)$. Since F_x is continuous, this dealer will only sell the asset if no other dealer responds to the RFQ. This happens with probability $(1-q)^{M-1}$. In this case, the dealer's utility will increase by $x(p_v(x) - p_c(x)) > 0$.

The following result gives the closed-form expression for the distribution function F_x that solves (3.17) for any x with $x(p_v(x) - p_c(x)) > 0$. The last inequality is a necessary condition for the existence of strictly positive benefits of trade between dealers on the platform and the investor. In the statement of Lemma 2, we will leave implicit that p_c and p_v depend on x. In Section 3.5, we will study a version of the model in which the dealers' cost p_c does not depend on x. Since Lemma 2 holds irrespective of what variables p_v and p_v depend on, we will state it without reference to any of those variables.

Lemma 2. *Let p_c be the dealers' cost of providing a certain quantity $x \in \mathbb{R} \setminus \{0\}$ of the asset and let p_v denote the investor's value of acquiring x units of the asset. Let the investor submit an RFQ to $M \geq 2$ dealers on the platform to trade quantity x with $x(p_v - p_c) > 0$. Let $q < 1$. Assume that dealers who get contacted know θ.*

If a dealer responds to an RFQ, he will charge a random price that is distributed according to the distribution function F_x. This function is defined by

$$F_x(p) := \frac{1}{q} - \frac{1-q}{q}\left(\frac{p_v - p_c}{p - p_c}\right)^{1/(M-1)}. \tag{3.18}$$

If $x > 0$, the support of F_x is given by $[\overline{p}_x, p_v]$, where $\overline{p}_x$ is determined by $F_x(\overline{p}_x) = 0$ and satisfies $\overline{p}_x > p_c$.

If $x < 0$, the support of F_x is given by $[p_v, \overline{p}_x]$, where $\overline{p}_x$ is again determined by $F_x(\overline{p}_x) = 0$ and satisfies $\overline{p}_x < p_c$.

In this case, the expected price the investor has to pay for the asset conditional on at least one response to the RFQ is given by

$$P(x) := \mathbb{E}(p_1 \mid x, \textit{at least one response}) = \int_{\mathrm{supp}(F_x)} p \, \mathrm{d}G_x(p) = p_c + \kappa(p_v - p_c), \tag{3.19}$$

where the distribution $G_x : \mathbb{R} \to [0,1]$ is defined by

$$G_x(p) := \frac{1 - (1 - qF_x(p))^M}{1 - (1-q)^M}$$

and

$$\kappa := \frac{Mq(1-q)^{M-1}}{1-(1-q)^M} \in [0,1) \tag{3.20}$$

If $q = 1$, Bertrand competition implies that dealers have to set a price that equal to their cost p_c. Thus, the above expression for $P(x)$ holds for all $q \in (0,1]$.

Equation (3.19) states that the expected price the investor receives on the platform is equal to the dealers cost p_c plus a fraction of the total gains from trade $p_v - p_c$. The fraction of this surplus that the investor has to pay is equal to κ, defined as in (3.20). Thus, κ can be viewed as the endogenously determined bargaining power of the dealers. By taking derivatives, it can be shown that κ is decreasing in M and q, which is consistent with economic intuition. As M becomes larger, competition among the dealers for the business of the investor increases. This competition is also higher, if the presence of other dealers on the platform becomes more likely.

Note that the results in Lemma 2 required the assumptions that $x(p_v - p_c) > 0$ and that contacted dealers observe $\theta + \delta$. In the remaining part of this section we will derive an optimal strategy of the investor that allows both assumptions to hold in equilibrium. We will restrict the possible strategies of the investor to strategies that are linear in the sum $\theta + \delta$. This means that the quantity the investor wants to trade is a (positive) multiple of $\theta + \delta$. It is obvious that dealers then can infer $\theta + \delta$ from the quantity the investor wants to trade. However, it is a nontrivial result that the investor finds it indeed optimal to reveal $\theta + \delta$ and the associated information about θ through his choice of the quantity x. The reason why such an equilibrium is possible, even as the private value δ becomes negligible, lies in the fact that the parameter a as defined in 1 is generally less than one. If the investor reveals a given value of $\theta + \delta$ to the dealers, the dealers expect a dividend payment equal to ξ as defined in (3.4). The price for the asset in the interdealer market will be $\xi a < \xi$ in expectation. This price in the interdealer market determines the cost for dealers of providing the asset, which according to Lemma 2 determines the expected price the investor receives on the platform. If $a < 1$, the quotes the investor gets on the platform are less sensitive to θ than the investor's utility. This makes an equilibrium possible in which the investor partially reveals his information θ to the dealers.

We now conjecture that the investor's demand for the asset on the platform is given by

$$x = \alpha(\theta + \delta), \tag{3.21}$$

for some $\alpha \in \mathbb{R}$. In the appendix we show that the expected price $P(x)$ from Proposition 2 is linear in x and ξ:

$$P(x) = \beta_1 \xi + \beta_2 x, \tag{3.22}$$

with $\beta_1, \beta_2 \in \mathbb{R}$. From the investor's conjectured strategy (3.21), the contacted dealers infer

$\theta + \delta = \frac{x}{\alpha}$. Using (3.2), (3.22) and (3.4), the investor's problem therefore becomes

$$\max_{x \in \mathbb{R}} \left[(\theta+\delta)x - x^2 \frac{\gamma_I}{2}\sigma_\varepsilon^2 - x\left(\beta_1 \frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\delta^2}\frac{x}{\alpha} + \beta_2 x\right)\right]. \tag{3.23}$$

Note that (3.23) considers the investor's expected payoff conditional on at least one response to the RFQ. Since the probability of this event is exogenous and always gives a zero payoff, it can be neglected. The first-order condition for (3.23) implies the investor's optimal demand schedule

$$x = (\theta+\delta)\frac{\alpha(\sigma_\theta^2+\sigma_\delta^2)}{2\alpha\beta_2(\sigma_\theta^2+\sigma_\delta^2)+\alpha\gamma_I\sigma_\varepsilon^2(\sigma_\theta^2+\sigma_\delta^2)+2\beta_1\sigma_\theta^2}. \tag{3.24}$$

Therefore, the investor's optimal demand is indeed linear in $\theta+\delta$. Matching the coefficient in (3.24) with the conjectured strategy (3.21) gives

$$\alpha = \frac{\sigma_\theta^2+\sigma_\delta^2-2\beta_1\sigma_\theta^2}{(2\beta_2+\gamma_I\sigma_\varepsilon^2)(\sigma_\theta^2+\sigma_\delta^2)}. \tag{3.25}$$

The following proposition summarizes these results and states formal conditions under which the equilibrium exists.

Proposition 2. *The expected price on the platform $P(x)$ from Lemma 2 is linear in ξ and x, as stated in (3.22). Let $M \geq 2$. If*

$$\kappa < \frac{1}{2}, \tag{3.26}$$

with κ as in (3.20), there is a threshold $\overline{a} > 0$, such that the equilibrium on the platform described below exists if and only if $a < \overline{a}$. The last condition holds as $N \to \infty$ and $\sigma_W^2 \to \infty$ or as $\sigma_\delta \to \infty$. The inequality in (3.26) will always hold for all $M \geq 2, \sigma_\theta^2, \sigma_\delta^2 > 0$ if $q \to 1$. If (3.26) does not hold, the equilibrium does not exist.

The equilibrium is characterized as follows. The investor submits a demand x as determined in equations (3.21) and (3.25). The dealers quote independently with probability q according to the distribution function F_x in (3.18).

One furthermore has $0 < \beta_1 < \frac{1}{2}\frac{\sigma_\theta^2+\sigma_\delta^2}{\sigma_\theta^2}$, $\beta_2 > -\frac{\gamma_I}{2}\sigma_\varepsilon^2$ and $\alpha > 0$ in each such equilibrium.

With the RFQ trading protocol, an equilibrium with linear strategies described in Proposition

2 is possible even though a linear equilibrium in double auctions and two strategic traders would not exist due to correlated values as Du and Zhu [2017] show. With the RFQ trading protocol, only the investor has the option to avoid price impact by reducing his demand. The dealers have to take the traded quantity as given and can merely charge a markup in addition to their cost of providing the asset.

We will illustrate the results derived so far with an example.

3.4.3 A brief example

For illustrative purposes we fix the exogenous parameters as follows: $N = 100$, $M = 10$, $\sigma_\varepsilon = 1$, $\sigma_W = N$, $\sigma_\theta = 1$, $\gamma_d = 1$, $\gamma_I = 1$, $q = 0.3$. To illustrate the economic mechanism of our model, we first consider the case in which $\theta + \delta = 1$, which corresponds to a realization one standard deviation above the mean. Afterwards we consider the case when $\theta + \delta = -1$. It is sufficient to only consider the sum of the common value and the investor's private, since both the investor's demand and the dealars' inferences depend only on this sum.

In Figure 3.3, $\theta + \delta$ has the high realization. In Panel (a) we plot the price $p_v(x)$ that the investor is willing to pay for x units of the asset. If the absolute value of x is small, this price is approximately equal to $\theta + \delta$, since the cost of bearing risk is small. The price $p_v(x)$ is linearly decreasing in x because of the quadratic cost of bearing risk. We also plot the dealer's cost $p_c(x)$ of providing x units of the asset, if they believe the dividend payment is normally distributed with mean ξ and precision σ_ξ^2, as defined in (3.4) and (3.5). One can see that this cost is slightly increasing in x, which represents the difficulty of offsetting the trade in the interdealer market or with the outside agent, respectively. The average price the investor can expect conditional on at least one response to the RFQ, $P(x)$, is between the other two curves.

In Panel (b) of Figure 3.3 we keep the investor's reservation price $p_v(x)$, but now look at the average price dealers quote when they the expected dividend level ξ from the investors demand x one can see that this price increases faster in x than the cost $p_c(x)$ in Panel (a). We also plot the profit the investor investor gets for demanding a certain quantity x. This profit is the solution to problem (3.23) weighted by the probability of at least one response to the RFQ. We see that the optimum is approximately at $x = 0.69$. This also turns out to be the value of α. Thus, Panel (b) illustrates, that the investor has indeed no incentive to deviate from the equilibrium strategy determined in the last section.

In Figure 3.4, we consider the low realization of $\theta + \delta$. Comparing panel (a) to Panel (a) in Figure 3.3, we observe that all curves have been shifted downwards by a constant. The curve of $p_v(x)$ has been shifted downwards more than the curve of $p_c(x)$. This has two reasons. First, the dealers expectation ξ is a weighted average between $\theta + \delta$ and zero, as (3.4) shows. Second, the dealers do not find it as costly to hold a bad asset as the investor does. The dealers expect to be able to resell the asset again at a favourable price, since there are many uninformed dealers in the interdealer market. Panel (b) of Figure 3.4 shows a similar picture as Panel (b) of

Figure 3.3 – High realization of $\theta + \delta$

(a) Expected and reservation prices if ξ is held fixed. (b) Dealers infer ξ from the investor's demand.

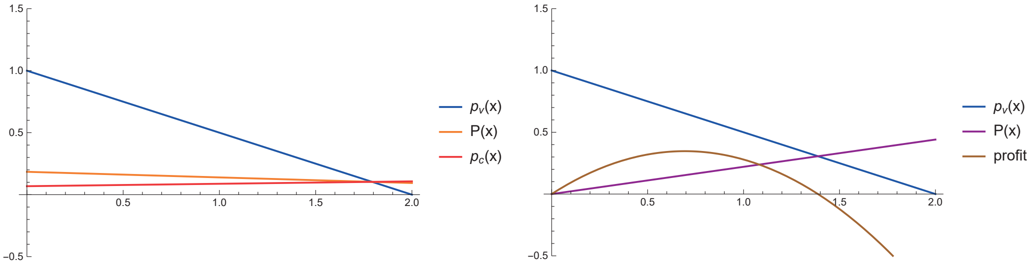


Figure 3.4 – Low realization of $\theta + \delta$

(a) Expected and reservation prices if ξ is held fixed. (b) Dealers infer ξ from the investor's demand.

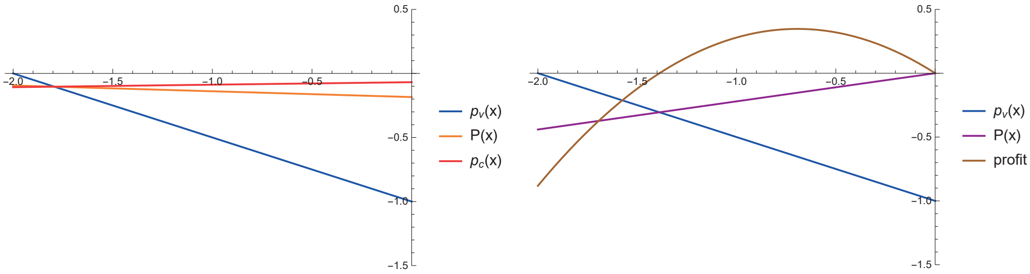


Figure 3.3. In Figure 3.4, however, the investor sells the asset at a negative expected price. The investor finds it profitable to do so, since $p_v(x)$ indicates that he would be willing to sell the asset at an even lower price due to the negative expected dividend. The equilibrium strategies have not changed in Figure 3.3 and Figure 3.4. Therefore, the optimal demand in Figure 3.4 is the negative of the optimal demand in Figure 3.3, since the respective realizations of $\theta + \delta$ have the same absolute value in both cases.

3.4.4 Competition vs. information leakage

In this section we take a closer look at the equilibrium described in Proposition 2. Specifically, we take a look how the investor's profits from trading on the platform are affected by varying the number of dealers who are contacted on the trading platform.

We define π_I as the investor's ex-ante expected payoff in the equilibrium described in Proposition 2. By the investor's utility function (3.2), his equilibrium strategy (3.25) and (3.22), one has

$$\pi_I = \mathbb{E}\left[(1-(1-q)^M)(\theta+\delta)^2 \frac{1}{2}\alpha\right]. \tag{3.27}$$

Equation (3.27) takes into account that the investor does not receive any quote with probability $(1-q)^M$ and that dealers infer ξ from the investor's demand.

Our first goal is to study the role of M, the number of recipients of each RFQ. Increasing M has three major effects that determine the investor's profit:

- As is evident from (3.27) a higher M increases the probability of a trade $1-(1-q)^M$, whenever $q<1$. Holding everything else equal, this increases expected profits.
- A higher M increases the fraction of informed dealers in the interdealer market. One can verify that a as defined in (3.13) is strictly increasing in M for $M<N$. This makes prices in the interdealer market more informative and it therefore becomes more difficult to offset any inventory that was acquired on the platform.
- A higher M decreases κ, as mentioned in the discussion after Lemma 2. Therefore, the bargaining power of the investor increases, which has a positive effect on his profit.

Considering these three bullet points, the investor's profit should be maximal for $M=2$, if $q=1$. If $q=1$, one has $\kappa=0$ and $1-(1-q)^M=1$, i.e. the investor's bargaining power is maximal and a trade happens with probability 1. Then the first and third bullet point above become irrelevant and increasing M is only associated with the cost of information leakage, discussed in the second bullet point. The following proposition formally confirms that the conclusion of the above heuristic reasoning is indeed true. Since we focus on the cost of information leakage we assume for better algebraic tractability that there are no private benefits, i.e. $\sigma_\delta=0$.

Proposition 3. *Let $2\le M$ and $q=1$ and $\sigma_\delta=0$. The equilibrium described in Proposition 2 exists if and only if a is below a certain threshold $\overline{a}$, with $\overline{a}<\frac{1}{2}$. In this equilibrium, one has $\beta_1,\beta_2,\alpha>0$.*

Furthermore, the equilibrium exists for any other choice of the number M' of dealers to contact with $2\le M'<M$. If $M=2$, the payoff for the investor is higher than in any other possible equilibrium with $M>2$.

When $q<1$, the investor has incentive to contact more dealers, i.e. $M\ge 2$. Because when $q\neq 1$, the first and third effects turn out to be relevant: increasing M will improve the probability of trading , as well as the bargaining power of the investor. But at the same time, the cost of information leakage is also increased (second bullet point).

The following proposition states that M sometimes has to be larger than a certain threshold in order for an equilibrium to exist in the first place. If q is relatively small, the bargaining power κ of dealers may be so high that investors do not want to incur any price impact they have on the trading platform. Increasing M lowers this bargaining power. Under the condition that prices in the interdealer market remain sufficiently uninformative, an equilibrium exists for a sufficiently large M. On the other hand, there is a clear upper bound on the possible number of dealers that are contacted on the platform for which an equilibrium exists. In particular, if more than half of the dealers are contacted and there is strong asymmetric information about the asset's payoff ($\sigma_\delta = 0$), an equilibrium cannot exist, because information leakage on the platform is too strong.

Proposition 4. *Let $\sigma_\delta^2 = 0$. If $M > \frac{1}{2}N$, there is no equilibrium on the trading platform as described in Proposition 2.*

If $q < \frac{1+\sqrt{1-\frac{2\left(a(2)^2-5a(2)+2\right)}{a(2)^2-5a(2)+4}}}{2}$, there is no such equilibrium with $M = 2$. If furthermore $a < \bar{a}$, for an $\bar{a} \in (0, \frac{1}{2})$, then there is such an equilibrium with $M \geq 3$.

3.4.5 Price impact

In this section we want to relate our theoretical results to the empirical findings of Collin-Dufresne et al. [2017]. In particular, we want to study the price impact that an investor faces on the trading platfrom and the price impact that dealers face in the interdealer market. The total price impact a trader faces can be decomposed as

$$\text{price impact} = \text{permanent impact} + \text{transitory impact}.$$

Collin-Dufresne et al. [2017] find that price impact in the D2C segment is higher than in the D2D segment. This difference is largely due to a differene in the permanent price impact.

We now want to derive the price impact and find analogues in our model that correspond to a permanent component and a transitory component. As commonly argued in theoretical studies [Sannikov and Skrzypacz, 2016, Kyle et al., 2017], the study of price impact is an off-equilibrium analysis. We will therefore assume an equilibrium as described in Proposition 2 and examine how the price a trader faces changes if the demanded quantity changes.

Equation (3.19) in Lemma 2 directly provides an expression of the expected price an investor receives on the platform. If the investor changes his demanded quantity x, then p_v and p_c in (3.19) and consequently the expected price for this quantity will change. Since the model presented in this paper is static, we have to find a decomposition of this price impact that would correspond to a decomposition into a permanent and a transitory component in a dynamic model. In empirical studies in Market Microstructure, it is generally assumed that the transitory component reflects a markup of the dealers, whereas the permanent component reflects the cost of the dealers of providing the asset due to future price changes. In our following analysis, we adopt this interpretation. We say that the price impact is permanent, if it was caused by a change in the dealers' cost of providing the asset.[5] Therefore, we define

$$PI := \frac{\partial}{\partial x} p_c(x) \tag{3.28}$$

as the permanent price impact of the investor, because (3.28) reflects the change in the price that is due to an increase in the dealers' cost of trading the asset.

In the following, we will consider p_c as defined in (3.15). Due to adverse selection, we also need to take into account that the dealers form their expectation ξ about the dividend payment based on (3.4) and (3.21). The following proposition contains some statements about the price impact on the platform and in the interdealer market.

Proposition 5. *The (permanent) price impact an informed dealer faces in the interdealer market is given by $-b$, where b is defined as in Proposition 1. Without adverse selection (dealers do not update their belief ξ), one has*

$$PI < -b,$$

i.e. the permanent impact on the trading platform is smaller then the permanent impact the dealers face in the interdealer market. In the presence of adverse selection and $\rho = \frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\delta^2} > 1/4$, one has

$$PI > -b.$$

The dealers' permanent price impact $-b$ derived in Proposition 5 is due to a change in the uninformed dealers' belief about the dividend payment and a permanent change in the aggregate inventory held by other dealers in the interdealer market. Proposition 5 shows that the permanent price impact on the trading platform higher than the permanent impact in the interdealer market if and only if investors know more about the asset than dealers do. If there is no adverse selection and dealers do not update their belief about the asset's payoff, the dealers' cost of providing the asset changes by a lower rate than the price in the interdealer market would when trading the same quantity. This result is due to the dealers' optimal portfolio choice in period 2. A dealer could always offset the investor's demand in the interdealer market with price impact $-b$. If the investor however changed his demanded quantity, the dealer would, due to risk sharing considerations, in general not offset the total amount of this quantity in the interdealer market. Due to optimality of the dealer's portfolio choice, the dealer must be able to provide the quantity at a lower price than the one he would pay for this quantity in the interdealer market.

In the presence of adverse selection, however, the permanent price impact on the trading

[5] This assumes that changes in the dealers' cost have no transitory component. Transitory changes in the dealers' cost may arise due to inventory holding costs or order processing costs. In our model, dealers can immediately offset their inventory and the interdealer market is competitive. Order processing is costless. Therefore, such transitory components of dealers' costs are not present in our model.

platform is higher than the permanent price impact the dealers face in the interdealer market. This makes our model (which assumes information asymmetries) consistent with the findings of Collin-Dufresne et al. [2017] that the permanent price impact is higher on the trading platform than in the interdealer market.

