Imbalance-Based Option Pricing*

Yuan Zhang†

December 5, 2017

Abstract

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.

JEL Classification: G12, G13

Keywords: nonlinear risk imbalance, market fragmentation, market power, wealth effect, option demand, option pricing

*I would like to especially thank my supervisor Semyon Malamud and members of my dissertation committee, Pierre Collin-Dufresne and Julien Hugonnier for their helpful comments and advices. I also thank Hui Chen, Jerome Detemple, Ruediger Fahlenbrach, Zhiguo He, Leonid Kogan, Norman Schürhoff, and seminar participants at the SFI Academic Job Market Workshop, SFI Research Day and MIT Finance Student Lunch seminar for helpful comments. All errors are my own.

†Swiss Finance Institute, EPF Lausanne; Email: yuan.zhang@epfl.ch; Website: www.yuanzhangsfi.com.

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-tier 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. [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

[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)).

[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).

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.

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 trading in an almost competitive D2C exchange. With little market power, this 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, Joslin, and Ni, 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 and the variance risk implied from

[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, Pedersen, and Poteshman (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).

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 [7] are determined by the distribution of the nonlinear risk 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 nonlinear risk imbalance explains the second puzzle that the risk-neutral density departs from the physical density.

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, Joslin, and Ni (forthcoming) argue that dealers become more risk-averse [8] 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 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)

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

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

and extend their model to allow for an intermediate bargaining power of the dealers.[9]

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 (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, Shin, and Zigrand, 2012; Etula, 2013; Constantinides and Lian, 2015), equity financing constraints (Brunnermeier and Sannikov, 2014; He and Krishnamurthy, 2013; He, Kelly, and Manela, 2016), 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, Stapleton, and Subrahmanyam (1998); that is, the nonlinear risks render options non-redundant [10]. My assumption on fragmented markets [11] is borrowed from the literature on over-the-counter markets (e.g., Duffie, Garleanu, and Pedersen, 2005; Duffie, Malamud, and Manso, 2015; Atkeson, Eisfeldt, and Weill, 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.[12]

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, Fishe, and Harris (2003).

Fourth, option prices imply a skewed and fat-tailed risk-neutral distribution for the underlying returns (Buraschi and Jackwerth, 2001; Bakshi, Kapadia, and Madan, 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, Tauchen, and Zhou, 2009; Drechsler and Yaron, 2011; Drechsler, 2013; Bekaert and Engstrom, 2015). I take a different approach by showing how an implied volatility

[9] 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.

[10] Options are also non-redundant if agents have heterogeneous utilities (Bates, 2008; Baker and Routledge, 2016) or heterogeneous beliefs (Liu, Pan, and Wang, 2005; Buraschi and Jiltsov, 2006).

[11] Papers that have also assumed market fragmentation but with different trading protocols include Basak and Cuoco (1998), Edmond and Weill (2012), and Goldstein, Li, and Yang (2014). For endogenous market fragmentation, see Alvarez, Atkeson, and Kehoe (2002) and Babus and Parlatore (2016).

[12] 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.

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 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, Joslin, and Ni (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, Fodor, and Weinbaum (2015), Hu (2014), Muravyev (2016), and Malamud, Tseng, and Zhang (2017) show that the option order-imbalance measure is correlated with option premium and/or underlying returns, Pan and Poteshman (2006) show that the put-call ratio is correlated with next-period single equity returns, Roll, Schwartz, and Subrahmanyam (2010, 2014) and Ge, Lin, and Pearson (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.

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.

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[13] 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

[13]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.

of bilateral D2C exchanges [14], indexed by a pair (i, j) and a single D2D exchange. Throughout this paper, the following assumption is always present:

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:

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

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

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

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

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

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 [15], I assume that all agents in the economy share the same utility function U defined on an interval $(\underline{\mathrm{X}}, +\infty)$ with some $\underline{\mathrm{X}} \geq -\infty$, satisfying the standard Inada conditions $U'(\underline{\mathrm{X}}) = +\infty$, $U'(+\infty) = 0$. 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.

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

[14] 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.

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

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

$$\bar{\nu}_i \equiv \max_{\beta_i} \mathrm{E}[U(F_i^{\mathrm{C}}(X) + \beta_i(X - s\mathrm{e}^r))]\,. \tag{1}$$

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

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

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

$$\bar{\nu}_j \equiv \max_{G_j} \{\mathrm{E}[U(F_j^{\mathrm{D}}(X) + G_j(X))] : \mathrm{E}[M^{\mathrm{D2D}}(X)G_j(X)] = 0\}\,. \tag{2}$$

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}.