3.5 Centralized trading vs. electronic trading via RFQs

This section develops the market-design implications of our model. Our final goal is to characterize situations in which investors are better off trading in a centralized market and when an OTC market can improve their utility. In order to do this, we extend our previous model from to the case in which there is a is a continuum of investors of measure μ[6] who all know the realization of θ in the beginning of period 1. The investors' utility function is still give by (3.2). The risk-aversion parameter γ_I is the same for all investors. The investors receive a private benefit δ_i, where as before $\delta_i \sim \mathcal{N}(0, \sigma_\delta^2)$. The private benefits for different investors are essentially pairwise independent for different investors. This assumptions lets us apply the exact law of large numbers of Sun [2006]. The model assumptions about the dealers are as in Section 3.3, except that we do not assume the presence of an outside agent in this section. Before we establish an equilibrium in the OTC market, we quickly describe how the investors would trade in a centralized market.

3.5.1 The centralized-market benchmark

Investors trade through double auctions in the centralized market. In these double auctions, each investor specifies a demand schedule, i.e. conditional on each price $p \in \mathbb{R}$ the investor specifies a quantity he wants to trade. The equilibrium price in the centralized market will be the market-clearing price. The market clearing price will be the unique price for which the investors' aggregate demand is equal to the aggregate supply of the asset (zero). The specification of the investor's utility function (3.2) gives the following maximization problem for each investor for each $p \in \mathbb{R}$:

$$\max_{x_i \in \mathbb{R}} \left[x_i(\theta + \delta_i - p) - \frac{\gamma_I \sigma_\epsilon^2}{2} x_i^2 \right],$$

where x_i denotes the quantity the investor demands given the price p on the exchange.

The sufficient first-order condition for the above optimization problem gives

$$x_i = \frac{\theta + \delta_i - p}{\gamma_I \sigma_\epsilon^2},$$

[6]Formally, let $(\Omega, \mathscr{F})$ denote the measurable space of investors. Then there is a bijective measurable map $\Phi : \Omega \to [0, \mu]$ and the measure of any set of investors $F \in \mathscr{F}$ is equal to the Lebesgue measure of the set $\Phi(F)$.

To determine the market-clearing price, we substitute each investor's demand schedule x_i into the market clearing condition, $\int x_i \mathrm{d}i = 0$.[7] We get

$$0 \;=\; \mu\frac{\theta}{\gamma_I\sigma_\epsilon^2} - \mu\frac{p}{\gamma_I\sigma_\epsilon^2} \Leftrightarrow p = \theta\,,$$

where we have used the fact that $\int \delta_i \mathrm{d}i = 0$ almost surely by the exact law of large numbers.

Using each investor's optimal demand schedule, the utility function (3.2) and the fact that the market clearing price is given by θ, we can define each investor's ex-ante payoff:

$$\pi_i^c \;:= \mathbb{E}\left(\frac{1}{2}\frac{(\theta+\delta_i-p)^2}{\gamma_I\sigma_\epsilon^2}\right) = \frac{1}{2}\frac{\sigma_\delta^2}{\gamma_I\sigma_\epsilon^2}\,. \tag{3.29}$$

Equation (3.29) states that the centralized market realizes all the gains from trade that arise due to dispersed private values. When all investors have the same valuation of the asset ($\sigma_\delta^2 = 0$), no trade happens and the investor's profits become zero. Each investor's profit decreases if the cost of bearing risk increases.

3.5.2 Electronic trading with a continuum of investors

The model with a continuum of investors is very similar to the model with one investor. It will turn out that a continuum of investors allows us to derive an equilibrium without the assumption of an outside agent. We will let the mass of investors have a measure $\mu \in (0,\infty)$. In period 1, all investors submit RFQs to M dealers. Afterwards, dealers trade in the interdealer market. All investors contact the same M dealers at the same time. The dealers then independently respond with a probability q to each RFQ. As before, we will determine the equilibrium in this model by backward induction.

Since there is no outside agent anymore in this section, uninformed dealers in the interdealer market take into account that the aggregate supply of the asset is correlated with the investors' information about the dividend level θ. We will conjecture that each investor demands a quantity x_i on the trading platform, where

$$x_i = \alpha_1\theta + \alpha_2\delta_i\,, \tag{3.30}$$

for some $\alpha_1, \alpha_2 \in \mathbb{R}$. As in Section 3.4, it will turn out that an investor always trades the asset if he receives a quote on the trading platform. Since each dealer responds independently with probability q to each RFQ, the an investor is able to trade the asset with probability $\mathbb{P}(trade) = 1-(1-q)^M$. By the exact law of large numbers and (3.30), the investors' aggregate

[7] The notation di means that we integrate with respect to the measure on set of investors defined in Footnote 6.

demand traded on the platform given by

$$X^{agg} := \int \mathbb{P}(trade)(\alpha_1\theta+\alpha_2\delta_i)\mathrm{d}i = (1-(1-q)^M)\int(\alpha_1\theta+\alpha_2\delta_i)\mathrm{d}i = (1-(1-q)^M)\mu\alpha_1\theta, \quad (3.31)$$

where the last equality holds almost surely. By symmetry, each dealer gets an equal fraction of this aggregate demand. We define $X_k := -\frac{X^{agg}}{M}$ as the inventory of each dealer $k \leq M$ who gets contacted on the trading platform. From the dealers' utility function (3.1), one obtains the optimal demand schedule q_k for each dealer $k \leq M$:

$$q_k = \frac{\theta - p_2}{\gamma_d \sigma_\varepsilon^2} + X_K. \quad (3.32)$$

Notice that X_k is a multiple of θ, this will simplify the inference problem that the uninformed dealers face in the interdealer market. Analogously to Section 3.4, we conjecture that the market-clearing price in the interdealer market is given by

$$p_2 = a\theta + bW, \quad (3.33)$$

where W is the noise in the aggregate supply of the asset. The uninformed dealers use the normal projection theorem obtain the distribution of the dividend payment conditional on the market-clearing price p_2. The dealers' utility function (3.1) now gives the optimal demand

$$q_k = \frac{\mathbb{E}(D|p_2) - p_2}{\mathbb{V}(D|p_2)} \quad (3.34)$$

for the uniformed dealers who do not get contacted on the trading platform. Analogously to Proposition 1, we now state the equilibrium in the interdealer market, conditional on the investors' trading strategy (3.30).

Proposition 6. *For any given α_1, there is a rational expectations equilibrium such that the market clearing price is given by (3.33). Define*

$$\begin{aligned}
\varphi : &= (1-(1-q)^M)\mu, && (3.35)\\
\tau_u : &= \frac{1}{Var(D|p_2)} = \frac{1}{\sigma_\theta^2+\sigma_\epsilon^2-\psi\sigma_\theta^2}, && (3.36)\\
\psi : &= \frac{a^2\sigma_\theta^2}{a^2\sigma_\theta^2+b^2\sigma_W^2} = \frac{\sigma_\theta^2}{\sigma_\theta^2+\left(\frac{\gamma_d}{M\tau_\epsilon+\gamma_d\varphi\alpha_1}\right)^2\sigma_W^2} && (3.37)
\end{aligned}$$

Then a and b are given by

$$a = \frac{M\tau_\epsilon+\gamma_d\alpha_1\varphi+(N-M)\psi\tau_u}{M\tau_\epsilon+(N-M)\tau_u} \tag{3.38}$$

and

$$b = -\frac{\gamma_d}{M\tau_\epsilon+\gamma_d\varphi\alpha_1}a, \tag{3.39}$$

One has $a > 0$. One also has $a \leq 1$ with a strict inequality if $M < N$.

Lemma 2 gives the dealers optimal quoting strategy for any aggregate quantity X_k that dealer k trades with the investors and any demand x_i they face from an individual investor i. There is only a slight difference between the case in Section 3.4 and the setup considered here. Whereas the aggregate demand a dealer faced was equal to the demand by the single investor in Section 3.4, the quantities x_i and X_k are different here. Using Lemma 2 and taking account of this difference gives the expected price $P(x_i)$, investor i gets for his demand x_i conditional on at least one response to the RFQ:

$$P(x_i) = p_c(X_k)+(p_v(x_i)-p_c(X_k))\frac{Mq(1-q)^{M-1}}{1-(1-q)^M}. \tag{3.40}$$

We already determined in (3.30) which form each investor's demand x_i takes. We also know the quantity X_k given these individual demand schedules. We now determine the values of $p_v(x_i)$ and $p_c(X_k)$, so that we can use (3.40) to determine the expected price that each investor faces for his demand.

We define this dealer's value function that maps his inventory after period 1 X_k to expected utility as

$$V_{k,1}(\theta, X_k) := \mathbb{E}_k\left[D(q_k - X_k) - p_2 q_k - \frac{\gamma_d}{2}\sigma_\varepsilon^2 (q_k - X_k)^2\right], \tag{3.41}$$

where we used the dealer's utility function (3.1).

Having obtained the dealer's utility $V_{k,1}(\theta, X_k)$ when holding X_k units of the asset, we define the dealer's break even price $p_c(X)$ such the payment compensates for the marginal cost of holding an additional marginal unit of the asset. The resulting expression is stated in the following Lemma.

Lemma 3. *Conditional on the equilibrium inventory of dealer $k \leq M$, the dealer's equilibrium break-even price for for the asset is given by*

$$p_c(X_k) := -\frac{\partial}{\partial X_k} V_{k,1}(\xi, X) = a\theta - b\frac{(1-a)\theta}{\gamma_d \sigma_\varepsilon^2} + b\frac{\varphi}{M}\alpha_1\theta, \tag{3.42}$$

where a, b, φ are defined as in Proposition 6.

Given that a dealer inferred the realization of θ from the investors' demand, a dealer can infer the private value δ_i of investor i from this investor's individual demand and (3.30). Given the investor's demand for x_i units of the asset, a dealer can infer the maximum price the investor is willing to pay for these x_i units by using (3.2):

$$p_v(x_i) := \theta + \delta_i - \frac{\gamma_I}{2} x \sigma_\varepsilon^2. \tag{3.43}$$

Using (3.31), (3.42) and (3.42), one can rewrite (3.40) as

$$P(x_i) = \beta_1\theta + \beta_2 x_i, \tag{3.44}$$

for some $\beta_1, \beta_2 \in \mathbb{R}$ stated in the appendix. We now determine the optimal amount x_i that an investor wants to demand given that the expected price he faces on the platform is given by

(3.44). The maximization problem of investor i is given by

$$\max_{x\in\mathbb{R}}\left[(\theta+\delta_i)x_i - x^2\frac{\gamma_I}{2}\sigma_\varepsilon^2 - x_i\left(\beta_1\theta+\beta_2 x_i\right)\right]. \tag{3.45}$$

The expression in (3.45) considers the investor's expected payoff conditional on at least one response to the RFQ, since the investor's payoff is maximized when his payoff conditional on at least one response is maximized. The first-order condition to the problem in (3.45) gives the investor's optimal demand schedule

$$x = \theta\frac{1-\beta_1}{2\beta_2+\gamma_I\sigma_\varepsilon^2} + \delta\frac{1}{2\beta_2+\gamma_I\sigma_\varepsilon^2}. \tag{3.46}$$

One can immediately determine α_1 and α_2 from (3.30) by looking at (3.46):

$$\alpha_1 = \frac{1-\beta_1}{2\beta_2+\gamma_I\sigma_\varepsilon^2}, \tag{3.47}$$

$$\alpha_2 = \frac{1}{2\beta_2+\gamma_I\sigma_\varepsilon^2}. \tag{3.48}$$

We are now ready to establish the existence of an equilibrium.

Proposition 7. *The expected price on the platform $P(x_i)$ that an investor gets on the platform for his demand x_i is given by (3.44) for some $\beta_1, \beta_2 \in \mathbb{R}$. Let $M \geq 2$. There is an equilibrium on the platform described below if and only if $M < N$ and*

$$\kappa = \frac{Mq(1-q)^{M-1}}{1-(1-q)^M} < \frac{1}{2}.$$

The equilibrium is characterized as follows. The investor submits a demand x_i as determined in equations (3.30) with $\alpha_1, \alpha_2 \in \mathbb{R}$, with $0 < \alpha_1 < 1$ and $\alpha_1 \leq \alpha_2$. The dealers quote independently with probability q according to the distribution function F in (3.18) with $p_c(X_k)$ and $p_v(x_i)$ given by (3.42) and (3.43).

3.5.3 Market design

In this section we will use the results derived in Section 3.5.1 and Section 3.5.2 and study when investors prefer the centralized market and when they prefer the OTC market with an electronic trading platform. Proposition 7 states that there cannot be an equilibrium on the electronic trading platform if $\kappa \geq \frac{1}{2}$ or $N = M$. In this case, there is only an equilibrium in the centralized market. Therefore, we restrict our further discussion to the case in which $\kappa < \frac{1}{2}$ and $M < N$. The following claim follows from (3.29) and Proposition 7.

Proposition 8. *Let* $0 < \kappa < \frac{1}{2}$ *and* $2 \leq M < N$. *As* $\sigma_\delta^2 \to 0$, *investors prefer to trade in on the trading platform in the OTC market. As* $\sigma_\delta^2 \to \infty$, *investors prefer to trade in the centralized market.*

If $\sigma_\delta^2 \to 0$, equation (3.29) implies that investors' gains from trading in the centralized market go to zero. However, due to information asymmetries between dealers and investors, investors can still benefit from trading in the OTC market.

Suppose on the other hand, that $\sigma_\delta^2 > 0$ and the mass of investors μ becomes very large. Then holding everything else constant, the investors' demand will be very sensitive to variations in θ. In this case, an equilibrium is only possible if α_1, the coefficient in the investors' demand on θ is very small and investors will mainly trade based on their private value of holding the asset. If markups in the interdealer market are positive, investors will therefore prefer to trade in the centralized market instead. The following proposition proofs this statement formally.

Proposition 9. *Let* $0 < \kappa < \frac{1}{2}$, $2 \leq M < N$ *and* $\sigma_\delta^2 > 0$. *As* $\mu \to \infty$, *investors prefer to trade in the centralized market.*

The proof of Proposition 9 shows that $\alpha_1 \to 0$ as $\mu \to 0$. According to (3.47), this is equivalent to $\beta_1 \to 1$, holding everything else equal and noting that by (C.27), β_2 is unaffected by μ. Thus, (3.44) implies that the expected price an investor receives on the platform when $\mu \to \infty$ is approximately the sum of the common value θ of the dividend payment and a markup. In this case, the investors' gains from trade are derived mostly from their private values.

So far, we assumed that $\kappa > 0$, which lead to positive expected markups for the dealers when quoting on the trading platform. In the following we consider the case in which $q \to 1$, which leads to $\kappa \to 0$. If $\kappa \to 0$, these markups become negligible and dealers efficiently intermediate trades between their customers as if these customers were trading in a centralized market. Furthermore, the probability of not receiving a quote goes to zero as $q \to 1$. Thus, all the gains from trade that could be realized in the centralized market would also be realized in the OTC market. However, investors can still benefit from information asymmetries between them and the dealers in the OTC market. As $q \to 1$ investors therefore prefer to trade in the OTC market. This claim is formally proved in the next proposition.

Proposition 10. *Let* $2 \leq M < N$. *As* $q \to 1$, *investors prefer to trade on the trading platform.*

3.6 Discussion and concluding remarks

Electronic trading platforms play a central role in today's OTC markets. The implications of our model are consistent with recent empirical research that studies OTC markets with electronic trading platforms. One important feature of our model is information leakage which is studied in Hendershott and Madhavan [2015] and Hagströmer and Menkveld [2016]. We also showed that information asymmetries between dealers and investors are a sufficient and necessary condition to generate the price impact patterns observed in Collin-Dufresne et al. [2017]. Therefore, the first part of this paper can be viewed as a theoretical foundation of several empirical findings in recent research. The model can also be used to evaluate the impact of recent financial regulation on investors' trading profits. The Dodd-Frank Act mandates that the most liquid index CDS in the US are trades on electronic platforms. An RFQ furthermore should be sent to at least three dealers.[8] We show that increasing the number of contacted dealers may decrease investor's profits if the cost of information-leakage is high. On the other hand, the number of contacted dealers has to be sufficiently high in order for an equilibrium to exist, if competition among dealers on the platform (in terms of response rates) is low.

In the second part of the paper, we considered a hypothetical scenario in which there is either a centralized exchange or an OTC market and studied the respective implications on investor welfare. Some of our results are consistent with the recent theoretical literature in the area of market design. That investor welfare is generally higher on exchanges if the investors associate strong private values with holding the asset, can be viewed as an analogue to the result of Babus and Parlatore [2017] that there is only a centralized-market equilibrium if the investors' values of holding the asset are sufficiently independent. We also emphasize the role of information asymmetries that becomes important in OTC markets. In this respect our paper is related to Glode and Opp [2017]. However, the specific trading protocol on electronic trading platforms features some aspects that are not present in other models of OTC markets. As the RFQ response rate q of dealers becomes high, our model shows that electronic trading platforms indeed become similar to exchanges, in the sense that dealers efficiently intermediate the demand from their customers. This result justifies the common opinion that electronic trading platforms represent a natural compromise between exchanges and OTC markets.[9]

To conclude, we want to make some general remarks on our model assumptions. As every theoretical model, also the one presented in this paper is build on some simplifying assumptions trading-off analytical tractability against appropriate representation of the real world. The fact that all investors are equally informed about the asset's payoff is certainly not completely realistic, but should capture the general information asymmetry between investors and dealers that in many markets seem to exist. To justify the way we model trading in the interdealer market, we want to refer to the event that made both the academic world and international regulatory authorities focus so much on OTC markets in the first place: the recent financial crisis. Arguably, demand for certain credit derivatives originated from informed hedge funds who wanted to bet against a credit bubble in the US credit market. Some investment banks may have learned about the value of certain securities from this informed demand and may have tried to use this knowledge against other less informed investment banks or other clients

[8]See Collin-Dufresne et al. [2017] for an overview of the regulatory changes in the US CDS market.
[9]See Stafford [2016].

(which may be represented by noise traders in our model).

This example also suggests to interpret the welfare results derived from our model with a slight grain of salt. In this paper, we exclusively focused on investor welfare. While this approach may be viewed as standard in market design, it does not take into account financial stability considerations that may be important when determining the optimal level of transparency in the market. If losses to dealers or noise traders are large, the financial system may very well be affected in ways that cannot be captured in the model presented here. While the trade-off between the efficient allocation of assets and financial stability is a common theme in banking, examining the trade-off between investor welfare and financial stability in OTC markets may be a theme for future research.

A Appendix to Chapter 1

A.1 Micro-Found the Trading Protocol

In this section, I micro-found the trading protocol by incorporating the fact that in practice, customers have uncertainty regarding the degree of competition in the market. Specifically, I split the D2C trading round into two sub periods. In the first sub-period, customer can decide whether to direct the order to a dealer (monopolistic pricing), or start a flash auction with uncertainty on the number of participants. If customer decides to have a flash auction, then with probability $1-\theta$, the auction is competitive, and customer's order is able to trade at the centralized inter-dealer pricing kernel M^{D2D}. However, with probability θ, there is only one response and customer has to trade at the monopolistic pricing kernel $\hat{M}^{\text{D2C}}_{(\text{i,j})}$. Hence, ex-ante, customer's utility from a flash auction is given by,

$$(1-\theta)\nu_{(\text{i,j})}[M^{\text{D2D}}] + \theta\nu_{(\text{i,j})}[\hat{M}^{\text{D2C}}_{(\text{i,j})}] \, .$$

Now, in sub-period one, the dealer then quotes a pricing kernel, $M^{\text{D2D}}_{(\text{i,j})}$, such that customer breaks even between a directed order and a flash auction. I assume that in case of break-even, customer trades with the dealer using directed orders. The dealer's quoting problem is again a monopolistic pricing rule, and he maximizes his indirect utility, subject to the customer's participation constraint

$$\nu_{(\text{i,j})}[M^{\text{D2C}}_{(\text{i,j})}] \;\geq\; (1-\theta)\nu_{(\text{i,j})}[M^{\text{D2D}}] + \theta\nu_{(i,j)}[\hat{M}^{\text{D2C}}_{(\text{i,j})}] \, ,$$

as well as the two no-arbitrage constraints. Interestingly, the first-order condition of this maximization problem coincides with the optimal condition in the relaxed problem (1.12), with $\pi_{(\text{i,j})}$ being endogenized as the Lagrange multiplier of the customer's participation constraint.

A.2 Expansion

I use log-utility agents as an example. For other utility functions, the steps are similar.

Definition 4. *The linear operator on a nonlinear risk function F is defined as*

$$\mathscr{M}[F] \equiv M^{(0)}\left(M^{(0)}F - E[M^{(0)}F] - \left(M^{(0)} - E[M^{(0)}]\right)\frac{\text{Cov}[M^{(0)}, M^{(0)}F]}{\text{Var}[M^{(0)}]}\right) = M^{(0)}\varepsilon_F .$$

Lemma 6. *The linear operator $\mathscr{M}[\cdot]$ satisfies: (i) $\mathscr{M}[F] = 0$ if F is linear in X; (ii) $E[\mathscr{M}[F]X] = E[\mathscr{M}[F]] = 0$.*

For ease of notation, I omit the exchange subscript (i, j) and the superscript $D2C$ and $D2D$. Instead, I refer the D2D exchange pricing kernel to be N, and any D2C exchange pricing kernel to be M (There are many D2C exchanges, however they share the same 'kind' of pricing kernel formula.).

The following lemma help me to further reduce the number of endogenous parameters. It reads the Lagrange multipliers for the two no-arbitrage conditions, (1.6) and (1.7), are linearly related.

Lemma 7. *The Lagrange multiplier for the no-arbitrage conditions satisfy*

$$\mu_{(i,j),r} = -\mathrm{e}^r s\mu_{(i,j),s} .$$

Assumption 5. *I assume the subsistence parameter c is chosen such that the customer's out-side option has an interior solution in the fragmented equilibrium with competitive D2C exchanges.*

I make the following change of variables: $w = \lambda^{-1}$ for all $D2C$ pair (i, j). Then the F.O.C. of the bargaining problem (1.12) becomes,

$$0 = -M^{-1}\pi w + M^{-2}Nw - (F + c)(\kappa - \pi) - \mu(X - s\mathrm{e}^r) .$$

For each D2C exchange, the endogenous parameters satisfy (1.8), (1.13), (1.7):

$$\begin{aligned} w &= E[M(F + c)] , \\ \kappa &= E[M^{-1}N] , \\ \mathrm{e}^{-r} &= E[M] . \end{aligned}$$

Next, for the D2D exchange, the global endogenous parameters satisfy

$$\begin{aligned} s &= E[NX] , \\ s &= \iint_{[0,1]^2} \kappa w \, di \, dj + w_D - 2c\mathrm{e}^{-r} . \end{aligned}$$

Monopolistic Dealers

Proposition 9. *When $F^{(0)} = \alpha X$ and $\pi^{(0)} = 0$, the fragmented equilibrium with monopolistic dealers coincides with the centralized, competitive equilibrium.*

Lemma 8. *The future price of the risk security is*

$$e^r s^{(0)} = \frac{1}{E[(X+2c)^{-1}]} - 2c.$$

Hence, when dealers are option sellers, the shape of the mid-pricing kernel is determined by the option buyers nonlinear endowments, implying a positive variance risk premium.

Proposition 10. *When the nonlinear risk $F^{(1)}$ is small, there exist a unique equilibrium. The D2D pricing kernel is $N = M^{(0)} + \epsilon_F N^{(1)}$, in which* [1]

$$N^{(1)} = \frac{1}{s^{(0)} + 2ce^{-r} + w_D^{(0)}} \mathcal{M}\left[-F_J^{(1)}\right].$$

The pricing kernel for each of the D2C exchanges is $M^{(1)} = M^{(0)} + \epsilon_F M^{(1)}$, in which

$$M^{(1)} = -\frac{1}{2w^{(0)}} \mathcal{M}\left[F^{(1)}\right] + \frac{1}{2} N^{(1)}.$$

Corollary 2. *The first-order effect of nonlinear exposures and market power on the risk premium of the underlying asset is zero. The effect would be non-zero if either the risk-free asset or the risky asset is not available to all customers.*

Lemma 9. *For each of the customers, the equilibrium demand on the risky security based on the outside option is,*

$$b^{(1)} = \frac{1}{e^r s^{(0)} + 2c} \frac{\mathrm{Cov}\left[M^{(0)}, F^{(1)} M^{(0)}\right]}{\mathrm{Var}\left[M^{(0)}\right]} = \mu^{(1)}.$$

Competitive Dealers For simplicity, assume the dealers have the same market power $\pi^{(1)} > 0$, in an 'almost' competitive D2C exchange, I have $\pi = 1 - \epsilon_\pi \pi^{(1)}$.

Proposition 11. *When the market power shock is small, there exists a unique equilibrium. The D2C pricing kernel is $N = M^{(0)} + \epsilon_\pi N^{(1)}$, in which*

$$N^{(1)} = \frac{1}{s^{(0)} + 2ce^{-r}} \mathcal{M}[-F_J]\pi^{(1)}.$$

The pricing kernel for each of the D2C exchange is $M = M^{(0)} + \epsilon_\pi M^{(1)}$, in which

$$M^{(1)} = -\frac{1}{w^{(0)}} \mathcal{M}[F]\pi^{(1)} + N^{(1)}.$$

Lemma 10. *Suppose the nonlinear risk F is a convex, and continuous twice differentiable*

[1] I define $F_J^{(1)} \equiv -\iint_{[0,1]^2} F^{(1)}\, di\, dj$.

function defined on $[X_{\min}, X_{\max}]$*, and satisfies*

$$\lim_{X\to X_{\min}}\left(-F+(X+2c)F'+\frac{\mathrm{Cov}[M^{(0)},M^{(0)}F]}{\mathrm{Var}[M^{(0)}]}\right)<0\,,$$
$$\lim_{X\to X_{\max}}\left(-F+(X+2c)F'+\frac{\mathrm{Cov}[M^{(0)},M^{(0)}F]}{\mathrm{Var}[M^{(0)}]}\right)>0\,,$$

then the following results hold

- *the linear operator* $\mathcal{M}[F]$ *has only one critical point* $X^* \in [X_{\min}, X_{\max}]$;
- *the equation* $\mathcal{M}[F]=0$ *has two roots* $X_1, X_2 \in [X_{\min}, X_{\max}]$; [2]
- *the linear operator* $\mathcal{M}[F]$ *is positive for* $X < X_1$ *or* $X > X_2$*, and is negative for* $X \in [X_1, X_2]$;
- $\mathcal{M}[F]$ *is decreasing for* $X \in [X_{\min}, X^*]$*, and increasing in* $X \in [X^*, X_{\max}]$*;*

Proof. The critical point of linear operator $\mathcal{M}[F]$ can be determined by

$$-(X+2c)^{-2}F+(X+2c)^{-1}F'+(X+2c)^{-2}\frac{\mathrm{Cov}[M^{(0)},M^{(0)}F]}{\mathrm{Var}[M^{(0)}]} \;=\; 0\,.$$

As the pricing kernel $M^{(0)}>0$, I multiply both sides by $(X+2c)^2$ to get

$$-F+(X+2c)F'+\frac{\mathrm{Cov}[M^{(0)},M^{(0)}F]}{\mathrm{Var}[M^{(0)}]} \;=\; 0\,.$$

The left-hand side is monotonic as its first-order derivative is

$$-F'+(X+2c)F''+F' \;=\; (X+2c)F''\,.$$

Hence, F'' determines whether the equation is increasing or decreasing. For convex F, the left-hand side is monotonically increasing. Hence, for the system to have a solution $X \in [X_{\min}, X_{\max}]$, I need

$$\lim_{X\to X_{\min}} -F+(X+2c)F'+\frac{\mathrm{Cov}[M^{(0)},M^{(0)}F]}{\mathrm{Var}[M^{(0)}]}<0,$$

as well as

$$\lim_{X\to X_{\max}} -F+(X+2c)F'+\frac{\mathrm{Cov}[M^{(0)},M^{(0)}F]}{\mathrm{Var}[M^{(0)}]}>0.$$

Therefore, $\mathcal{M}[F]$ first increases then decreases. As $E[\mathcal{M}[F]]=0$, there are two solutions $X_1, X_2 \in [X_{\min}, X_{\max}]$ to the following equation

$$\mathcal{M}[F] \;=\; 0\,.$$

[2] Without loss of generality, I assume $X_1 < X_2$.

□

In practice, customers can buy OTM puts and sell OTM calls, which is covered by this lemma. The net effects depend on the particular choice of the physical density, as well as the shape of the two functions. For example, if the dealers long OTM puts and short OTM calls, then in equilibrium, customers buy OTM puts and sell OTM calls; this demand creates downward pressure on the skewness of the risk-neutral density (measured by the 'mid' price), while the risk-neutral variance depends on the relative selling pressure between the calls and the puts.

Lemma 11. *For small nonlinear risk, the equilibrium price for the risky security is*

$$s = s^{(0)} + \epsilon^2 s^{(2)},$$

in which

$$s^{(2)} = -e^r \operatorname{Cov}\left[M^{(0)}, (M^{(0)})^{-2}\left(\iint (M^{(1)})^2 w^{(0)}\, di\, dj + (N^{(1)})^2 w_D^{(0)}\right)\right].$$

A.3 Proofs

Proof of Proposition 1. Suppose the claim holds, then $M^{\text{D2C}}_{(i,j)} = M^{\text{D2D}}$ for all D2C exchanges. Then from the inter-dealer market clearing condition I get,

$$X = \int_0^1 J(\lambda_{(i,j)} M^{\text{D2D}})\, di\, dj + \int_0^1 J(\lambda_{(j,i)} M^{\text{D2D}})\, di\, dj,$$

where the two Lagrange multipliers are given by their corresponding budget constraints,

$$\begin{aligned}
0 &= E[M^{\text{D2D}} J(\lambda_{(i,j)} M^{\text{D2D}})] - E[M^{\text{D2D}} F_i^{\text{C}}], \\
0 &= E[M^{\text{D2D}} J(\lambda_{(j,i)} M^{\text{D2D}})] - E[M^{\text{D2D}} F_j^{\text{D}}] + E[M^{\text{D2D}} G^*_{(i,j)}].
\end{aligned}$$

Note that as the pricing kernels are the same across exchanges, the last term in the dealer j's budget constraint becomes customer i's budget constraint, which is zero.

Next, I need to verify that indeed M^{d2d} solves the Nash bargaining problem (1.5). This is indeed the case.

Hence, I conclude that the fragmented equilibrium is equivalent to the all-to-all competitive equilibrium. Furthermore, this equilibrium allocation is unique. To see this, Suppose now that there is another solution that also solves the Nash bargaining F.O.C., then multiply M^{D2D} on both sides, and take expectation to obtain,

$$E\left[J'(\lambda_{(i,j)} M^{\text{D2C}}_{(i,j)}) M^{\text{D2C}}_{(i,j)} M^{\text{D2D}}\right]^2 = E\left[J'(\lambda_{(i,j)} M^{\text{D2C}}_{(i,j)})(M^{\text{D2D}})^2\right] E\left[J'(\lambda_{(i,j)} M^{\text{D2C}}_{(i,j)})(M^{\text{D2C}}_{(i,j)})^2\right],$$

where I have used Lemma 7. Hence, by Cauchy-Schwartz inequality, the equality holds only when $M^{\text{D2C}}_{(\text{i,j})} \propto M^{\text{D2D}}$ (i.e., $M^{\text{D2C}}_{(\text{i,j})} = M^{\text{D2D}}$). □

Proof of Theorem 1. Then the inter-dealer market clearing condition is

$$X = \int_{[0,1]^2} (\lambda_{(\text{i,j})} M^{\text{D2C}}_{(\text{i,j})})^{-1}\, di\,dj + (M^{\text{D2D}})^{-1} \int_{[0,1]^2} (\lambda_{(\text{j,i})})^{-1}\, di\,dj - 2c.$$

Suppose there are only two types of customers, then

$$X = \sum_{i=1,2} \alpha_i (\lambda_{(\text{i,j})} M^{\text{D2C}}_{(\text{i,j})})^{-1} + (\lambda_D M^{\text{D2D}})^{-1} - 2c,$$

where α_i is the population of type i customer, and

$$\lambda_D^{-1} \equiv \sum_{i=1,2} \alpha_i \lambda^{-1}_{(\text{j,i})}.$$

$$0 = A^2 (M^{\text{D2D}})^4 + 2AB(M^{\text{D2D}})^3 + D(M^{\text{D2D}})^2 + EM^{\text{D2D}} + F.$$

$$\begin{aligned}
A &= (X+2c)^2,\\
B &= -(X+2c)\left(\frac{\alpha_1 \pi_{(1,j)}}{\lambda_{(1,j)}} + \frac{\alpha_2 \pi_{(2,j)}}{\lambda_{(2,j)}} + \frac{2}{\lambda_D}\right) + \frac{Z_{(1,j)}\alpha_1^2}{\lambda_{(1,j)}} + \frac{Z_{(2,j)}\alpha_2^2}{\lambda_{(2,j)}},\\
C &= \frac{\alpha_1 \pi_{(1,j)} \alpha_2 \pi_{(2,j)}}{2\lambda_{(1,j)}\lambda_{(2,j)}} + \frac{\alpha_1 \pi_{(1,j)}}{\lambda_D \lambda_{(1,j)}} + \frac{\alpha_2 \pi_{(2,j)}}{\lambda_D \lambda_{(2,j)}} + \frac{1}{\lambda_D^2},\\
D &= B^2 + 2AC - \frac{4Z_{(1,j)} Z_{(2,j)} \alpha_1^2 \alpha_2^2}{\lambda_{(1,j)}\lambda_{(2,j)}},\\
E &= 2BC + \frac{Z_{(1,j)}\alpha_1^2 \alpha_2^2 \pi_{(2,j)}^2}{\lambda_{(1,j)}\lambda_{(2,j)}^2} + \frac{Z_{(2,j)}\alpha_1^2 \alpha_2^2 \pi_{(1,j)}^2}{\lambda_{(1,j)}^2 \lambda_{(2,j)}},\\
F &= C^2 - \frac{\alpha_1^2 \alpha_2^2 \pi_{(1,j)}^2 \pi_{(2,j)}^2}{4\lambda_{(1,j)}^2 \lambda_{(2,j)}^2}.
\end{aligned}$$

□

Proof of Proposition 2. From Lemma 10, there exists a point K^*, such that $\int_0^{K^*} \mathscr{M}[F]\,\mathrm{dX} = 0$. Suppose the two roots are given by $K_1 < K^* < K_2$, then a call option with strike $K \in (K^*, K_2)$

has price increment

$$\begin{aligned}
E[\mathcal{M}[F](X-K)^+] &= \int_K^{K_{\max}} \mathcal{M}[F](X-K)\,\mathrm{dX}, \\
&= \int_K^{K_2} \mathcal{M}[F](X-K)\,\mathrm{dX} + \int_{K_2}^{K_{\max}} \mathcal{M}[F](X-K)\,\mathrm{dX}, \\
&= \int_K^{K_2} \mathcal{M}[F](X-K_2)\,\mathrm{dX} + (K_2-K)\int_K^{K_{\max}} \mathcal{M}[F]\,\mathrm{dX} + E[\mathcal{M}[F](X-K_2)^+]\,.
\end{aligned}$$

All the three terms are positive in the last expression. The argument is similar for any $K \in (K_1, K^*)$. This proofs that option prices are all positive. As the variance risk premium is the positively weighted-sum of all available option prices [Bakshi et al., 2003], this immediately suggests a positive variance risk premium.

$$m_2^{(1)} \;=\; \mathrm{e}^r E\left[\mathrm{M}[F]\left(\log X - m_1^{(0)}\right)^2\right] > 0\,.$$

□

Proof of Proposition 3. The first-order effect of a 'small' nonlinear shock F_J on the skewness risk premium derived from the average D2C pricing kernel is proportional to

$$E\left[\mathcal{M}[F_J]\left(\left(\log\frac{X}{s^{(0)}} - m_1^{\mathbb{Q}}\right)^3 - 3m_2^{\mathbb{Q}}\log\frac{X}{s^{(0)}}\right)\right]\,.$$

Then take the functional derivative with respect to F_J to obtain

$$P\mathcal{M}\left[\left(\log\frac{X}{s^{(0)}} - m_1^{\mathbb{Q}}\right)^3 - 3m_2^{\mathbb{Q}}\log\frac{X}{s^{(0)}}\right]\,.$$

As P is positive, only the second term matters at determining the sign of the first-order effect on the skewness risk premium. The derivative of the second term with respect to X yields a cubic equation for $\log\frac{X}{s^{(0)}}$. Hence, the second term of the functional derivative has at most three critical points (i.e., M shape) and at least one critical point. Furthermore, when $X \to 0$, the functional derivative converges to $-\infty$. Therefore, demand on options that are far out-of-the-money pushes down the skewness risk premium.

Next, as the linear operator $\mathcal{M}[\cdot]$ has mean of zero, suggesting that at least one critical point has to be above zero. Hence, options' demand nearby this point may push up the skewness risk premium.

Now, assume P is log-normally distributed, then direct computation shows that the second term satisfies the following properties: (i) when $X \to \infty$, the functional derivative converges to a negative constant; (ii) for σ smaller than a threshold $\bar{\sigma}$, demand on OTM call options with strikes slightly above the future price, $\mathrm{e}^r s^{(0)}$, pushes up the skewness risk premium; (iii) for the same threshold, the last time the functional derivative crosses x-axis at $\hat{X} \gg e^r s^{(0)}$, hence call option demand at that resion has negligible effects on the skewness risk premium.

□

Proof of Proposition 4. See proofs for Proposition 5 and 6. □

Proof of Proposition 5. For an almost competitive D2C exchange, s.t. ϵ_π, the customer's option demand is

$$-F'' + \epsilon_\pi \left(-\frac{M^{(1)} w^{(0)}}{(M^{(0)})^2} \right)'' = -F'' + \epsilon_\pi \left(F'' \pi^{(1)} + \frac{w^{(0)}}{s^{(0)} + 2ce^{-r}} F_J'' \pi^{(1)} \right).$$

Aggregate over all customers to get,

$$F_J'' + \epsilon_\pi \left(-\frac{w_D^{(0)}}{s^{(0)} + 2ce^{-r}} F_J'' \pi^{(1)} \right).$$

For convex F_J, customers overall buy more options when dealers' wealth is reduced. Meanwhile, the mid pricing kernel is,

$$M^{(0)} + \epsilon_\pi \left(\frac{1}{s^{(0)} + 2ce^{-r}} \frac{w_D^{(0)}}{s^{(0)} + 2ce^{-r} - w_D^{(0)}} \mathcal{M}[F_J] \pi^{(1)} \right).$$

Hence, for convex F_J, reducing dealers' wealth reduces the average price for customers to buy options. □

Proof of Proposition 6. According to the Carr-Madan formula, customer's option demand is the second-order derivative of the demand function G,

$$G'' = \epsilon_F \left(-\frac{w^{(0)}}{(M^{(0)})^2} M^{(1)} - F \right)''.$$

Plug-in the formula for $M^{(1)}$ to get,

$$\epsilon_F \left(-\frac{1}{2} F'' + \frac{1}{2} \frac{w^{(0)}}{s^{(0)} + 2ce^{-r} + w_D^{(0)}} \left(F_J^{(1)} \right)'' \right).$$

Then aggregate among all customers to get *customers' net buy of options*,

$$\epsilon_F \left(\frac{s^{(0)} + 2ce^{-r}}{s^{(0)} + 2ce^{-r} + w_D^{(0)}} \left(F_J^{(1)} \right)'' \right).$$

The first term in the product is decreasing in $w_D^{(0)}$, hence customers buy more options when dealers' wealth decreases, for a convex $F_J^{(1)}$. Meanwhile, the mid pricing kernel is

$$M^{(0)} + \epsilon_F \left(\frac{\iint_{[0,1]^2} w^{(0)} M^{(1)} \, di \, dj}{s^{(0)} + 2c\mathrm{e}^{-r} - w_D^{(0)}} \right).$$

Plug-in the definitions to get

$$M^{(0)} + \epsilon_F \left(\frac{w_D^{(0)}}{(s^{(0)} + 2c\mathrm{e}^{-r})^2 - (w_D^{(0)})^2} \mathscr{M}\left[F_J^{(1)}\right] \right).$$

Similarly, for a convex $F_J^{(1)}$, customers pay less to buy options from dealers, if $w_D^{(0)}$ is reduced. □

Proof of Proposition 7. It follows directly from Proposition 2 and 3. □

Proof of Proposition 8. When the D2C exchanges are 'almost' competitive, the effective percentage bid-ask spreads are given by

$$\epsilon_\pi \frac{|E[(M^{(1)} - \bar{M}^{(1)})O(K)]|}{E[M^{(0)} O(K)]} + \mathscr{O}(\epsilon_\pi^2).$$

Plug-in the definitions to get the difference in the pricing kernel,

$$\left(-\frac{1}{w^{(0)}} \mathscr{M}[F] + \frac{1}{s^{(0)} + 2c\mathrm{e}^{-r} - w_D^{(0)}} \mathscr{M}[-F_J] \right) \pi^{(1)}$$

When the nonlinear risk is 'small' and dealers are monopolists, the spreads are given by

$$\epsilon_F \frac{|E[(M^{(1)} - \bar{M}^{(1)})O(K)]|}{E[M^{(0)} O(K)]} + \mathscr{O}(\epsilon_F^2).$$

Plug-in the definitions to get the difference in the pricing kernel,

$$\frac{1}{2} \left(-\frac{1}{w^{(0)}} \mathscr{M}\left[F^{(1)}\right] + \frac{1}{s^{(0)} + 2c\mathrm{e}^{-r} - w_D^{(0)}} \mathscr{M}\left[-F_J^{(1)}\right] \right).$$

□

Proof of Proposition 9. The endogenous parameters are solved explicitly, [3]

$$
\begin{aligned}
w^{(0)} &= \alpha s^{(0)} + c\mathrm{e}^{-r}\,, \\
\kappa^{(0)} &= 1\,, \\
\mu^{(0)} &= 0\,, \\
s^{(0)} &= E[M^{(0)}X]\,, \\
w_D^{(0)} &= \alpha_D s^{(0)} + c\mathrm{e}^{-r}\,.
\end{aligned}
$$

□

Proof of Corollary 2. First-order expand on the first-order condition (1.12) of the bargaining problem to get

$$
\begin{aligned}
0 = &\frac{1}{(M^{(0)})^3}\left(M^{(0)}M^{(0)}w^{(1)} + M^{(0)}N^{(1)}w^{(0)} - 2M^{(1)}M^{(0)}w^{(0)}\right) \\
&- \frac{1}{(M^{(0)})^2}\left(M^{(0)}\pi^{(0)}w^{(1)} + M^{(0)}\pi^{(1)}w^{(0)} - M^{(1)}\pi^{(0)}w^{(0)}\right) \\
&- F^{(0)}\kappa^{(1)} + F^{(0)}\pi^{(1)} - F^{(1)}\kappa^{(0)} + F^{(1)}\pi^{(0)} - X\mu^{(1)} - \kappa^{(1)}c + \mathrm{e}^r\mu^{(0)}s^{(1)} + \mathrm{e}^r\mu^{(1)}s^{(0)} + \pi^{(1)}c\,.
\end{aligned}
$$

Given that $\kappa^{(0)} = 1$, $\mu^{(0)} = 0$ and $M^{(0)} = M^{(0)}$, I simplify the equation to get,

$$
\begin{aligned}
0 = &\frac{1}{(M^{(0)})^2}\left(M^{(0)}w^{(1)} - 2M^{(1)}w^{(0)} + N^{(1)}w^{(0)}\right) \\
&- \frac{1}{(M^{(0)})^2}\left(M^{(0)}\pi^{(0)}w^{(1)} + M^{(0)}\pi^{(1)}w^{(0)} - M^{(1)}\pi^{(0)}w^{(0)}\right) \\
&- F^{(0)}\kappa^{(1)} + F^{(0)}\pi^{(1)} + F^{(1)}\pi^{(0)} - F^{(1)} - X\mu^{(1)} - \kappa^{(1)}c + \mathrm{e}^r\mu^{(1)}s^{(0)} + \pi^{(1)}c\,.
\end{aligned}
$$

Then suppose the D2D exchange is monopolistic, then $\pi^{(0)} = \pi^{(1)} = 0$.

$$
\begin{aligned}
0 = &\frac{1}{(M^{(0)})^2}\left(M^{(0)}w^{(1)} - 2M^{(1)}w^{(0)} + N^{(1)}w^{(0)}\right) \\
&- F^{(0)}\kappa^{(1)} - F^{(1)} - X\mu^{(1)} - \kappa^{(1)}c + \mathrm{e}^r\mu^{(1)}s^{(0)}\,.
\end{aligned}
$$

Rearrange and solve for $M^{(1)}$ to get,

$$
M^{(1)} = \frac{1}{2w^{(0)}}\left((M^{(0)})^2(-F^{(0)}\kappa^{(1)} - F^{(1)} - X\mu^{(1)} - \kappa^{(1)}c + \mathrm{e}^r\mu^{(1)}s^{(0)}) + M^{(0)}w^{(1)} + N^{(1)}w^{(0)}\right).
$$

The customer budget constraint implies

$$
w^{(1)} = E[F^{(0)}M^{(1)} + F^{(1)}M^{(0)} + M^{(1)}c]\,.
$$

[3] I define $\alpha_D \equiv 1 - \iint_{[0,1]^2}\alpha\, di\, dj$.