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 $M^{\mathrm{D2C}}_{(\mathrm{i,j})}$ 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 $G_{(\mathrm{i,j})}(M^{\mathrm{D2C}}_{(\mathrm{i,j})})$ to dealer j. Without loss of generality, I assume that $G_{(\mathrm{i,j})}$ satisfies $\mathrm{E}[M^{\mathrm{D2C}}_{(\mathrm{i,j})}G_{(\mathrm{i,j})}] = 0$.[16]

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

$$\nu_{(\mathrm{i,j})}[M^{\mathrm{D2C}}_{(\mathrm{i,j})}] \equiv \max_{G_{(\mathrm{i,j})}} \{\mathrm{E}[U(F_i^{\mathrm{C}} + G_{(\mathrm{i,j})})] : \mathrm{E}[M^{\mathrm{D2C}}_{(\mathrm{i,j})}G_{(\mathrm{i,j})}] = 0\}\,. \tag{3}$$

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:

$$\nu_{(\mathrm{j,i})}[M^{\mathrm{D2C}}_{(\mathrm{i,j})}] \equiv \max_{G_{(\mathrm{j,i})}} \{\mathrm{E}[U(F_j^{\mathrm{D}} - G^*_{(\mathrm{i,j})}(M^{\mathrm{D2C}}_{(\mathrm{i,j})}) + G_{(\mathrm{j,i})})] : \mathrm{E}[M^{\mathrm{D2D}}G_{(\mathrm{j,i})}] = 0\}\,, \tag{4}$$

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

Given dealer j's bargaining power θ, the pair bargain and choose the pricing kernel that solves

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

a version of the Nash bargaining problem, that is, maximizing the weighted surplus from trade, [17]

$$\max_{M^{\text{D2C}}_{(\text{i,j})}} (1-\theta)\log(\nu_{(\text{i,j})}[M^{\text{D2C}}_{(\text{i,j})}] - \bar{\nu}_i) + \theta \log(\nu_{(\text{j,i})}[M^{\text{D2C}}_{(\text{i,j})}] - \bar{\nu}_j)\,, \tag{5}$$

subject to the no-arbitrage constraints

$$0 = \mathrm{E}[M^{\text{D2C}}_{(\text{i,j})} X] - \mathrm{E}[M^{\text{D2D}} X]\,, \tag{6}$$

$$0 = \mathrm{E}[M^{\text{D2C}}_{(\text{i,j})}] - \mathrm{E}[M^{\text{D2D}}]\,. \tag{7}$$

Note that when dealers' bargaining power $\theta \neq 0$, the participation constraints $\nu_{(\text{i,j})} \geq \bar{\nu}_i$ and $\nu_{(\text{j,i})} \geq \bar{\nu}_j$ 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.*

2.5 Equilibrium

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

- $G^*_{(\text{i,j})}$ maximizes customer i's utility in (3),
- $G^*_{(\text{j,i})}$ maximizes dealer j's utility in (4) conditional on the customer's demand schedule $G^*_{(\text{i,j})}(M^{\text{D2C}}_{(\text{i,j})})$ as a function of the quoted pricing kernel $M^{\text{D2C}}_{(\text{i,j})}$,
- β^*_i maximizes customer i's autarky utility in (1),
- $\bar{G}_{(\text{j,i})}$ maximizes dealer j's autarky utility in (2),
- $M^{\text{D2C}}_{(\text{i,j})}$ maximizes the Nash bargaining protocol (5) given constraints (6), (7),
- the D2D market clears, $0 = \int_{[0,1]^2} G^*_{(\text{j,i})} didj$.

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.

[17] 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, I use the uncertainty of getting a competitive execution in a two-stage trading game to micro-found the trading protocol.

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

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

The Lagrangian for this problem is

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

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 $\mathrm{E}[MJ(\lambda_a M)] = \mathrm{E}[MF_a]$.

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 $\mathrm{E}^{\mathbb{Q}}$ to denote the corresponding expectation. I use $m_1^{\mathbb{P}} = \mathrm{E}[\log X]$, $m_1^{\mathbb{Q}} = \mathrm{E}^{\mathbb{Q}}[\log X]$ and

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

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*

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

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 $m_3^{\mathbb{Q}} = m_3^{\mathbb{P}} = 0$.