From the no-arbitrage conditions, we know that $E[M^{(1)}] = 0$ and $E[M^{(1)}X] = s^{(1)}$. Hence

$$w^{(1)} \;=\; \alpha s^{(1)} + E[F^{(1)} M^{(0)}]\,.$$

The definition for the endogenous parameter κ implies

$$\kappa^{(1)} \;=\; -E\left[\frac{1}{M^{(0)}}\left(M^{(1)} - N^{(1)}\right)\right].$$

Hence $\kappa^{(1)} = 0$, then I get

$$M^{(1)} = \frac{1}{2w^{(0)}}\left((M^{(0)})^2(-F^{(1)} - X\mu^{(1)} + \mathrm{e}^r \mu^{(1)} s^{(0)}) + M^{(0)} w^{(1)} + N^{(1)} w^{(0)}\right).$$

From the fact that $E[M^{(1)}] = E[N^{(1)}] = 0$, I get

$$\mu^{(1)} \;=\; \frac{E[M^{(0)} w^{(1)} - (M^{(0)})^2 F^{(1)}]}{E[(M^{(0)})^2 (X - \mathrm{e}^r s^{(0)})]}\,.$$

Under the D2D market clearing condition

$$0 \;=\; \left(M^{(0)} w_D^{(1)} - N^{(1)} w_D^{(0)}\right) + \iint_{[0,1]^2} \left(M^{(0)} w^{(1)} - M^{(1)} w^{(0)}\right) di\,dj\,.$$

Take expectation to get

$$w_D^{(1)} \;=\; -\iint_{[0,1]^2} w^{(1)}\, di\,dj\,.$$

Multiply by X and take expectation on the D2D market clearing condition to get

$$s^{(1)} \;=\; 0\,.$$

Then solve for the D2D pricing kernel,

$$\begin{aligned} N^{(1)} = {} & \left((1+\alpha_D)\, s^{(0)} + 3c\mathrm{e}^{-r}\right)^{-1} \\ & \times \iint_{[0,1]^2} (M^{(0)})^2 (F^{(1)} + \mu^{(1)}(X - \mathrm{e}^r s^{(0)})) - M^{(0)} w^{(1)}\, di\,dj\,. \end{aligned}$$

The D2C pricing kernel is,

$$M^{(1)} = \frac{1}{2w^{(0)}}\left((M^{(0)})^2(-F^{(1)} - \mu^{(1)}(X - \mathrm{e}^r s^{(0)})) + M^{(0)} w^{(1)} + N^{(1)} w^{(0)}\right).$$

The endogenous parameters are given by

$$
\begin{aligned}
w^{(1)} &= E\left[F^{(1)} M^{(0)}\right], \\
\kappa^{(1)} &= 0, \\
\mu^{(1)} &= \frac{1}{\mathrm{e}^r s^{(0)} + 2c} \frac{\mathrm{Cov}\left[M^{(0)}, M^{(0)} F^{(1)}\right]}{\mathrm{Var}\left[M^{(0)}\right]}, \\
s^{(1)} &= 0. \\
w_D^{(1)} &= -\iint_{[0,1]^2} w^{(1)} \, di \, dj.
\end{aligned}
$$

□

Proof of Proposition 11. From the bargaining first-order condition

$$
\begin{aligned}
0 = &-\frac{1}{(M^{(0)})^2}\left(M^{(0)}\pi^{(1)} w^{(0)} + M^{(1)} w^{(0)} - N^{(1)} w^{(0)}\right) \\
&- F\kappa^{(1)} + F\pi^{(1)} - X\mu^{(1)} - \kappa^{(1)} c + \mathrm{e}^r \mu^{(1)} s^{(0)} + \pi^{(1)} c
\end{aligned}
$$

We get $\kappa^{(1)} = 0$ from its definition. Next we solve for the D2C exchange pricing kernel,

$$
M^{(1)} = \frac{1}{w^{(0)}}\left((M^{(0)})^2\left((F+c)\pi^{(1)} - X\mu^{(1)} + \mathrm{e}^r \mu^{(1)} s^{(0)}\right) + w^{(0)}\left(-M^{(0)}\pi^{(1)} + N^{(1)}\right)\right)
$$

From the customer's budget constraint, I get

$$
w^{(1)} = E[FM^{(1)}].
$$

From the no-arbitrage condition, I get

$$
\mu^{(1)} = \pi^{(1)} \frac{E[(M^{(0)})^2(F+c)] - E[M^{(0)}]w^{(0)}}{E[(M^{(0)})^2 X] - E[(M^{(0)})^2]\mathrm{e}^r s^{(0)}}.
$$

From the D2D market-clearing condition

$$
\begin{aligned}
N^{(1)} = &(s^{(0)} + 2c\mathrm{e}^{-r})^{-1} \\
&\times \left(-\iint_{[0,1]^2} (M^{(0)})^2\left((F+c)\pi^{(1)} - \mu^{(1)}(X - \mathrm{e}^r s^{(0)})\right) - w^{(0)} M^{(0)} \pi^{(1)} \, di \, dj\right).
\end{aligned}
$$

The endogenous parameters are given by

$$
\begin{aligned}
w^{(1)} &= E[FM^{(1)}]\,, \\
\kappa^{(1)} &= 0\,, \\
\mu^{(1)} &= \frac{\pi^{(1)}}{e^r s^{(0)} + 2c} \frac{\mathrm{Cov}[M^{(0)}, M^{(0)}(F+c)]}{\mathrm{Var}[M^{(0)}]}\,, \\
s^{(1)} &= 0\,. \\
w_D^{(1)} &= -\iint_{[0,1]^2} w^{(1)}\, di\, dj\,.
\end{aligned}
$$

□

Proof of Lemma 11. The first-part comes directly from the first-order expansion. The second-part comes from the second-order expansion for the D2D market clearing condition, For the customer's consumption,

$$
\left(2(M^{(0)})^2 w^{(2)} - 2M^{(0)} M^{(1)} w^{(1)} - 2M^{(0)} M^{(2)} w^{(0)} + 2(M^{(1)})^2 w^{(0)}\right).
$$

For the dealer's consumption,

$$
\left(2(M^{(0)})^2 w_D^{(2)} - 2M^{(0)} N^{(1)} w_D^{(1)} - 2M^{(0)} N^{(2)} w_D^{(0)} + 2(N^{(1)})^2 w_D^{(0)}\right).
$$

□

Lemma 12. *The functional derivatives for the indirect utilities are*

$$
\begin{aligned}
\frac{\delta v_{(i,j)}[M^{\mathrm{D2C}}_{(i,j)}]}{\delta M^{\mathrm{D2C}}_{(i,j)}} &= -P\lambda_{(i,j)}(J(\lambda_{(i,j)} M^{\mathrm{D2C}}_{(i,j)}) - F_i^{\mathrm{C}}), \\
\frac{\delta v_{(j,i)}[M^{\mathrm{D2C}}_{(i,j)}]}{\delta M^{\mathrm{D2C}}_{(i,j)}} &= P\lambda_{(j,i)}\kappa_{(i,j)}(J(\lambda_{(i,j)} M^{\mathrm{D2C}}_{(i,j)}) - F_i^{\mathrm{C}}) \\
&\quad + P\lambda_{(j,i)}\lambda_{(i,j)} J'(\lambda_{(i,j)} M^{\mathrm{D2C}}_{(i,j)})(\kappa_{(i,j)} M^{\mathrm{D2C}}_{(i,j)} - M^{\mathrm{D2D}}).
\end{aligned}
$$

where $\kappa_{(i,j)}$ *is defined in* (1.13).

Proof. Direct calculation yields the results. □

Lemma 13. *Dealer j's profit from D2C trading is* $-E[M^{\mathrm{D2D}} G^*_{(i,j)}(M^{\mathrm{D2C}}_{(i,j)})]$*, and it is decreasing in the inverse market power* $\pi_{(i,j)}$.

Proof. From the Nash bargaining first-order condition, I get

$$
G^*_{(\mathrm{i,j})} = -\frac{\lambda_{(\mathrm{i,j})} J'(\lambda_{(\mathrm{i,j})} M^{\mathrm{D2C}}_{(\mathrm{i,j})})(\kappa_{(\mathrm{i,j})} M^{\mathrm{D2C}}_{(\mathrm{i,j})} - M^{\mathrm{D2D}})}{\kappa_{(\mathrm{i,j})} - \pi_{(\mathrm{i,j})}}\,.
$$

Taking expectations and plug-in the definition to get

$$-E[M^{\text{D2D}}G^*_{(i,j)}] = \lambda_{(i,j)} \frac{E[M^{\text{D2C}}_{(i,j)} M^{\text{D2D}} J'(\lambda_{(i,j)} M^{\text{D2C}}_{(i,j)})]^2 - E[(M^{\text{D2D}})^2 J'(\lambda_{(i,j)} M^{\text{D2C}}_{(i,j)})] E[(M^{\text{D2C}}_{(i,j)})^2 J'(\lambda_{(i,j)} M^{\text{D2C}}_{(i,j)})]}{(\kappa_{(i,j)} - \pi_{(i,j)}) E[M^{\text{D2C}}_{(i,j)} J'(\lambda_{(i,j)} M^{\text{D2C}}_{(i,j)})]} > 0$$

The inequality comes from Holder inequality. Also, dealer will agree to trade only when this term is positive, hence, I need that $\kappa_{(i,j)} > \pi_{(i,j)}$. □

A.4 Miscellaneous

The ETF sample is selected based on the average daily volume during 2015 on ISE exchange.

- Equity Sector: XRT, SMH, XBI, XLY, IBB, XLV, XLI, XLU, XLE, IYR, XLF;
- Equity Index: DIA, IWM, SPY, QQQ;
- Equity International: ASHR, RSX, EWJ, DXJ, EFA, EWZ, FXI, EEM;
- Fixed-income: HYG, TLT;
- Commodity: UNG, OIH, SLV, GDX, GLD, USO, XOP;
- Currency: FXE, UUP;

Risk-Neutral Moments According to the Bakshi et al. [2003], the risk-neutral variance is given by

$$\text{Variance}^{\mathbb{Q}}_t(T) = \frac{e^{r(T-t)} V_t(T) - \mu_t(T)^2}{T-t}$$

The risk-neutral skewness is given by

$$\text{Skew}^{\mathbb{Q}}_t(T) = \frac{e^{r(T-t)} W_t(T) - 3\mu_t(T) e^{r(T-t)} V_t(T) + 2\mu_t(T)^3}{\left(e^{r(T-t)} V_t(T) - \mu_t(T)^2\right)^{3/2}}.$$

The time t prices of the time T quadratic, cubic and quartic payoffs are given as the weighted sum of OTM calls and puts,

$$V_t(T) = \int_{S_t}^{\infty} \frac{2\left(1 - \log\frac{K}{S_t}\right)}{K^2} O_t(C,K,T)\,dK + \int_0^{S_t} \frac{2\left(1 + \log\frac{S_t}{K}\right)}{K^2} O_t(P,K,T)\,dK,$$

$$W_t(T) = \int_{S_t}^{\infty} \frac{6\log\frac{K}{S_t} - 3\left(\log\frac{K}{S_t}\right)^2}{K^2} O_t(C,K,T)\,dK - \int_0^{S_t} \frac{6\log\frac{S_t}{K} + 3\left(\log\frac{S_t}{K}\right)^2}{K^2} O_t(P,K,T)\,dK,$$

$$X_t(T) = \int_{S_t}^{\infty} \frac{12\left(\log\frac{K}{S_t}\right)^2 - 4\left(\log\frac{K}{S_t}\right)^3}{K^2} O_t(C,K,T)\,dK + \int_0^{S_t} \frac{12\left(\log\frac{S_t}{K}\right)^2 + 4\left(\log\frac{S_t}{K}\right)^3}{K^2} O_t(P,K,T)\,dK.$$

and

$$\mu_t(T) \approx \mathrm{e}^{r(T-t)} - 1 - \frac{\mathrm{e}^{r(T-t)}}{2} V_t(T) - \frac{\mathrm{e}^{r(T-t)}}{6} W_t(T) - \frac{\mathrm{e}^{r(T-t)}}{24} X_t(T).$$

B Appendix to Chapter 2

This appendix contains all proofs for Chapter 2.

Proof of Proposition 1. Suppose that the variance of ε_q goes to zero, then we have that the covariance between pricing kernel and the spot price P being solely determined by the variation in ε_p. In this case, we know that the pricing kernel is a decreasing function in ε_p, while the price is an increasing function. Hence, we conclude that the covariance is negative.

Suppose that the variance of ε_p goes to zero, then we have that the covariance between pricing kernel and the spot price P being solely determined by the variation in ε_q. In this case, we have that both the pricing kernel and the spot price being a decreasing function of the quantity shock ε_q. Hence, the covariance is positive. □

Proof of Lemma 4. The first-order expansion for equation (2.6) gives

$$\left(E[M^{(1)}]\kappa_* - \delta\kappa^{(1)} + \kappa^{(1)}\right) e^{-\delta \log(\kappa_*)} = E[M^{(1)}\delta + \alpha^{(1)}\delta \log(Z)].$$

Note that $E[M^{(1)}] = 0$. □

Proof of Proposition 5. Assume that

$$X_{p,\epsilon} = X_{p,\epsilon} + \epsilon X_p^{(1)}.$$

Expand the optimal consumption demand of the producer to the first-order, we get

$$X_p^{(1)}(P) = -\frac{\delta - \lambda_p}{\alpha}\kappa_*^{\delta} \log P\alpha^{(1)} + const_p.$$

For the consumer, similarly, we can find his optimal consumption demand as

$$X_c^{(1)}(P) = \frac{\delta - 1 + \lambda_c}{\alpha}\kappa_*^{\delta} \log P\alpha^{(1)} + const_c.$$

Then, we take the second-order derivative with respect to P to get the claim. □

Proof of Corollary 6. The aggregate put option demand from the hedgers is

$$\int_{P_{\min}+\epsilon P_{\min}^{(1)}}^{F_\epsilon} D_\epsilon(P)\,dP = \epsilon\frac{1-\lambda_p-\lambda_c}{\alpha}\kappa_*^\delta\left(P_{\min}^{-1}-F^{-1}\right)\alpha^{(1)}.$$

The aggregate call option demand from the hedgers is

$$\int_{F_\epsilon}^{P_{\max}+\epsilon P_{\max}^{(1)}} D_\epsilon(P)\,dP = \epsilon\frac{1-\lambda_p-\lambda_c}{\alpha}\kappa_*^\delta\left(F^{-1}-P_{\max}^{-1}\right)\alpha^{(1)}.$$

Note that we don't need to compute the first-order terms for F_ϵ and P_ϵ, as they would show up only in the second order terms after the integration. □

Proof of Lemma 5. The first-order perturbation for the pricing kernel is

$$M^{(1)} = \lambda_M^{-1}\lambda_M^{(1)} - \frac{\alpha^{(1)}\gamma\kappa e^{\delta\log(\kappa)}\log(Z)-\delta\gamma\kappa\kappa^{(1)}+\delta\gamma\kappa^{(1)}e^{\delta\log(\kappa)}}{-\delta\kappa^2+\delta\kappa w+\kappa e^{\delta\log(\kappa)}}$$

Then we compute the $\lambda_M^{(1)}$,

$$\lambda_M^{(1)} = \lambda_M\frac{\alpha^{(1)}\gamma\kappa e^{\delta\log(\kappa)}E[\log(Z)]-\delta\gamma\kappa\kappa^{(1)}+\delta\gamma\kappa^{(1)}e^{\delta\log(\kappa)}}{-\delta\kappa^2+\delta\kappa w+\kappa e^{\delta\log(\kappa)}}$$

Then, plug the result in the pricing kernel, we get the claim. □

C Appendix to Chapter 3

This appendix contains all proofs for Chapter 3.

Proof of Proposition 1. By the conjecture (3.8), the market clearing price p_2 is jointly normally distributed with θ. By the definition of ξ in (3.4), one has

$$\mathrm{Cov}(D, p_2) = \mathrm{Cov}\left(\theta + \varepsilon, a\frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\delta^2}(\theta + \delta)\right) = a\rho\sigma_\theta^2. \tag{C.1}$$

Furthermore, one has

$$\mathbb{V}(\xi) = \mathbb{V}\left(\frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\delta^2}(\theta + \delta)\right) = \rho\sigma_\theta^2. \tag{C.2}$$

Now (C.1), (C.2) and the normal projection theorem give

$$\mathbb{E}(D|p_2) = \frac{a\rho\sigma_\theta^2}{a^2\rho\sigma_\theta^2 + b^2\sigma_W^2}p_2 = \frac{\psi}{a}(a\xi + bW), \tag{C.3}$$

$$\mathbb{V}(D|p_2) = \frac{1}{\tau_u} = \sigma_\theta^2 + \sigma_\varepsilon^2 - \frac{a^2\rho^2\sigma_\theta^4}{a^2\rho\sigma_\theta^2 + b^2\sigma_W^2} = \sigma_\theta^2 + \sigma_\varepsilon^2 - \psi\rho\sigma_\theta^2. \tag{C.4}$$

Plugging (C.3) and (C.4) into (3.7), using the result with (3.6) in the market-clearing condition

(3.9):

$$\frac{M\tau_\xi(\xi - p_2)}{\gamma_d} + \frac{(N-M)\tau_u(\psi\xi + \frac{\psi b}{a}W - p_2)}{\gamma_d} = W$$

solving for p_2 and matching coefficients with (3.8) yields

$$M\tau_\xi + (N-M)\psi\tau_u = [M\tau_\xi + (N-M)\tau_u]a \tag{C.5}$$

$$(N-M)\tau_u\frac{\psi b}{a} - \gamma_d = [M\tau_\xi + (N-M)\tau_u]b. \tag{C.6}$$

Substituting $\psi = \frac{a^2\rho\sigma_\theta^2}{a^2\rho\sigma_\theta^2 + b^2\sigma_W^2}$ into equations (C.5) and (C.6) and solve for a and b gives the expressions in (3.13) and (3.14).

It is immediately clear from (3.13) that $a > 0$ if $M > 0$, since both numerator and denominator are always positive in this case. Since $\psi > 0$ it follows also that $a \leq 1$, with an equality only if $N = M$. □

Proof of Lemma 1. The dealer's optimal demand schedule follows directly from the first-order condition (3.3) by substituting $\overline{\omega}_k = q_k - x$. The demand schedules of other informed dealers do not change, since they do not make inferences from the price in the interdealer market. The dealers who have not been contacted by the investor perform inferences as described in Proposition 1. One can now conjecture $p_2 = a\xi + b(W - x)$. Thus using dealer k's demand schedule and demand schedules (3.6) and (3.7) for the other dealers in the market clearing condition and following the exact procedure described in the proof of Proposition 1 determines a and b as in 1.

□

Proof of Lemma 2. Let F_x denote the dealers' optimal quoting strategy. This means dealers quote a price p_0 that is a random variable with the distribution function F_x.

Let $x > 0$. Then $x(p_v - p_c) > 0$ implies $p_v > p_c$. If the dealers' optimal strategy were such that there is a $p^* \in (p_v, p_c)$ such that dealers quote a price $p \leq p^*$ with a probability of 1, then a dealer could profitably deviate from this strategy by quoting p_v. This would contradict optimality. On the other hand, quoting a prices greater than p_v with any positive probability cannot be optimal, since the investor would not buy the asset at that price. Thus, one obtains $\sup \operatorname{supp}(F_x) = p_v$.

Now we show that F_x must be continuous, i.e. there cannot be any atoms in the distribution of p_0. Clearly, quoting a price less than or equal to p_c with any positive probability cannot be optimal, since a dealer would not make any positive profit by doing so, whereas he would make a positive expected profit by quoting p_v. Now, suppose there is a price p' with $p_v \geq p' > p_c$ that is quoted with probability $\rho > 0$ by all dealers.

Then a single dealer could again profitably deviate from this strategy which contradicts optimality. The profitable deviation is constructed as follows. Since the number of prices charged with positive probability must be countable, one can find for each $\delta > 0$ an ε_δ, such that $\delta \geq \varepsilon_\delta > 0$ and the price $p' - \varepsilon_\delta$ is charged with probability zero by all dealers. The deviating dealer can now charge price $p' - \varepsilon_\delta$ with probability ρ and charge price p' with probability zero. Using the fact that $\lim_{\delta \to 0} F_x(p' - \varepsilon_\delta) = F_x(p') - \rho$, one can express the difference Δ in profits between the original strategy and the proposed deviation as follows. A dealer quoting p' only makes a positive profit if no other dealer on the platform quotes a lower price. If no other dealer quotes a lower price, there might be $j = 0, 1, \ldots, M-1$ dealers who quote p' as well. In the latter case, each of the $j+1$ is equally likely to be chosen by the investor for trading the asset. The calculation below considers the cases in which j dealers quote price p on the platform separately.

$$\begin{aligned}
\Delta \quad = \quad & (1 - qF_x(p' - \varepsilon_\delta) - q\rho)^{M-1}(p' - \varepsilon_\delta - p_c)x \\
& -(1 - qF_x(p'))^{M-1}(p - c)x \\
& + \sum_{j=1}^{M-1} \binom{M-1}{j} (1 - qF_x(p' - \varepsilon_\delta) - q\rho)^{M-1-j}(q\rho)^j (p' - \varepsilon_\delta - p_c)x \\
& - \sum_{j=1}^{M-1} \binom{M-1}{j} (1 - qF_x(p'))^{M-1-j}(q\rho)^j (p' - p_c)\frac{x}{j+1}.
\end{aligned}$$

The first two lines in the above expression compare expected profits from quoting $p' - \varepsilon_\delta$ and expected profits from quoting p' in the event that all other dealers quote a price above p'. Since $\lim_{\delta \to 0} F_x(p' - \varepsilon_\delta) = F_x(p') - \rho$, the difference in these two lines goes to zero as δ goes to zero. The last two lines compare the respective profits in the cases in which $j > 0$ other dealers quote p'. Since $M \geq 2$, the deviating dealer can get a jump in expected trading volume in this case, since he can avoid ties with other dealers. Therefore one obtains

$$\Delta \to \sum_{j=1}^{M-1} \binom{M-1}{j} (p' - p_c)\frac{jx}{j+1}(1 - qF_x(p'))^{M-1-j}(q\rho)^j > 0 \quad \text{as } \delta \to 0.$$

Thus, the proposed deviation is profitable for a small δ. In equilibrium, F_x cannot have any atoms.

If $x < 0$, one verifies analogously to the case of $x > 0$, that $\inf \operatorname{supp}(F_x) = p_v$ must hold for any optimal strategy. That the distribution cannot have any atoms follows analogously as well.

The dealers are only willing to randomize over prices if they earn the same profit in expectation with each price in the support of F_x. This profit must be equal to the profit in which the dealer quotes $p_v(x)$. This gives the indifference condition expressed in (3.17).

Using the binomial formula $(x+y)^n = \sum_{k=0}^{n}\binom{n}{k}x^{n-k}y^k$, (3.17) simplifies to

$$(p-p_c)x(1-qF_x(p))^{M-1} = (1-q)^{M-1}(p_v(x)-p_c)x,$$

which can be solved for F_x. The solution is given by (3.18).

Using (3.18) and solving $F_x(\overline{p}_x)=0$ for $\overline{p}_x$ gives

$$\overline{p}_x = p_c + (p_v - p_c)(1-q)^{M-1}.$$

Since $x(p_v - p_c) > 0$, one obtains $\overline{p}_x > p_c$ for $x > 0$ and $\overline{p}_x < p_c$ for $x < 0$.

The event that at least one dealer is on the platform happens with probability $1-(1-q)^M$, since all dealers respond independently with probability q. The unconditional probability that no dealer quotes above $p \in \mathrm{supp}(F_x)$ can be expressed by $(1-qF_s(p))^M$. Therefore, the conditional distribution G_x has to satisfy $G(p)(1-(1-q)^M) = 1-(1-qF_x(p))^M$. Performing a change of variables $p = p_c + \frac{(p_v-p_c)(1-q)^{M-1}}{(1-(1-(1-q)^M)u)^{(M-1)/M}}$, one can calculate

$$\int_{\mathrm{supp}(F_x)} p\mathrm{d}G_x(p) = \int_0^1 \left[p_c + \frac{(p_v-p_c)(1-q)^{M-1}}{(1-(1-(1-q)^M)u)^{(M-1)/M}}\right]\mathrm{d}u = p_c + \frac{(p_v-p_c)(1-q)^{M-1}}{1-(1-q)^M}Mq.$$

The claim that $0 \le \kappa < 1$, can be shown as follows. That $0 \le \kappa$ is immediately clear from the definition (3.20). The other inequality can be seen as follows.

- κ as a function of q is strictly decreasing in q for all $q \in (0,1]$, since

  $$\frac{\partial \kappa}{\partial q} = \frac{-M(1-q)^{M-2}\left[(1-q)^M + Mq - 1\right]}{(1-(1-q)^M)^2} < 0$$

  for $q \in (0,1]$.

- By L'Hospital's rule, one has

  $$\lim_{q\to 0}\kappa = \frac{\lim_{q\to 0}(M(1-q)^{M-1} - M(M-1)q(1-q)^{M-2})}{\lim_{q\to 0} M(1-q)^M} = \frac{M}{M} = 1.$$

The last two bullet points imply $\kappa < 1$ for all $q \in (0,1]$.

This proves all statements in the lemma.

$\square$

Proof of Proposition 2. Claim 1: The expected price on the platform is linear in ξ and x. Define

$$\beta_1 := \left[1 - \frac{Mq(1-q)^{M-1}}{1-(1-q)^M}\right]\left[a - \frac{(1-a)b}{\gamma_d \sigma_\xi^2}\right] + \frac{Mq(1-q)^{M-1}}{1-(1-q)^M}\left(1 + \frac{\sigma_\delta^2}{\sigma_\theta^2}\right) \tag{C.7}$$

and

$$\beta_2 := \left[1 - \frac{Mq(1-q)^{M-1}}{1-(1-q)^M}\right]\left[-b - \frac{b^2}{2\gamma_d \sigma_\xi^2}\right] + \frac{Mq(1-q)^{M-1}}{1-(1-q)^M}\left(-\frac{\gamma_I \sigma_\epsilon^2}{2}\right). \tag{C.8}$$

Using the definitions of $p_c(x)$ and $p_v(x)$, it follows by direct computation that $P(x)$ as defined in Lemma 2 is given by (3.22).

Claim 2: Let $\frac{M(1-q)^{M-1}}{1-(1-q)^M} < \frac{1}{2}$. An equilibrium exists if and only if $a < \overline{a}$ for some $\overline{a} \in \mathbb{R}$.

To verify the existence of the described equilibrium, there are several things to check. The strategy (3.25) is well-defined if

$$2\beta_2 + \gamma_I \sigma_\varepsilon^2 \neq 0. \tag{C.9}$$

Furthermore, the investor's second-order condition from the maximization problem (3.23) requires

$$-\gamma_I \sigma_\varepsilon^2 - \left(2 \frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\delta^2} \frac{\beta_1}{\alpha} + 2\beta_2\right) < 0. \tag{C.10}$$

In order to apply Lemma 2 we also need to verify that

$$x(p_v(x) - p_c(x)) > 0 \tag{C.11}$$

for holds for any $x \neq 0$ demanded by the investor in the proposed equilibrium.

If (C.9), (C.10) and (C.11) hold, one can use Lemma 2 to see that there exist optimal strategies for the dealers that yield $P(x)$ as the expected price on the platform conditional on at least one response. As demonstrated in the text, the stated strategy for the investor (3.25) indeed solves the first order condition (3.24), given that dealers rationally infer $\theta+\delta$ from the investor's demand. Thus, both dealers and the investor behave optimally given the strategies of the others and an equilibrium is established.

The strategy of the proof of this claim is as follows. We will assume that the average price is given by the expression in Lemma 2. We then show that the investor's strategy is well-defined so that the first-order and second-order conditions of the maximization problem (3.23) are satisfied. We that verify that in this case

In order to prove our claim, we first note that (C.11) is satisfied in this case, so that dealers indeed find it optimal to quote as described in Lemma 2.

For the following proof, it is worth noting that Lemma 2 states that

$$0 \le \kappa < 1 \tag{C.12}$$

for all $q \in (0,1]$ and $M \ge 2$.

„⇒": *Proof that equilibrium exists under the stated conditions.*

Let now $\kappa < \frac{1}{2}$.

We rewrite (C.7) and (C.8) as using $b = -\frac{\gamma_d \sigma_\xi^2}{M} a$:

$$\beta_1 = (1-\kappa)\left(a + \frac{(1-a)a}{M}\right) + \kappa \frac{\sigma_\theta^2 + \sigma_\delta^2}{\sigma_\theta^2} > 0, \tag{C.13}$$

$$\beta_2 = (1-\kappa)\left(\frac{\gamma_d \sigma_\xi^2}{M} a - \frac{a^2 \gamma_d \sigma_\xi^2}{2M^2}\right) - \kappa\left(\frac{\gamma_I \sigma_\epsilon^2}{2}\right) > -\frac{\gamma_I \sigma_\epsilon^2}{2}, \tag{C.14}$$

where the inequalities follow from (C.12) and $1 \le a > 0$.

Define

$$\Psi := \frac{\sigma_\theta^2 + \sigma_\delta^2}{\sigma_\theta^2} \frac{\frac{1}{2} - \kappa}{1-\kappa} > 0$$

and define $\overline{a}$ as the smaller solution to the quadratic equation

$$a + \frac{(1-a)a}{M} = \Psi,$$

if there is a real solution to the equation. Set $\overline{a} = 1$ otherwise. Then it follows from (C.13) that

$$\beta_1 < \frac{1}{2}\frac{\sigma_\theta^2 + \sigma_\delta^2}{\sigma_\theta^2},$$

if $a < \overline{a}$.

Thus, all that remains to show is that there is an equilibrium if the last inequality involving β_1 holds. As described above it is sufficient to check that (C.9), (C.10) and (C.11) hold. It is immediatly clear from (C.14) that (C.9) always holds for any set of parameters.

Regarding (C.10), note that using (3.25), (C.14) and the assumption on β_1 imply $\alpha > 0$. Using (C.13), one therefore obtains

$$-\gamma_I \sigma_\varepsilon^2 - \left(2\frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\delta^2}\frac{\beta_1}{\alpha} + 2\beta_2\right) < 2\frac{\sigma_\theta^2}{\sigma_\theta^2 + \sigma_\delta^2}\frac{\beta_1}{\alpha} + 2\beta_2 < 0.$$

Thus, the investor's second-order condition holds if $\beta_1 < \frac{1}{2}\frac{\sigma_\theta^2 + \sigma_\delta^2}{\sigma_\theta^2}$.

Lastly, we check that (C.11) holds which justifies the use of Lemma 2 for determining the expected price on the platform. Note that by optimality of the investor's choice of x and $\alpha > 0$, it follows that the investor makes a positive profit if $x \neq 0$ This can be seen, since the investor could always make a zero profit by not trading, but instead chooses a different x. By the convexity of the maximization problem (3.23), the optimal quantity is uniquely determined and therefore must give a positive profit. This implies

$$x(p_v(x) - p_c(x)) \geq x(p_v(x) - p_c(x))(1-\kappa) = x(p_v(x) - P(x)) > 0.$$

Therefore (C.11) indeed holds and Lemma 2 can be used to determine the dealer's quoting strategies on the platform.

Since (C.9), (C.10) and (C.11) indeed hold, the equilibrium exists.

„$\Leftarrow$": *Proof that equilibrium does not exist if $a \geq \overline{a}$.*

The definition of $\overline{a}$ and β_1 imply that $\beta_1 \geq \frac{1}{2}\frac{\sigma_\theta^2 + \sigma_\delta^2}{\sigma_\theta^2}$ if $a \geq \overline{a}$. If the last inequality is an equlity, it follows that $\alpha = 0$. This means, the investor does not trade and the quoting strategies of the dealers are not defined. Let the inequality be strict. Note that by (C.13), $\kappa \in [0,1]$ and $a \in [0,1]$, one has $\beta_1 \leq a + (1-a) = 1$. This in turn implies

$$\frac{1}{1-2\beta_1\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\delta^2}} > -1.$$

One now obtains

$$\begin{aligned}
& -\gamma_I\sigma_\varepsilon^2 - \left(2\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\delta^2}\frac{\beta_1}{\alpha}+2\beta_2\right) \\
= \;& -\gamma_I\sigma_\varepsilon^2 + 2\beta_2 - 2\frac{2\beta_2+\gamma_I\sigma_\varepsilon^2}{1-2\beta_1}\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\delta^2}\beta_1 \\
> \;& -\gamma_I\sigma_\varepsilon^2 + 2\beta_2 + 2(2\beta_2+\gamma_I\sigma_\varepsilon^2)\frac{\sigma_\theta^2}{\sigma_\theta^2+\sigma_\delta^2}\beta_1 \\
> \;& -\gamma_I\sigma_\varepsilon^2 + 2\beta_2 + (2\beta_2+\gamma_I\sigma_\varepsilon^2) = 0.
\end{aligned}$$

Therefore, the second-order condition for the investor's maximization problem (3.23) is not satisfied. Thus, the investor's strategy is clearly not optimal and the described equilibrium does not exist.

Claim 3: The equilibrium does not exist if $\kappa \geq \frac{1}{2}$

In this case, $a \geq 0$ and (C.13) imply $\beta_1 \geq \frac{1}{2}\frac{\sigma_\theta^2+\sigma_\delta^2}{\sigma_\theta^2}$. The prove that the equilibrium doe not exist is identical to the proof in Claim 2.

Claim 4: $a \to 0$ as $N \to \infty$ and $\sigma_W \to \infty$.

By equation (3.13), one can see that

$$\begin{aligned}
\lim_{\sigma_W\to\infty}\lim_{N\to\infty} a &= \lim_{\sigma_W\to\infty}\lim_{N\to\infty}\frac{\frac{M\tau_\xi}{N-M}+\psi\tau_u}{\frac{M\tau_\xi}{N-M}+\tau_u} \\
&= \lim_{\sigma_W\to\infty}\psi \\
&= 0.
\end{aligned}$$

Claim 5: An equilibrium exist if $\kappa < \frac{1}{2}$ and $\sigma_\delta \to \infty$. As $\sigma_\delta \to \infty$, one has $\Psi \to \infty$. This means $a + \frac{(1-a)a}{2M} < \Psi$ for all $a \in \mathbb{R}$ and in particular for all $a \in (0,1]$ As shown in the proof of Claim 2, this implies $\beta_1 < \frac{1}{2}\frac{\sigma_\theta^2+\sigma_\delta^2}{\sigma_\theta^2}$ and the equilibrium exists. □

Proof of Proposition 3. We divided this proof into several steps. The first step is an auxiliary result that will be used later in the proof.

Step 1: $\frac{\partial a}{\partial M} \geq \frac{1}{M} a(1-a)$.

Since $\sigma_\delta = 0$, we get $\sigma_\xi = \sigma_\varepsilon$ and $\rho = \sigma_\theta^2$. We now rewrite a as defined in (3.13) as

$$a = 1 - \frac{\gamma_d^2 \sigma_\varepsilon^4 \sigma_W^2 (N-M)}{M^2 N \sigma_\theta^2 + \gamma_d^2 M \sigma_\varepsilon^2 \sigma_\theta^2 \sigma_W^2 + \gamma_d^2 N \sigma_\varepsilon^4 \sigma_W^2} \tag{C.15}$$

Using the expression in (C.15), one obtains by direct calculation and simplifying terms that

$$\begin{aligned} \frac{\partial a}{\partial M} - \frac{1}{M} a(1-a) &= \frac{\gamma_d^2 M \sigma_\varepsilon^4 \sigma_W^2 \left(\gamma_d^2 \sigma_\varepsilon^2 \sigma_W^2 \left(\sigma_\varepsilon^2 + \sigma_\theta^2\right) + N^2 \sigma_\theta^2\right)}{\left(M^2 N \sigma_\theta^2 + \gamma_d^2 M \sigma_\varepsilon^2 \sigma_\theta^2 \sigma_W^2 + \gamma_d^2 N \sigma_\varepsilon^4 \sigma_W^2\right)^2} \\ &\geq 0. \end{aligned}$$

This proves the first step

Step 2: The equilibrium exists if and only if a is below a certain threshold.

This result follows directly from Proposition 2 by noting that κ as defined (C.12) is equal to zero if $q = 1$. Furthermore, since $q = 1$, one has $\Psi = \frac{1}{2}$, where Ψ is defined in the proof of Proposition 1. Defining $\overline{a}$ as in the proof of Proposition 1, one gets that $\overline{a}$ is the smaller real solution to

$$a + \frac{(1-a)a}{M} = \frac{1}{2},$$

which is always greater than zero and less than $\frac{1}{2}$.

Step 3: There is an equilibrium for all $M' < M$.

In the proof of Proposition 2 it was established that the described equilibrium exists if and only if $\beta_1 < \frac{1}{2}\frac{\sigma_\theta^2+\sigma_\delta^2}{\sigma_\theta^2}$. If an equilibrium exists when M dealers get contacted, it consequently must be the case that $\beta_1 < \frac{1}{2}$. If furthermore, $\beta_1 < \frac{1}{2}$ for all $M' < M$, the result follows. The last claim will be shown next. If $q = 1$, one has

$$
\begin{aligned}
\frac{\partial \beta_1}{\partial M} &= \frac{(1-2a)\frac{\partial a}{\partial M}}{M} + \frac{\partial a}{\partial M} - \frac{(1-a)a}{M^2} \\
&= \frac{M+1-2a}{M}\frac{\partial a}{\partial M} - \frac{(1-a)a}{M^2} \\
&\geq \frac{M+1-2a}{M}\frac{1}{M}a(1-a) - \frac{(1-a)a}{M^2} \\
&> (3(1-a)+a^2)\frac{(1-a)a}{M^2} \\
&> 0.
\end{aligned}
$$

The third line follows from Step 1. Therefore, one has $0 \leq \beta_1 < \frac{1}{2}\frac{\sigma_\theta^2+\sigma_\delta^2}{\sigma_\theta^2}$ for all $M' < M$. Note that event though M represents an integer in the model, β_1 can be interpreted as a function in $C^1(\mathbb{R})$.

Step 4: The investor's payoff is highest if $M' = 2$ compared to all other $M'' \leq M$.

Since we know that a nonzero-trade equilibrium exists for all $M' < M$, we can calculate the investor's equilibrium payoff as defined by (3.27).