3.2 Economy with Frictions

At the D2C trading round, the Lagrangian for customer i's optimization problem (3) is

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

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

$$0 = \mathrm{E}[M_{(\mathrm{i,j})}^{\mathrm{D2C}} G_{(\mathrm{i,j})}^*]\,. \tag{8}$$

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

$$\nu_{(\mathrm{i,j})}[M_{(\mathrm{i,j})}^{\mathrm{D2C}}] = \mathrm{E}[U(J(\lambda_{(\mathrm{i,j})} M_{(\mathrm{i,j})}^{\mathrm{D2C}}))]. \tag{9}$$

At the D2D trading round, the Lagrangian for dealer j's optimization problem (4) is

$$\mathrm{E}[U(G_{(\mathrm{j,i})} + F_j^{\mathrm{D}} - G_{(\mathrm{i,j})}^*(M_{(\mathrm{i,j})}^{\mathrm{D2C}}))] - \lambda_{(\mathrm{j,i})} \mathrm{E}[M^{\mathrm{D2D}} G_{(\mathrm{j,i})}],$$

where again $\lambda_{(\mathrm{j,i})}$ 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 $G_{(\mathrm{j,i})}^* = J(\lambda_{(\mathrm{j,i})} M^{\mathrm{D2D}}) -$

$F_j^{\text{D}} + G^*_{(\text{i,j})}(M^{\text{D2C}}_{(\text{i,j})})$, where $\lambda_{(\text{j,i})}$ is determined by

$$0 \;=\; \text{E}[M^{\text{D2D}} G^*_{(\text{j,i})}]\,. \tag{10}$$

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

$$\nu_{(\text{j,i})}[M^{\text{D2C}}_{(\text{i,j})}] \;=\; \text{E}[U(J(\lambda_{(\text{j,i})} M^{\text{D2D}}))]\,. \tag{11}$$

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 (5). Its Lagrangian is

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

The second line consists of the no-arbitrage constraint for the risky asset (6) and that for the risk-free bond (7), where $\mu_{(\text{i,j}),\text{s}}$ and $\mu_{(\text{i,j}),\text{r}}$ are the corresponding Lagrange multipliers. I differentiate the Lagrangian function with respect to the D2C pricing kernel to get

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

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

$$\pi_{(\text{i,j})} \;\equiv\; \frac{\lambda_{(\text{i,j})}}{\lambda_{(\text{j,i})}} \frac{1-\theta}{\theta} \frac{\nu_{(\text{j,i})} - \bar{\nu}_j}{\nu_{(\text{i,j})} - \bar{\nu}_i}\,.$$

The first-order condition can be rewritten as [18]

$$0 \;=\; \pi_{(\text{i,j})} \frac{\delta \nu_{(\text{i,j})}[M^{\text{D2C}}_{(\text{i,j})}]}{\delta M^{\text{D2C}}_{(\text{i,j})}} + \frac{\delta \nu_{(\text{j,i})}[M^{\text{D2C}}_{(\text{i,j})}]}{\delta M^{\text{D2C}}_{(\text{i,j})}} - \mu_{(\text{i,j}),\text{s}} PX - \mu_{(\text{i,j}),\text{r}} P\,. \tag{12}$$

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

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

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 $\theta_{(\mathrm{i,j})}$. Due to the one-to-one mapping between $\theta_{(\mathrm{i,j})}$ and $\pi_{(\mathrm{i,j})}$, I can treat $\pi_{(\mathrm{i,j})}$ as exogenous and infer the bargaining parameter $\theta_{(\mathrm{i,j})}$, as well as other indirect utilities (1), (2), (9), and (11) after computing the equilibrium.

Given the relaxed and simplified D2C problem, I next compute the functional derivatives for the pricing kernel $M^{\mathrm{D2C}}_{(\mathrm{i,j})}$ in Appendix C and plug them into the relaxed first-order condition (12) to get

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

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

$$\kappa_{(\mathrm{i,j})} \;\equiv\; \frac{\mathrm{E}\left[J'(\lambda_{(\mathrm{i,j})}M^{\mathrm{D2C}}_{(\mathrm{i,j})})M^{\mathrm{D2C}}_{(\mathrm{i,j})}M^{\mathrm{D2D}}\right]}{\mathrm{E}\left[J'(\lambda_{(\mathrm{i,j})}M^{\mathrm{D2C}}_{(\mathrm{i,j})})(M^{\mathrm{D2C}}_{(\mathrm{i,j})})^2\right]}. \tag{13}$$

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

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,

$$0 \;=\; \iint_{[0,1]^2} G^{*}_{(\mathrm{j,i})}\, didj\,; \tag{14}$$

I arrive at the following theorem.