Using the expressions for β_1 and β_2 stated in Step 2 and using the definition of α from (3.25), one gets

$$
\pi_I = \frac{\sigma_\theta^2 M\left(2a^2 - 2a(M+1) + M\right)}{2\sigma_\varepsilon^2\left(2a\gamma_d M + \gamma_I M^2 - a^2\gamma_d\right)}.
$$

In equilibrium, one has $\pi_I > 0$. Since $a < \overline{a}$, the numerator in the above expression for π_I is positive. Therefore, the denominator must be positive as well. Interpreting π_I as a function in $C^1(\mathbb{R})$, one can show that π_I is strictly decreasing in M by showing that $\ln(\pi_I)$ is strictly decreasing in M. It then follows that the lowest possible M', i.e. $M' = 2$ is profit maximizing among all possible values less than M.

$$
\begin{aligned}
\frac{\partial}{\partial M}\ln \pi_I &= \frac{M\left(4a\frac{\partial a}{\partial M} - 2(M+1)\frac{\partial a}{\partial M} - 2a + 1\right) + \left(2a^2 - 2(M+1)a + M\right)}{M\left(2a^2 - 2(M+1)a + M\right)} \\
&\quad - \frac{2\gamma_d M\frac{\partial a}{\partial M} - 2\gamma_d a\frac{\partial a}{\partial M} + 2\gamma_d a + 2\gamma_I M}{2\gamma_d M a - \gamma_d a^2 + \gamma_I M^2}
\end{aligned}
$$

Collecting terms gives

$$\begin{aligned}\frac{\partial}{\partial M}\ln\pi_I &= \frac{1}{M}+\frac{1}{2a^2-2(M+1)a+M}-\frac{2a}{2a^2-2(M+1)a+M}\\ &-\frac{2\gamma_I M}{2\gamma_d Ma-\gamma_d a^2+\gamma_I M^2}-\frac{2\gamma_d a}{2\gamma_d Ma-\gamma_d a^2+\gamma_I M^2}+\\ &+\Big(\frac{4a}{2a^2-2(M+1)a+M}-\frac{2(M+1)}{2a^2-2(M+1)a+M}\\ &\frac{2\gamma_d a}{2\gamma_d Ma-\gamma_d a^2+\gamma_I M^2}-\frac{2\gamma_d M}{2\gamma_d Ma-\gamma_d a^2+\gamma_I M^2}\Big)\frac{\partial a}{\partial M}\end{aligned}$$

Since $a<\overline{a}$ one has $a^2-a(M+1)+M/2>0$. One can now see that the term in front of $\frac{\partial a}{\partial M}$ is negative. Therefore, one can obtain an upper bound for the $\frac{\partial a}{\partial M}\ln\pi_I$ by plugging in the result from Step 1 for $\frac{\partial a}{\partial M}$. Simplifying gives

$$\frac{\partial}{\partial M}\leq\frac{-2a\left(-2a^2(\gamma_d-\gamma_I M)+Ma(\gamma_d-\gamma_I(M+2))+\gamma_d a^3+\gamma_I M^2\right)}{\left(2a^2-2(M+1)a+M\right)\left(2\gamma_d Ma-\gamma_d a^2+\gamma_I M^2\right)}.$$

The denominator is positive due to $a<\overline{a}$. Simplifying the numerator gives

$$-2\gamma_I a\left(2Ma^2-(M+2)Ma+M^2\right)-2\gamma_d a\left(a^3-2a^2+Ma\right)<0.$$

Therefore one has $\frac{\partial}{\partial M}\ln\pi_I<0$ and the claim follows.

Step 5: $M=2$ is profit-maximizing among all possible values.

Assume there would be an $M'>2$ such that $M=M'$ gives a higher profit than $M=2$ in equilibrium. By Step 3, it must be the case that $a<\overline{a}$ for $M=M'$. Now it follows by Step 3 that having $M=2$ gives a higher profit for the investor than having $M=M'$. Thus, contacting only 2 dealers is indeed profit-maximizing.

□

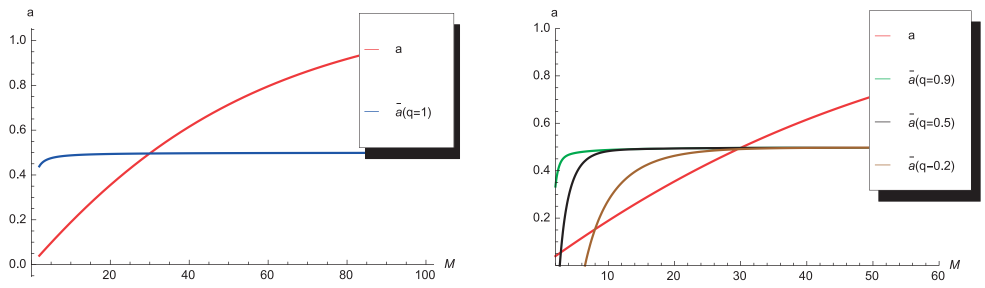
Proof of Proposition 4. It is shown in proposition 3 that equilibrium exists when $a<\bar{a}<\frac{1}{2}$. We now replace a as defined in equation (3.13): $\frac{M^2N\sigma_\theta^2+\gamma_d^2\sigma_\epsilon^2\sigma_\theta^2\sigma_W^2M+\gamma_d^2\sigma_\epsilon^4\sigma_W^2M}{M^2N\sigma_\theta^2+\gamma_d^2\sigma_\epsilon^2\sigma_\theta^2\sigma_W^2M+\gamma_d^2\sigma_\epsilon^4\sigma_W^2N}<\frac{1}{2}$. Equivalently,

$$M<\frac{1}{2}\frac{M^2N\sigma_\theta^2+\gamma_d^2\sigma_\epsilon^2\sigma_\theta^2\sigma_W^2M+\gamma_d^2\sigma_\epsilon^4\sigma_W^2N}{MN\sigma_\theta^2+\gamma_d^2\sigma_\epsilon^2\sigma_\theta^2\sigma_W^2+\gamma_d^2\sigma_\epsilon^4\sigma_W^2}<\frac{1}{2}N.$$

Therefore, if $M>\frac{1}{2}N$, one has $a\geq\bar{a}$, which implies that the equilibrium does not exist from proposition 3.

Figure C.1 – $a(M)$ and $\bar{a}(M,q)$

(a) $a(M)$ and $\bar{a}(M,q=1)$ (b) $a(M)$ and $\bar{a}(M,q=0.9,0.5,0.2)$



In the following, we show that the equilibrium existence condition $a < \bar{a}$ is equivalent to $M \in (\bar{M}_1, \bar{M}_2)$, where $\bar{M}_1$ and $\bar{M}_2$ are roots to the equation $a(M) = \bar{a}(M)$.

First, $a(M)$ is an increasing function of M and $\lim_{M\to N} a(M) = 1$. In terms of $\bar{a}(M,q)$, one can calculate the two derivatives

$$\frac{\partial \bar{a}}{\partial M} = 1 - \frac{2M + \frac{\kappa(1-\kappa)+\frac{M\partial\kappa}{\partial M}}{(1-\kappa)^2}}{\sqrt{4M^2+1+4M(1-2\Psi)}},$$

where $\frac{\partial\kappa}{\partial M} = \kappa\left[\frac{1}{M} + \frac{\ln(1-q)}{1-(1-q)^M}\right] < 0$, so $\frac{\partial\bar{a}}{\partial M} > 0$.

And

$$\frac{\partial \bar{a}}{\partial q} = -\frac{\frac{M}{(1-\kappa)^2}\frac{\partial\kappa}{\partial q}}{\sqrt{4M^2+1+4M[\frac{\kappa}{1-\kappa} - \frac{\sigma_\delta^2}{\sigma_\theta^2}]}},$$

where $\frac{\partial\kappa}{\partial q} = \frac{M(1-q)^{M-2}(1-Mq-(1-q)^M)}{(1-(1-q)^M)^2} < 0$, so $\frac{\partial\bar{a}}{\partial q} > 0$. Thus, $\bar{a}$ is an increasing function of both M and q.

Moreover, comparing the the value of $a(M)$ and $\bar{a}(M,q)$ at the limits, one gets

$$\begin{aligned}
\lim_{M\to N} a(M) = 1 \quad &> \quad \frac{1}{2} > \lim_{M\to N} \bar{a}(M,q),\\
\lim_{M\to 2} a(M) \quad &< \quad \lim_{M\to 2, q\to 1} \bar{a}(M,q),\\
\lim_{M\to 2} a(M) > 0 \quad &> \quad \lim_{M\to 2, q\to 0} \bar{a}(M,q).
\end{aligned}$$

So there are maximum two roots to the equation $a(M) = \bar{a}(M,q)$ for $M \in [2,N]$. As has been shown and demonstrated by figure (C.1) that there exists at least one root when $q = 1$, since

$\bar{a}(M,q)$ decreases when q decreases, the larger root $\bar{M}_2$ also decreases. Note that $\frac{\partial^2 \bar{a}(M,q)}{\partial q^2} < 0$, implies that the concavity of $\bar{a}(M,q)$ becomes larger , so the smaller root $\bar{M}_1$ increases when q decreases. More specifically,

(1) When $q = 1$, $\bar{M}_1 < 0$ and $\bar{M}_2 > 2$.

(2) When $q \in (\frac{1+\sqrt{1-\frac{2\left(a(2)^2-5a(2)+2\right)}{a(2)^2-5a(2)+4}}}{2}, 1]$, $\bar{M}_1 < 2$ and $\bar{M}_2 > 2$.

(3) When $q = \frac{1+\sqrt{1-\frac{2\left(a(2)^2-5a(2)+2\right)}{a(2)^2-5a(2)+4}}}{2}$, $\bar{M}_1 = 2$ and $\bar{M}_2 > 2$.

(4) When $q \in (\underline{q}, \frac{1+\sqrt{1-\frac{2\left(a(2)^2-5a(2)+2\right)}{a(2)^2-5a(2)+4}}}{2})$, $\bar{M}_1 > 2$ and $\bar{M}_2 > \bar{M}_1 > 2$.

(5) When $q = \underline{q}$, $\bar{M}_1 = N$, where $\underline{q}$ is the solution to the equation $\frac{Nq(1-q)^{N-1}}{1-(1-q)^N} = \frac{1}{2}$.

(6) When $q \in [0, \underline{q})$, there is no solution to $a(M) = \bar{a}(M)$ and $a(M) > \bar{a}(M)$.

The existence of equilibrium is summarized in figure (C.2).

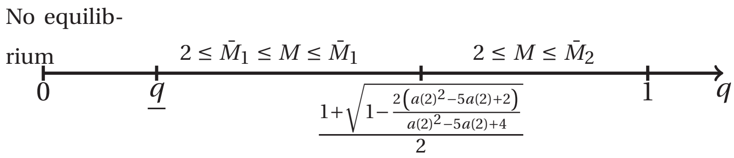


Figure C.2 – The existence of equilibrium

The above results show that when $q < \frac{1+\sqrt{1-\frac{2\left(a(2)^2-5a(2)+2\right)}{a(2)^2-5a(2)+4}}}{2}$, the equilibrium exists when $\bar{M}_1 < M < \bar{M}_2$. But $\bar{M}_1 \ge 2$, so there is no equilibrium when $M = 2$. Moreover, the minimum value of M such that the equilibrium exist is 3. Overall, we can conclude that when $q < \frac{1+\sqrt{1-\frac{2\left(a(2)^2-5a(2)+2\right)}{a(2)^2-5a(2)+4}}}{2}$, there exists an equilibrium with $M \ge 3$.

□

Proof of Proposition 5. Using the optimal demand schedule (3.6) for $M-1$ informed dealers and demand schedule (3.7) for the uninformed dealers, where the the conditional beliefs of uninformed dealers are formed as described in the proof of Proposition 1, the market clearing condition

$$q_k + \sum_{l \le M, l \neq k} q_l + \sum_{lM+1}^{N} = W$$

can be rearranged as

$$p_2 = b\left(W - \frac{(M-1)\xi}{\gamma_d \sigma_\xi^2} - q_k\right).$$

therefore, one has $\frac{\partial}{\partial q_k} p_2 = -b$, whenever dealer k is informed.

If dealers do not update their belief about ξ when the investor changes his demanded quantity x, taking the derivative of (3.15) w.r.t. x gives the following permanent price impact

$$\frac{\partial}{\partial x} p_c = -\frac{b^2}{2\gamma_d \sigma_\xi^2} - b < -b.$$

Using (3.15) and taking into account that the dealers form their expectation ξ about the dividend payment based on (3.4) and (3.21), one obtains

$$\begin{aligned}
\frac{\partial}{\partial x} p_c &= \frac{a\rho}{\alpha} - b - \frac{2(1-a)b\frac{\rho}{\alpha} + b^2}{2\gamma_d \sigma_\xi^2} \\
&= \frac{\rho\tilde{\beta}_1}{\alpha} + \tilde{\beta}_2 \\
&\geq \frac{\rho\tilde{\beta}_1}{\tilde{\alpha}} + \tilde{\beta}_2,
\end{aligned}$$

where $\tilde{\alpha}$, $\tilde{\beta}_1$ and $\tilde{\beta}_2$ denote the value of α, β_1 and β_2 when $q = 1$, respectively. The inequality holds since $\alpha = \tilde{\alpha} - \frac{\kappa}{1-\kappa}\frac{1}{2\tilde{\beta}_2 + \gamma_I \sigma_\epsilon^2}$. Replacing $\tilde{\alpha}$ by (3.25), one has

$$\begin{aligned}
\frac{\partial}{\partial x} p_c &\geq \frac{\rho\tilde{\beta}_1}{1-2\rho\tilde{\beta}_1}(2\tilde{\beta}_2 + \gamma_I \sigma_\epsilon^2) + \tilde{\beta}_2 \\
&> \frac{\rho\tilde{\beta}_1}{1-2\rho\tilde{\beta}_1} 2\tilde{\beta}_2 + \tilde{\beta}_2 \\
&= \frac{\tilde{\beta}_2}{1-2\rho\tilde{\beta}_1} \\
&= \frac{-b - \frac{b^2}{2\gamma_d \sigma_\xi^2}}{1 - 2\rho\left[a - \frac{(1-a)b}{\gamma_d \sigma_\xi^2}\right]} \\
&= \frac{-b\left(1 - \frac{a}{2M}\right)}{1 - 2\rho a\left(1 - \frac{1-a}{M}\right)} \\
&> -b.
\end{aligned}$$

The last inequality holds since $1 - \frac{a}{2M} > 1 - 2\rho a\left(1 - \frac{1-a}{M}\right)$, that is equivalent to $a > 0 > 1 - M + \frac{1}{4\rho}$. □

Proof of Proposition 6. By the conjecture (3.33), the market clearing price p_2 is jointly normally distributed with θ. One has

$$\mathrm{Cov}(D, p_2) = \mathrm{Cov}(\theta + \varepsilon, a\theta) = a\sigma_\theta^2. \tag{C.16}$$

Now (C.16), (3.37) and the normal projection theorem give

$$\mathbb{E}(D|p_2) = \frac{a\sigma_\theta^2}{a^2\sigma_\theta^2 + b^2\sigma_W^2} p_2 = \frac{\psi}{a}(a\theta + bW), \tag{C.17}$$

$$\mathbb{V}(D|p_2) = \frac{1}{\tau_u} = \sigma_\theta^2 + \sigma_\varepsilon^2 - \frac{a^2\sigma_\theta^4}{a^2\sigma_\theta^2 + b^2\sigma_W^2} = \sigma_\theta^2 + \sigma_\varepsilon^2 - \psi\sigma_\theta^2. \tag{C.18}$$

Plugging (C.17) and (C.18) into (3.34), using the result with (3.32) in the market-clearing condition (3.9):

$$\frac{M\tau_\epsilon(\theta - p_2)}{\gamma_d} + MX_k + \frac{(N-M)\tau_u(\psi\theta + \frac{\psi b}{a}W - p_2)}{\gamma_d} = W$$

solving for p_2 and matching coefficients with (3.33) yields

$$\begin{aligned} M\tau_\epsilon + (N-M)\psi\tau_u + \gamma_d\varphi\alpha_1 &= [M\tau_\xi + (N-M)\tau_u]a & \text{(C.19)} \\ (N-M)\tau_u\frac{\psi b}{a} - \gamma_d &= [M\tau_\xi + (N-M)\tau_u]b. & \text{(C.20)} \end{aligned}$$

Solving for a and b gives the expressions in (3.38) and (3.39).

It is immediately clear from (3.38) that $a > 0$, both numerator and denominator are always positive. Since $\psi > 0$ it follows also that $a \leq 1$ with a strict inequality only if $N = M$. □

Proof of Lemma 3. To show the second equality in (3.42), we note that the equilibrium price in the interdealer market depends on the aggregate inventory by market clearing. If market clearing holds, then

$$\sum_{l=1}^{M} q_l + \sum_{k=M+1}^{N} q_l + W,$$

where the demand schedules are defined as in (3.32) and (3.34). Using these definitions, the normal projection theorem to determine the conditional expectations gives and solving the previous equation for p_2 gives

$$p_2 = \frac{W - \sum_{l=1}^{M} X_l - \frac{\theta M}{\gamma_d \sigma_\varepsilon^2}}{\frac{a\sigma_\theta^2 (N-M)}{\gamma_d \left(a^2\sigma_\theta^2 + b^2\sigma_W^2\right)\left(-\frac{a^2\sigma_\theta^4}{a^2\sigma_\theta^2+b^2\sigma_W^2} + \sigma_\varepsilon^2 + \sigma_\theta^2\right)} - \frac{N-M}{\gamma_d\left(-\frac{a^2\sigma_\theta^4}{a^2\sigma_\theta^2+b^2\sigma_W^2} + \sigma_\varepsilon^2 + \sigma_\theta^2\right)} - \frac{M}{\gamma_d\sigma_\varepsilon^2}}.$$

Using the definition of a and b in Proposition 6, some algebra yields that the denominator on the right-hand side of the previous equation is equal to $\frac{1}{b}$. Therefore, it follows that

$$\frac{\partial}{\partial X_k} p_2 = -b. \tag{C.21}$$

Using $\mathbb{E}(p_2) = a\theta$, one can now calculate

$$\begin{aligned} & \frac{\partial}{\partial X_k}\mathbb{E}_k\left[D(q_k - X_k) - p_2 q_k - \frac{\gamma_d}{2}\sigma_\varepsilon^2 (q_k - X_k)^2\right] \\ = \quad & a\theta - b\frac{(1-a)\theta}{\gamma_d\sigma_\varepsilon^2} + bX_k. \end{aligned}$$

Since in equilibrium, one has $X_k = \frac{\varphi}{M}\alpha_1\theta$, the result follows. □

Proof of Proposition 7. Step 1: expressions of β_1 and β_2
Substituting equation $p_c(x_i)$ and $p_v(x)$ into the price $P(x)$ formula gives

$$\begin{aligned} \beta_1 &= \kappa\left(1 - \frac{\alpha_1}{\alpha_2}\right) + (1-\kappa)\left[a + \frac{b(1-a)}{\gamma_d\sigma_\epsilon^2} - \frac{b\varphi\alpha_1}{M}\right] && \text{(C.22)} \\ \beta_2 &= \kappa\left(\frac{1}{\alpha_2} - \frac{\gamma_I\sigma_\epsilon^2}{2}\right). && \text{(C.23)} \end{aligned}$$

Step 2: Solving α_1, α_2, β_1 and β_2
Combing the equations (3.47), (3.48), (C.22) and (C.23) and solving α_1, α_2, β_1 and β_2 leads to

the following:

$$\alpha_1 = \frac{1-a-\frac{b(1-a)}{\gamma_d\sigma_\epsilon^2}}{1-\frac{b\varphi}{M}\frac{1-2\kappa}{(1-\kappa)\gamma_I\sigma_\epsilon^2}}\frac{1-2\kappa}{(1-\kappa)\gamma_I\sigma_\epsilon^2}, \tag{C.24}$$

$$\alpha_2 = \frac{1-2\kappa}{(1-\kappa)\gamma_I\sigma_\epsilon^2}, \tag{C.25}$$

$$\beta_1 = \frac{a+\frac{b(1-a)}{\gamma_d\sigma_\epsilon^2}-\frac{b\varphi}{M}\frac{1-2\kappa}{(1-\kappa)\gamma_I\sigma_\epsilon^2}}{1-\frac{b\varphi}{M}\frac{1-2\kappa}{(1-\kappa)\gamma_I\sigma_\epsilon^2}}, \tag{C.26}$$

$$\beta_2 = \frac{\kappa}{2(1-2\kappa)}\gamma_I\sigma_\epsilon^2. \tag{C.27}$$

Note that when $\kappa < \frac{1}{2}$, one gets $\alpha_2 > 0$ and $\beta_2 > 0$.

Step 3: Show the existence of $\alpha_1 \in (0, \alpha_2]$ *and* $a \in (0, 1]$

We first show that α_1 is a decreasing function of a. Secondly, show that a is an increasing function of α_1, then prove that the two curves insect at $\{\alpha_1 \times a : (0, \alpha_2] \times (0, 1]\}$. Replacing b by equation (3.39) into formula (C.24) and derive the expression of α_1 as a function of a:

$$\alpha_1 = \frac{1}{2\gamma_d\varphi}\left[-M\tau_\epsilon - \gamma_d\varphi\alpha_2\left(\frac{a}{M}+a-1\right)+\sqrt{[M\tau_\epsilon+\gamma_d\varphi\alpha_2\left(\frac{a}{M}+a-1\right)]^2+4\gamma_d\varphi\alpha_2 M\tau_\epsilon(1-a)\left(\frac{a}{M}+1\right)}\right] \tag{C.28}$$

Once $a \le 1$, one could derive that $\alpha_1 > 0$ and further α_1 is monotonically decreasing on a. Since

$$\begin{aligned}\frac{\partial\alpha_1}{\partial a} &= \frac{\frac{\alpha_2}{2}(\frac{1}{M}+1)\left[M\tau_\epsilon+\gamma_d\varphi\alpha_2(\frac{a}{M}+a-1)-\sqrt{[M\tau_\epsilon+\gamma_d\varphi\alpha_2\left(\frac{a}{M}+a-1\right)]^2+4\gamma_d\varphi\alpha_2 M\tau_\epsilon(1-a)\left(\frac{a}{M}+1\right)}\right]}{\sqrt{[M\tau_\epsilon+\gamma_d\varphi\alpha_2\left(\frac{a}{M}+a-1\right)]^2+4\gamma_d\varphi\alpha_2 M\tau_\epsilon(1-a)\left(\frac{a}{M}+1\right)}}\\ &+ \frac{\alpha_2 M\tau_\epsilon(\frac{-2a}{M}+\frac{1}{M}-1)}{\sqrt{[M\tau_\epsilon+\gamma_d\varphi\alpha_2\left(\frac{a}{M}+a-1\right)]^2+4\gamma_d\varphi\alpha_2 M\tau_\epsilon(1-a)\left(\frac{a}{M}+1\right)}}\\ &< 0.\end{aligned}$$

Moreover, one has

$$\begin{aligned}\lim_{a\to 0}\alpha_1 &= \frac{1}{2\gamma_d\varphi}\left[-M\tau_\epsilon+\gamma_d\varphi\alpha_2+\sqrt{(M\tau_\epsilon-\gamma_d\varphi\alpha_2)^2+4\gamma_d\varphi\alpha_2 M\tau_\epsilon}\right] = \alpha_2,\\ \lim_{a\to 1}\alpha_1 &= \frac{1}{2\gamma_d\varphi}\left[-M\tau_\epsilon-\gamma_d\varphi\alpha_2+\sqrt{(M\tau_\epsilon+\gamma_d\varphi\alpha_2)^2}\right] = 0.\end{aligned}$$

Since α_1 is monotonically decreasing on a, one gets $\alpha_1 \in [0, \alpha_2)$.
In terms of a, one can rewrite a as a function of α_1 by substituting ψ and τ_u by equations

(3.37) and (3.36), and rearranging:

$$a = \frac{(M\tau_\epsilon + \gamma_d\varphi\alpha_1)^2(N\tau_\epsilon + \gamma_d\varphi\alpha_1) + \gamma_d^2\sigma_W^2(\tau_\theta + \tau_\epsilon)(M\tau_\epsilon + \gamma_d\varphi\alpha_1)}{(M\tau_\epsilon + \gamma_d\varphi\alpha_1)^2 N\tau_\epsilon + \gamma_d^2\sigma_W^2\tau_\epsilon(M\tau_\epsilon + N\tau_\theta)}$$

Next, one can compute the derivatives of a in terms of α_1 as

$$\begin{aligned}
\frac{\partial a}{\partial \alpha_1} &= \frac{(M\tau_\epsilon + \gamma_d\varphi\alpha_1)^4 N\tau_\epsilon + 2\gamma_d^2\sigma_W^2\tau_\epsilon(M\tau_\epsilon + \gamma_d\varphi\alpha_1)^2(M\tau_\epsilon + N\tau_\theta) + \gamma_d^4\sigma_W^4\tau_\epsilon(\tau_\theta + \tau_\epsilon)(M\tau_\epsilon + N\tau_\theta)}{\left[(M\tau_\epsilon + \gamma_d\varphi\alpha_1)^2 N\tau_\epsilon + \gamma_d^2\sigma_W^2\tau_\epsilon(M\tau_\epsilon + N\tau_\theta)\right]^2} \\
&+ \frac{\gamma_d^2\sigma_W^2(N-M)\tau_\epsilon^2(M\tau_\epsilon + \gamma_d\varphi\alpha_1)(M\tau_\epsilon + 2N\tau_\theta - \gamma_d\varphi\alpha_1)}{\left[(M\tau_\epsilon + \gamma_d\varphi\alpha_1)^2 N\tau_\epsilon + \gamma_d^2\sigma_W^2\tau_\epsilon(M\tau_\epsilon + N\tau_\theta)\right]^2} \\
&> 0.
\end{aligned}$$

Moreover, the values at the two bounds:

$$\begin{aligned}
\lim_{\alpha_1 \to 0} a &= \frac{M^2 N\tau_\epsilon^3 + M\gamma_d^2\sigma_W^2\tau_\epsilon(\tau_\theta + \tau_\epsilon)}{M^2 N\tau_\epsilon^3 + \gamma_d^2\sigma_W^2\tau_\epsilon(M\tau_\epsilon + N\tau_\theta)} < 1, \\
\lim_{\alpha_1 \to +\infty} a &= +\infty,
\end{aligned}$$

where the inequality above inequality follows from $M < N$. So, one gets that a is a monotonically increasing function of α_1 and $a \in (\frac{M^2 N\tau_\epsilon^3 + M\gamma_d^2\sigma_W^2\tau_\epsilon(\tau_\theta+\tau_\epsilon)}{M^2 N\tau_\epsilon^3 + \gamma_d^2\sigma_W^2\tau_\epsilon(M\tau_\epsilon + N\tau_\theta)}, +\infty)$

Since $\alpha_1(a)$ is monotonically decreasing on a and $\alpha_1 \in [0, \alpha_2)$, $a(\alpha_1)$ is monotonically increasing on α_1 and $a \in (\frac{M^2 N\tau_\epsilon^3 + M\gamma_d^2\sigma_W^2\tau_\epsilon(\tau_\theta+\tau_\epsilon)}{M^2 N\tau_\epsilon^3 + \gamma_d^2\sigma_W^2\tau_\epsilon(M\tau_\epsilon + N\tau_\theta)}, +\infty)$, by the fixed point theorem, there exists one unique solution (α_1^*, a^*) to the problem

$$\begin{cases} a(\alpha_1) = a \\ \alpha_1(a) = \alpha_1 \end{cases}$$

as demonstrated in figure(C.3).

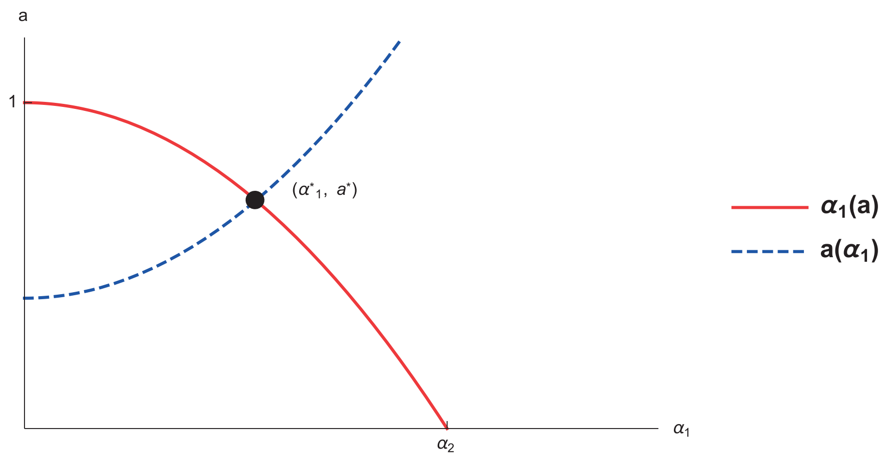


Figure C.3 – The curves of $\alpha_1(a)$ and $a(\alpha_1)$

Step 4: prove that $0 < \alpha_1^* \leq \alpha_2$ *and* $0 < a^* < 1$

First, it's obvious that $\alpha_1^* > 0$, we only need to prove that $\alpha_1^* \leq \alpha_2$. Suppose that $\alpha_1^* > \alpha_2$, then $a(\alpha_1^*) > a(\alpha_2)$, since a is an increasing function of α_1. Note that when $\alpha_1 = \alpha_2$, we have $\beta_1 = 1$ by equation(3.47), which implies that $a + \frac{b(1-a)}{\gamma_d \sigma_\epsilon^2} = 1$. Further, we get that either $a = 1$, or $a \neq 1$ and $b = \gamma_d \sigma_\epsilon^2$. But neither of the solution is consistent with the properties of α_1 and a function. On one hand, if $a = 1$, then $a(\alpha_1^*) > a(\alpha_2) = 1$, which is contrary to $a^* \leq 1$. On the other hand, if $b = \gamma_d \sigma_\epsilon^2$, then $a < 0$ by the equation (3.39), which is also contrary to $a > 0$. So $\alpha_1^* \leq \alpha_2$.

Next, we prove that $a^* < 1$. Assuming $a^* \geq 1$, then we should have $\alpha_1(a^*) \leq \alpha_1(1) = 0$, which is contrary to $\alpha_1^* > 0$. So $a^* < 1$. As in the equilibrium, $0 < \alpha_1 < \alpha_2$, one could get that $0 < \beta_1 < 1$ since $\alpha_1 = (1 - \beta_1)\alpha_2$.

Last, we verify that the second order condition of the maximization problem (3.45) is satisfied when $\kappa < \frac{1}{2}$:

$$-(2\beta_2 + \gamma_I \sigma_\epsilon^2) = -\left(\frac{\kappa}{1-2\kappa}\gamma_I \sigma_\epsilon^2 + \gamma_I \sigma_\epsilon^2\right) = -\frac{1-\kappa}{1-2\kappa}\gamma_I \sigma_\epsilon^2 < 0$$

The fact that Lemma 2 is applicable in order to derive the dealers' quoting strategies is proved as in the proof of Proposition 2.

Step 5: Show the equilibrium does not exist when $\kappa \geq \frac{1}{2}$

First, when $\kappa = \frac{1}{2}$, one has $\alpha_1 = 0$, $\alpha_2 = 0$ and $\beta_2 = +\infty$ from equation (C.24), (C.25) and (C.23). That is, the investor does not trade, and the price is not defined since $P(x_i) = \infty$. Thus, the equilibrium does not exist.

Second, when $\kappa > \frac{1}{2}$, one has $\beta_2 < 0$, and the second order condition of the optimization problem (3.45):

$$-(2\beta_2+\gamma_I\sigma_\epsilon^2) = -\Big(\frac{\kappa}{1-2\kappa}\gamma_I\sigma_\epsilon^2+\gamma_I\sigma_\epsilon^2\Big) = -\frac{1-\kappa}{1-2\kappa}\gamma_I\sigma_\epsilon^2 > 0.$$

This means that the investor's maximization problem does not have a solution and the equilibrium does not exist. □

Proof of Proposition 8. Claim 1: investors prefer to trade on the platform as $\sigma_\delta^2 \to 0$.

Equation (3.29) implies that each investor's ex-ante profits go to zero in the centralized market if $\sigma_\delta \to 0$. Proposition 7 states that an equilibrium exists if $\kappa < \frac{1}{2}$. All that is left to show is that expected profits for each investor remain strictly positive as $\sigma_\delta^2 \to 0$. This can be seen as follows. From the definition of the investors' utility (3.2), the dealers expected quotes conditional on a response on the platform (3.44) and the investors' equilibrium strategy (3.30) one obtains the following expression for the expected profit π_i of an investor trying to trade on the platform:

$$\pi_i = (1-(1-q)^M)\mathbb{E}\left[(\alpha_1\theta+\alpha_2\delta_i)\left(\theta+\delta-\beta_1\theta-(\alpha_1\theta+\alpha_2\delta_i)\left(\frac{\gamma_I\sigma_\varepsilon^2}{2}+\beta_2\right)\right)\right]. \tag{C.29}$$

As $\sigma_\delta^2 \to 0$, one obtains from (C.29) that

$$\begin{aligned}
\pi_i &\to (1-(1-q)^M)\alpha_1\sigma_\theta^2\left[(1-\beta_1)-\alpha_1\left(\frac{\gamma_i\sigma_\varepsilon^2}{2}+\beta_2\right)\right]\\
&= (1-(1-q)^M)\alpha_1\sigma_\theta^2\left[(1-\beta_1)-\alpha_1\frac{1}{2\alpha_2}\right]\\
&= (1-(1-q)^M)\alpha_1\sigma_\theta^2\left[(1-\beta_1)-\frac{1-\beta_1}{2}\right]\\
&> 0,
\end{aligned}$$

where the second line follows from the expressions for β_1 and α_2 in (C.24) and (C.27). The third line follows from the expressions for α_1, α_2 in (3.47) and (3.48). The inequality follows from $\alpha_1 > 0$ and (C.26), which implies $\beta_1 < 1$, since $b < 0$ and $a < 1$ by the proof of Proposition 7. This proves the first claim.

Claim 2: investors prefer to trade in the centralized market as $\sigma_\delta^2 \to \infty$.

Computing the expectation in (C.29) gives

$$\pi_i = (1-(1-q)^M)\left\{\underbrace{\alpha_1\sigma_\theta^2\left[(1-\beta_1)-\alpha_1\left(\frac{\gamma_i\sigma_\varepsilon^2}{2}+\beta_2\right)\right]}_{A}+\underbrace{\alpha_2\sigma_\delta^2\left[1-\alpha_2\left(\frac{\gamma_i\sigma_\varepsilon^2}{2}+\beta_2\right)\right]}_{B}\right\} \tag{C.30}$$

In the(C.30), A is not affected by σ_δ^2. Using the expressions for β_1 and α_2 in (C.24) and (C.27), one gets

$$B = \frac{1-2\kappa}{1-\kappa}\frac{\sigma_\delta^2}{2\gamma_I\sigma_\varepsilon^2} = \frac{1-2\kappa}{1-\kappa}\pi_i^c, \tag{C.31}$$

where π_i^c is the expected profit of the investor in the centralized market as defined in (3.29). It trivially follows that $\pi_i^c \to \infty$ as $\sigma_\delta^2 \to \infty$. Therefore, it follows that

$$\lim_{\sigma_\delta^2\to\infty}\frac{\pi_i}{\pi_i^c} = \lim_{\sigma_\delta^2\to\infty}(1-(1-q)^M)\frac{A+\frac{1-2\kappa}{1-\kappa}\pi_i^c}{\pi_i^c} = (1-(1-q)^M)\frac{1-2\kappa}{1-\kappa} < 1,$$

because of our assumption $\kappa > 0$. Therefore, investors will have a higher expected payoff in the centralized market as $\sigma_\delta^2 \to \infty$.

□

Proof of Proposition 9. We will show that the term denoted by A in (C.30) goes to zero as $\mu \to \infty$. Then it follows from (C.31) and $\kappa > 0$ that $\pi_i < \pi_i^c$ as $\mu \to \infty$, with $\pi_i < \pi_i^c$ defined as in (C.29) and (3.29).

In order to show $A \to 0$ as $mu \to \infty$, it is sufficient to show that $\alpha_1 \to 0$ as $mu \to \infty$, since β_2 is by (C.27) unaffected by μ and β_1 is by (C.26) between zero and one.

We show In order to show $A \to 0$ as $mu \to \infty$ as follows. Define the function $a(\cdot)$ as in the proof of Proposition 7. It has been shown in the proof of Proposition 7 that $\alpha_1 > 0$ for any $\mu > 0$ must hold in equilibrium. For any fixed $\alpha_1 > 0$, one has $a(\alpha_1) \to \infty$ for $\mu \to \infty$. The equilibrium condition $a(\alpha_1) = a < 1$ can only hold if $\alpha_1 \to 0$ for $\mu \to \infty$ (since $a(\cdot)$ is monotone increasing with $\lim \alpha_1 a(\alpha_1) \in (0,1)$). This proves the claim.

□

Proof of Proposition 10. Using (C.30), (C.31) and (3.29), one gets

$$\lim_{q \to 1} (\pi_i - \pi_i^c) = \lim_{q \to 1} A,$$

where A is defined as in (C.30). We proceed as in the proof of Proposition 8:

$$\begin{aligned}
\lim_{q \to 1} A &= \lim_{q \to 1} \alpha_1 \sigma_\theta^2 \left[(1 - \beta_1) - \alpha_1 \left(\frac{\gamma_i \sigma_\varepsilon^2}{2} + \beta_2 \right) \right] \\
&= \lim_{q \to 1} \alpha_1 \sigma_\theta^2 \left[(1 - \beta_1) - \alpha_1 \frac{1}{2\alpha_2} \right] \\
&= \lim_{q \to 1} \alpha_1 \sigma_\theta^2 \left[(1 - \beta_1) - \frac{1 - \beta_1}{2} \right] \\
&> 0,
\end{aligned}$$

where the second line follows from the expressions for β_1 and α_2 in (C.24) and (C.27). The third line follows from the expressions for α_1, α_2 in (3.47) and (3.48). The inequality follows from $\alpha_1 > 0$ and (C.26), which implies $\beta_1 < 1$ as $q \to 1$, since $b < 0$ and $a < 1$ by the proof of Proposition 7.

□

Bibliography

Tobias Adrian and Nina Boyarchenko. Intermediary Leverage Cycles and Financial Stability. *Federal Reserve Bank of New York Staff Reports*, 2012.

Tobias Adrian and Hyun Song Shin. Liquidity and leverage. *Journal of Financial Intermediation*, 19(3):418–437, 2010.

Fernando Alvarez, Andrew Atkeson, and Patrick J Kehoe. Money, Interest Rates, and Exchange Rates with Endogenously Segmented Markets. *Journal of Political Economy*, 110(1):73–112, 2002.

Yakov Amihud and Haim Mendelson. Dealership Market. Market-Making with Inventory. *Journal of Financial Economics*, 8(1):31–53, 1980.

Ronald W. Anderson and Jean-Pierre Danthine. The Time Pattern of Hedging and the Volatility of Futures Prices. *The Review of Economic Studies*, pages 249–266, 1983.

Daniel Andrei and Julien Cujean. Information Percolation, Momentum, and Reversal. *Journal of Financial Economics*, 123:617–645, 2017.

Andrew G. Atkeson, Andrea L. Eisfeldt, and Pierre-Olivier Weill. Entry and Exit in OTC Derivatives Markets. *Econometrica*, 83(6):2231–2292, 2015.

Ana Babus and Péter Kondor. Trading and Information Diffusion in Over-the-Counter Markets. *working paper*, 2016.

Ana Babus and Cecilia Parlatore. Strategic Fragmented Markets. *working paper*, 2017.

Steven Baker and Bryan Routledge. The Price of Oil Risk. *working paper*, 2016.

Gurdip Bakshi and Dilip Madan. Spanning and Derivative-Security Valuation. *Journal of Financial Economics*, 55(2):205–238, 2000.

Gurdip Bakshi, Nikunj Kapadia, and Dilip Madan. Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options. *The Review of Financial Studies*, 16(1):101–143, 2003.

Suleyman Basak and Domenico Cuoco. An Equilibrium Model with Restricted Stock Market Participation. *The Review of Financial Studies*, 11(2):309–341, 1998.

David S. Bates. Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options. *The Review of Financial Studies*, 9(1):69–107, 1996.

David S. Bates. Empirical Option Pricing: A Retrospection. *Journal of Econometrics*, 116(1-2): 387–404, 2003.

David S. Bates. The Market for Crash Risk. *Journal of Economic Dynamics and Control*, 32(7): 2291–2321, 2008.

Geert Bekaert and Eric Engstrom. Asset Return Dynamics under Habits and Bad Environment–Good Environment Fundamentals. *Journal of Political Economy*, 125(3):713–760, 2017.

Ben S. Bernanke and Mark Gertler. Agency Costs, Net Worth, and Business Fluctuations. *American Economic Review*, 79(1):pp. 14–31, 1989.

Geir Bjønnes, Carol L. Osler, and Dagfinn Rime. Asymmetric Information in the Interbank Foreign Exchange Market. *working paper*, 2008.

Geir Bjønnes, Neophytos Kathitziotis, and Carol L. Osler. Bid-Ask Spreads in OTC Markets. *working paper*, 2016.

Nicolas P. B. Bollen and Robert E. Whaley. Does Net Buying Pressure Affect the Shape of Implied Volatility Functions? *The Journal of Finance*, 59(2):711–753, 2004.

Tim Bollerslev, George Tauchen, and Hao Zhou. Expected Stock Returns and Variance Risk Premia. *The Review of Financial Studies*, 22(11):4463–4492, 2009.

Gregory W. Brown and Klaus Bjerre Toft. How Firms Should Hedge. *The Review of Financial Studies*, 15(4):1283–1324, 2002.

Markus K. Brunnermeier and Lasse Heje Pedersen. Market Liquidity and Funding Liquidity. *The Review of Financial Studies*, 22(6):2201–2238, 2009.

Markus K. Brunnermeier and Yuliy Sannikov. A Macroeconomic Model with a Financial Sector. *American Economic Review*, 104(2):379–421, 2014.

Andrea Buraschi and Jens Carsten Jackwerth. The Price of a Smile: Hedging and Spanning in Option Markets. *The Review of Financial Studies*, 14(2):495–527, 2001.

Andrea Buraschi and Alexei Jiltsov. Model Uncertainty and Option Markets with Heterogeneous Beliefs. *The Journal of Finance*, 61(6):2841–2897, 2006.

Kenneth Burdett and Kenneth L. Judd. Equilibrium Price Dispersion. *Econometrica*, 51(4): 955–969, 1983.

Peter Carr and Dilip Madan. *Towards a Theory of Volatility Trading*, page 458–476. Cambridge University Press, 2001.

Peter Carr and Liuren Wu. Variance Risk Premiums. *The Review of Financial Studies*, 22(3): 1311–1341, 2009.

Eric C. Chang. Returns to Speculators and the Theory of Normal Backwardation. *The Journal of Finance*, 40(1):193–208, 1985.

Hui Chen, Scott Joslin, and Sophie Ni. Demand for Crash Insurance, Intermediary Constraints, and Risk Premia in Financial Markets. *The Review of Financial Studies, forthcoming*.

Ing-Haw Cheng and Wei Xiong. Financialization of Commodity Markets. *Annual Review of Financial Economics*, 6(1):419–441, 2014.

Young-Hye Cho and Robert Engle. Modeling the Impacts of Market Activity on Bid-Ask Spreads in the Option Market. *NBER working paper*, 1999.

Peter Christoffersen, Ruslan Goyenko, Kris Jacobs, and Mehdi Karoui. Illiquidity Premia in the Equity Options Market. *The Review of Financial Studies, forthcoming*.

Pierre Collin-Dufresne, Benjamin Junge, and Anders B. Trolle. Market Structure and Transaction Costs of Index CDSs. *working paper*, 2017.

George M Constantinides and Darrell Duffie. Asset Pricing with Heterogeneous Consumers. *Journal of Political Economy*, 104(2):219, 1996.

George M Constantinides and Lei Lian. The Supply and Demand of S&P 500 Put Options. *working paper*, 2015.

Paul H. Cootner. Returns to Speculators: Telser versus Keynes. *Journal of Political Economy*, 68(4):396–404, 1960.

Paul H. Cootner. Speculation and Hedging. *working paper*, 1967.

Thomas E. Copeland and Dan Galai. Information Effects on the Bid-Ask Spread. *The Journal of Finance*, 38(5):1457–1469, 1983.

Martijn Cremers, Andy Fodor, and David Weinbaum. Where Do Informed Traders Trade First? Option Trading Activity, News Releases, and Stock Return Predictability. *working paper*, 2015.

Jon Danielsson, Hyun Song Shin, and Jean-Pierre Zigrand. Procyclical Leverage and Endogenous Risk. *working paper*, 2012.

Patrick De Fontnouvelle, Raymond P.H. Fishe, and Jeffrey H. Harris. The Behavior of Bid-Ask Spreads and Volume in Options Markets during the Competition for Listings in 1999. *The Journal of Finance*, 58(6):2437–2463, 2003.

M. Ángeles De Frutos and Carolina Manzano. Risk Aversion, Transparency, and Market Performance. *The Journal of Finance*, 57(2):959–984, 2002.

Emanuel Derman Michael Kamal Demeterfi, Kresimir and Joseph Zou. More Than You Ever Wanted to Know About Volatility Swaps. *Goldman Sachs Quantitative Strategies Research Notes*, 1999.

Douglas W. Diamond and Robert E. Verrecchia. Information Aggregation in a Noisy Rational Expectations Economy. *Journal of Financial Economics*, 9(3):221–235, 1981.

James S. Doran and Ehud I. Ronn. Computing the Market Price of Volatility Risk in the Energy Commodity Markets. *Journal of Banking and Finance*, 32(12):2541–2552, 2008.

Itamar Drechsler. Uncertainty, Time-Varying Fear, and Asset Prices. *The Journal of Finance*, 68 (5):1843–1889, 2013.

Itamar Drechsler and Amir Yaron. What's Vol Got to Do with It. *The Review of Financial Studies*, 24(1):1–45, 2011.

Songzi Du and Haoxiang Zhu. Bilateral Trading in Divisible Double Auctions. *Journal of Economic Theory*, 167:285–311, 2017.

Darrel Duffie, Piotr Dworczak, and Haoxiang Zhu. Benchmarks in Search Markets. *The Journal of Finance*, 72(5):1983–2044, 2017.

Darrell Duffie and Gustavo Manso. Information Percolation in Large Markets. *American Economic Review*, 97(2):203–209, 2007.

Darrell Duffie, Nicolae Garleanu, and Lasse Heje Pedersen. Over-the-Counter Markets. *Econometrica*, 73(6):1815–1847, 2005.

Darrell Duffie, Semyon Malamud, and Gustavo Manso. Information Percolation with Equilibrium Search Dynamics. *Econometrica*, 77(5):1513–1574, 2009.

Darrell Duffie, Semyon Malamud, and Gustavo Manso. Information Percolation in Segmented Markets. *Journal of Economic Theory*, 158(PB):838–869, 2015.

Philip H. Dybvig and Stephen A. Ross. Chapter 10 Arbitrage, State Prices and Portfolio Theory, 2003.

David Easley and Maureen O'Hara. Price, Trade Size, and Information in Securities Markets. *Journal of Financial Economics*, 19:69–90, 1987.

David Easley, Maureen O'hara, and Pulle Subrahmanya Srinivas. Option Volume and Stock Prices: Evidence on where Informed Traders Trade. *The Journal of Finance*, 53(2):431–465, 1998.

Chris Edmond and Pierre Olivier Weill. Aggregate Implications of Micro Asset Market Segmentation. *Journal of Monetary Economics*, 59(4):319–335, 2012.

Reinhard Ellwanger. The Tail Risk Premium in the Crude Oil Futures Market. *working paper*, 2015.

Erkko Etula. Broker-Dealer Risk Appetite and Commodity Returns. *Journal of Financial Econometrics*, 11(3):486–521, 2013.

Peter Feldhütter. The Same Bond at Different Prices: Identifying Search Frictions and Selling Pressures. *The Review of Financial Studies*, 25(4):1155–1206, 2005.

Mathieu Fournier and Kris Jacobs. Inventory Risk , Market-Maker Wealth , and the Variance Risk Premium : Theory and Evidence. *working paper*, 2016.

Gunter Franke, Richard C Stapleton, and Marti G Subrahmanyam. Who Buys and Who Sells Options: The Role of Options in an Economy with Background Risk. *Journal of Economic Theory*, 82(1):89–109, 1998.

Nicolae Gârleanu. Portfolio Choice and Pricing in Illiquid Markets. *Journal of Economic Theory*, 144:532–564, 2009.

Nicolae Gârleanu, Lasse Heje Pedersen, and Allen M. Poteshman. Demand-Based Option Pricing. *The Review of Financial Studies*, 22(10):4259–4299, 2009.

Gerald D. Gay, Jouahn Nam, and Marian Turac. How Firms Manage Risk: the Optimal Mix of Linear and Non-linear Derivatives. *Journal of Applied Corporate Finance*, 14(4):82–93, 2002.

Gerald D. Gay, Jouahn Nam, and Marian Turac. On the Optimal Mix of Corporate Hedging Instruments: Linear versus Nonlinear Derivatives. *Journal of Futures Markets*, 23(3):217–239, 2003.

Li Ge, Tse-Chun Lin, and Neil D. Pearson. Why does Option to Stock Volume Predict Stock Returns? *Journal of Financial Economics*, 1(217), 2015.

John Geanakoplos. The Leverage Cycle. *working paper*, 2010.

Thomas J. George and Francis A. Longstaff. Bid-Ask Spreads and Trading Activity in the S&P 100 Index Options Market. *Journal of Financial and Quantitative Analysis*, 28(3):381–397, 1993.

Vincent Glode and Christian Opp. Over-the-Counter vs. Limit-Order Markets: The Role of Traders' Expertise. *working paper*, 2017.

Lawrence R. Glosten and Paul R. Milgrom. Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. *Journal of Financial Economics*, 14(1): 71–100, 1985.

Itay Goldstein, Yan Li, and Liyan Yang. Speculation and Hedging in Segmented Markets. *The Review of Financial Studies*, 27(3):881–922, 2014.

Denis Gromb and Dimitri Vayanos. Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs. *Journal of Financial Economics*, 66(2-3):361–407, 2002.

Sanford J. Grossman and Joseph E. Stiglitz. On the Impossibility of Informationally Efficient Markets. *American Economic Review*, 70(3):393–408, 1980.

Björn Hagströmer and Albert J. Menkveld. A Network Map of Information Percolation. *working paper*, 2016.

Hau Harald, Hoffmann Peter, Langfield Sam, and Timmer Yannick. Discriminatory Pricing of Over-The-Counter FX Derivatives. *working paper*, 2017.

Zhiguo He and Arvind Krishnamurthy. Intermediary Asset Pricing. *American Economic Review*, 103(2):732–70, 2013.

Zhiguo He, Bryan Kelly, and Asaf Manela. Intermediary asset pricing: New evidence from many asset classes. *Journal of Financial Economics*, 126(1):1 – 35, 2017.

Martin F. Hellwig. On the Aggregation of Information in Competitive Markets. *Journal of Economic Theory*, 22(3):477–498, 1980.

Terrence Hendershott and Ananth Madhavan. Click or Call? Auction versus Search in the Over-the-Counter Market. *The Journal of Finance*, 70(1):419–447, 2015.

Steven Heston. A Closed-Form Solution for Options with Stochastic Volatility with Applications to Bond and Currency Options. *The Review of Financial Studies*, 6(2):327–343, 1993.

John Richard Hicks. *Value and Capital*. Oxford University Press, 1939.

David Hirshleifer. Residual Risk, Trading Costs, and Commodity Futures Risk Premia. *The Review of Financial Studies*, 1(2):173–193, 1988a.

David Hirshleifer. Risk, Futures Pricing, and the Organization of Production in Commodity Markets. *Journal of Political Economy*, 96(6):1206–1220, 1988b.

David Hirshleifer. Hedging Pressure and Futures Price Movements in a General Equilibrium Model. *Econometrica*, pages 411–428, 1990.

Steffen Hitzemann, Michael Hofmann, Marliese Uhrig-Homburg, and Christian Wagner. Margin Requirements and Equity Option Returns. *working paper*, 2016.

Thomas S.Y. Ho and Hans R. Stoll. The Dynamics of Dealer Markets Under Competition. *The Journal of Finance*, 38(4):1053–1074, 1983.

Harrison Hong and Motohiro Yogo. What does futures market interest tell us about the macroeconomy and asset prices? *Journal of Financial Economics*, 105(3):473–490, 2012.

Jianfeng Hu. Does Option Trading Convey Stock Price Information. *Journal of Financial Economics*, 111(3):625–645, 2014.

Maarten C.W. Janssen, José Luis Moraga-González, and Matthijs R. Wildenbeest. Truly costly sequential search and oligopolistic pricing. *International Journal of Industrial Organization*, 23:451–466, 2005.

Maarten C.W. Janssen, José Luis Moraga-González, and Matthijs R. Wildenbeest. Oligopolistic markets with sequential search and production cost uncertainty. *The RAND Journal of Economics*, 42(3):444–470, 2011.

Travis L. Johnson and Eric C. So. The Option to Stock Volume Ratio and Future Returns. *Journal of Financial Economics*, 106(2):262–286, 2012.

Boyan Jovanovic and Albert J. Menkveld. Dispersion and Skewness of Bid Prices. *working paper*, 2015.

Wenjin Kang, K. Geert Rouwenhorst, and Ke Tang. A Tale of Two Premiums: The Role of Hedgers and Speculators in Commodity Futures Markets. *Yale International Center for Finance Working Paper*, 2017.

John Maynard Keynes. Some Aspects of Commodity Markets. *Manchester Guardian Commercial: European Reconstruction Series*, 13:784–786, 1923.

Albert S. Kyle. Continuous Auctions and Insider Trading. *Econometrica*, 53(6):1315–1335, 1985.

Albert S. Kyle. Informed Speculation with Imperfect Competition. *The Review of Economic Studies*, 56(3):317–355, 1989.

Albert S. Kyle, Anna A. Obizhaeva, and Yajun Wang. Smooth Trading with Overconfidence and Market Power . *The Review of Economic Studies*, pages 1–52, 2017.

Ricardo Lagos and Guillaume Rocheteau. Liquidity in Asset Markets with Search Frictions. *Econometrica*, 77(2):403–426, 2009.

Ricardo Lagos, Guillaume Rocheteau, and Pierre-Olivier Weill. Crises and Liquidity in Over-the-Counter Markets. *Journal of Economic Theory*, 146:2169–2205, 2011.

Harvey Lapan, Giancarlo Moschini, and Steven D. Hanson. Production, Hedging, and Speculative Decisions with Options and Futures Markets. *American Journal of Agricultural Economics*, 73(1):66–74, 1991.

Benjamin Lester, Guillaume Rocheteau, and Pierre-Olivier Weill. Competing for Order Flow in OTC Markets. *Journal of Money, Credit and Banking*, 47(2):77–126, 2015.

Benjamin Lester, Ali Shourideh, Venky Venkateswaran, and Ariel Zetlin-Jones. Screening and adverse selection in frictional markets. *working paper*, 2017.

Robert H Litzenberger and Nir Rabinowitz. Backwardation in Oil Futures Markets: Theory and Empirical Evidence. *The Journal of Finance*, 50(5):1517–1545, 1995.

Jun Liu, Jun Pan, and Tan Wang. An Equilibrium Model of Rare-Event Premia and Its Implication for Option Smirks. *The Review of Financial Studies*, 18(1):131–164, 2005.

Semyon Malamud and Marzena Rostek. Decentralized exchange. *American Economic Review*, 107(11):3320–62, November 2017.

Semyon Malamud and Andreas Schrimpf. Intermediation Markups and Monetary Policy Passthrough. *working paper*, 2017.

Semyon Malamud, Michael C. Tseng, and Yuan Zhang. The Demand for Commodity Options. *working paper*, 2017.

Robert C. Merton. Option Pricing when Underlying Stock Returns are Discontinuous. *Journal of Financial Economics*, 3(1-2):125–144, 1976.

Mohamed Mnasri, Georges Dionne, and Jean-Pierre Gueyie. How Do Firms Hedge Risks? Empirical Evidence from US Oil and Gas Producers. *working paper*, 2013.

John Moore and Nobuhiro Kiyotaki. Credit Cycles. *Journal of Political Economy*, 105(2): 211–248, 1997.

Tobias J. Moskowitz, Yao Hua Ooi, and Lasse Heje Pedersen. Time Series Momentum. *Journal of Financial Economics*, 104(2):228–250, 2012.

Dmitriy Muravyev. Order Flow and Expected Option Returns. *The Journal of Finance*, 71(2): 673–708, 2016.

Narayan Y. Naik, Anthony Neuberger, and S. Viswanathan. Trade Disclosure Regulation in Markets with Negotiated Trades. *The Review of Financial Studies*, 12(4):873–900, 1999.

Anthony Neuberger. Realized Skewness. *The Review of Financial Studies*, 25(11):3424–3455, 2012.

David M.G. Newberry and Joseph E. Stiglitz. *The Theory of Commodity Price Stabilization*. Clarendon Press, 1981.

David M.G. Newbery. The Manipulation of Futures Markets by a Dominant Producer. *working paper*, 1983.

Marco Pagano and Ailsa Röell. Transparency and Liquidity: A Comparison of Auction and Dealer Markets with Informed Trading. *The Journal of Finance*, 51(2):579–611, 1996.

Emiliano Pagnotta and Thomas Philippon. Competing on Speed. *working paper*, 2011.

Jun Pan and Allen M. Poteshman. The Information in Option Volume for Future Stock Prices. *The Review of Financial Studies*, 19(3):871–908, 2006a.

Jun Pan and Allen M. Poteshman. The Information in Option Volume for Future Stock Prices. *The Review of Financial Studies*, 19(3):871–908, 2006b.

Jacques Rolfo. Optimal Hedging under Price and Quantity Uncertainty: The Case of a Cocoa Producer. *Journal of Political Economy*, 88(1):100–116, 1980.

Richard Roll, Eduardo Schwartz, and Avanidhar Subrahmanyam. O/S: The Relative Trading Activity in Options and Stock. *Journal of Financial Economics*, 96(1):1–17, 2010a.

Richard Roll, Eduardo Schwartz, and Avanidhar Subrahmanyam. O/S: The Relative Trading Activity in Options and Stock. *Journal of Financial Economics*, 96(1):1–17, 2010b.

Richard Roll, Eduardo Schwartz, and Avanidhar Subrahmanyam. Trading Activity in the Equity Market and Its Contingent Claims: An Mmpirical Investigation. *Journal of Empirical Finance*, 28:13–35, 2014.

Steve Ross. The Recovery Theorem. *The Journal of Finance*, 70(2):615–648, 2015.

Mark Rubinstein. The Valuation of Uncertain Income Streams and the Pricing of Options. *The Bell Journal of Economics*, 7(2):407–425, 1976.

Yuliy Sannikov and Andrzej Skrzypacz. Dynamic trading: Price inertia and front-running. *working paper*, 2016.

Andrei Shleifer and Robert Vishny. The Limits to Arbitrage. *The Journal of Finance*, 52:35–55, 1997.

Philip Stafford. ICAP's Future Reflects Derivatives Market in Transition . *Financial Times*, October 5 2016.

Dale O. Stahl. Oligopolistic Pricing with Sequential Consumer Search. *American Economic Review*, 79(4):700–712, 1989.

Jerome L. Stein. The Simultaneous Determination of Spot and Futures Prices. *American Economic Review*, 51(5):1012–1025, 1961.

Hans R. Stoll. Commodity Futures and Spot Price Determination and Hedging in Capital Market Equilibrium. *Journal of Financial and Quantitative Analysis*, 14(4):873–894, 1979.

Yeneng Sun. The exact law of large numbers via Fubini extension and characterization of insurable risks. *Journal of Economic Theory*, 126:31–69, 2006.

Anders B. Trolle and Eduardo S. Schwartz. Variance Risk Premia in Energy Commodities. *The Journal of Derivatives*, 17(3):15–32, 2010.

Anders B. Trolle and Eduardo S. Schwartz. The Swaption Cube. *The Review of Financial Studies*, 27(8):2307–2353, 2014.

Hal R. Varian. A Model of Sales. *American Economic Review*, 70(4):651–659, 1980.

Xavier Vives. *Information and Learning in Markets: The Impact of Market Microstructure*. Princeton University Press, Princeton, New Jersey, 2010.

Pierre-Olivier Weill. Leaning Against the Wind. *The Review of Economic Studies*, 74:1329–1354, 2007.

Joshua D. Woodard and Thomas W. Sproul. Identification of Hedging Pressure Effects in Commodity Futures-Options Markets: Implications for Price Discovery. *working paper*, 2016.

Xiankang Yin. A Comparison of Centralized and Fragmented Markets with Costly Search. *The Journal of Finance*, 60(3):1567–1590, 2005.

Bart Zhou Yueshen. Uncertain Market Making. *working paper*, 2017.

Haoxiang Zhu. Finding a Good Price in Opaque Over-the-Counter Markets. *The Review of Financial Studies*, 25(4):1255–1285, 2012.

Yuan Zhang

Swiss Finance Institute at EPFL
Extranef 129, UNIL-Dorigny
Lausanne 1015 Switzerland

Mobile: +41 78 635 72 39
Office: +41 21 693 20 56

Email: yuan.zhang@epfl.ch
Website: www.yuanzhangsfi.com

Research Interests

Market Microstructure, Financial Economics, Options Market Liquidity

Education

2012 - PhD Candidate in Finance, Swiss Finance Institute at EPF Lausanne, Switzerland
Advisor: Prof. Semyon Malamud

Spring 2016 Visiting PhD Student, MIT Sloan, sponsor: Prof. Hui Chen

2009 - 2011 MS in Financial Engineering, EPF Lausanne, Switzerland

2005 - 2009 BA in Financial Economics, Sichuan University, China

Job Market Paper

Imbalance-Based Option Pricing
I develop an equilibrium model of fragmented options markets in which option prices and bid-ask spreads are determined by the nonlinear risk imbalance between dealers and customers. In my model, dealers optimally exploit their market power and charge higher spreads for deep out-of-the-money (OTM) options, leading to an endogenous skew in both prices and spreads. In stark contrast to theories of price pressure in option markets, I show how wealth effects can make customers' net demand for options be negatively correlated with option prices. Under natural conditions, the skewness risk premium is positively correlated with the variance risk premium, consistent with the data.

Working Papers

The Demand for Commodity Options, with Semyon Malamud and Michael C. Tseng

Electronic Trading in OTC Markets vs. Centralized Exchange, with Ying Liu and Sebastian Vogel

Working in Progress

News-Implied Co-Movement, with Damir Filipovic, Semyon Malamud, and Asaf Manela

Macro Announcements and Market Liquidity, with Hasan Fallahgoul

Academic Experience

Swiss Finance Institute at EPF Lausanne

2015 -	Teaching Assistant for Prof. Anders Trolle, *Advanced Derivatives*
2014	Teaching Assistant for Prof. Loriano Mancini, *Econometrics*
2013	Teaching Assistant for Prof. Semyon Malamud, *Stochastic Calculus I* and Prof. Florian Pelgrin, *Financial Econometrics*

CEPREMAP, École Normale Supérieure

2011 - 2012	Research Assistant for Prof. Gabrielle Demange, *Financial Networks*

Paper Presentations

2016	MIT Sloan Finance Student Lunch Seminar
2014	SFI Research Days in Gerzensee

Awards, Grants and Fellowships

2016	Best Teaching Assistant Award, Master in Finnancial Engineering, EPFL
2014	Travel Grant, Pacific Institute for the Mathematical Sciences, UBC
2012	PhD Program Grant, Swiss Finance Institute

Other Professional Experience

2011	Finlab S.A., Geneva, Switzerland
2008	CMC Markets, Beijing, China

Other Information

Languages	Chinese (native), English (fluent)
Computer Skills	C/C++, CUDA, Latex, Mathematica, Matlab, Python, R

Lausanne, July 4, 2018