Theorem 1. *Suppose all agents have $U(X) = \log(X + c)$ [19]. Suppose also that an equilibrium exists. Then, for each pair (i,j), the state-by-state D2C pricing kernel is given by*

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

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

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 (8) (13), (6), (7), and (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

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

fragmented equilibrium, it is not trivial to provide a general condition such that the equilibrium exists. However, as long as the endowment function F_i^{C} is such that customer i' outside option (1) has an interior solution in the competitive equilibrium, then a unique fragmented equilibrium exists for a certain range of market competitiveness $\pi_{(\mathrm{i,j})}$ and its local uniqueness is given by the implicit function theorem.[20]

4 Examples

In this section, I provide numerical examples [21] and use them to show how my model can generate empirically observed patterns in option prices and trading volume.

4.1 Primitives

Table 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.

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. [22] I use this competitive equilibrium as the benchmark.

For ease of illustration, the nonlinear risk is specified as, [23]

$$F_J \;=\; (\mathrm{e}^r s - X)^2 \;,\; X \leq \mathrm{e}^r s.$$

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) [24], 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,

[20] 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.

[21] For details of computation, see appendix C.

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

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

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

Lemma 5. *For a continuous twice differentiable function $F(X)$ defined on $X \in [\underline{X}, \infty)$, the following representation holds,* [25]

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

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, $\mathrm{e}^r s$. 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, $\mathrm{e}^r s$, 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 4.1. *From the dealers' perspective, any convex nonlinear risk F_J is said to be skewed to the left if*

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

with the linear operator $\mathcal{M}[\cdot]$ defined in B.1, 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 $X \geq \mathrm{e}^r s$). 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 [26]; 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.

[25] 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.

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

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 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 $F_i^{\text{C}}(0)$ 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 $X - \sum_{i=S,B} F_i^{\text{C}}$. Each dealer therefore starts with endowment $F_j^{\text{D}}(0) = 0.3X + 0.6F_J$. 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 shows the implied volatility [27] 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 4.2. *The 'mid' pricing kernel is the wealth-weighted average pricing kernel among all D2C exchanges,* $\bar{M}^{\text{D2C}} = \iint_{[0,1]^2} \lambda_{(i,j)}^{-1} M_{(i,j)}^{\text{D2C}}\, didj / \iint_{[0,1]^2} \lambda_{(i,j)}^{-1}\, didj$.

In my model, customers' demand for options is proportional to their wealth, $\lambda_{(\text{i,j})}^{-1}$. 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 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,

[27] 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.

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.

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), 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,

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

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,

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

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.

4.3 Option Premia

Now we look at the correlation between customers' option demand and option risk premia. Figure 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 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 2, column two, shows that, after the reduction of dealers' market power, customers buy more options. At the same time, the variance risk premium becomes smaller (see Figure 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 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 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.

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, Joslin, and Ni (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 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.

4.4 Cost of Trading

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

Definition 4.3. *The cost of trading is defined as the effective percentage bid-ask spreads,*

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

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

Figure 4 and 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 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 4 and 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 exchanges, and customers respond by trading fewer options. This market power effect limits the risk sharing between dealers and customers.

Figure 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 4 and 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 of multi-listed options, likely due to improved competition. Similarly, in a recent study, Christoffersen, Goyenko, Jacobs, and Karoui (2017) 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.

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, Joslin, and Ni (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.

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 (2006), Gârleanu et al. (2009), Chen, Joslin, and Ni (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 [28]: 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.

[28] 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.

To test my model, I select a sample [29] 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.[30]

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 2) as in Bollen and Whaley (2004). Formally, I denote the moneyness bin as $\mathcal{B} = \text{OTM}, \text{DOTM}, \text{ATM}, \text{ITM}, \text{DITM}$.

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). [31]

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

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

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

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

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

[31] 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 [32]

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

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.

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

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,

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

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,

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

Intuitively, customers' net buy of OTM call options or net sell of OTM put options drives up the skew. [33]

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

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

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

[32] 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.

[33] 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.

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 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 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 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, Joslin, and Ni (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.

5.4 Correlation Puzzle

Table 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: [34] 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,

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

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 7 shows that for certain ETF options the results seem to align with my model's predic-

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

tion. 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 8 shows no particular relationship between the shape measure and the correlation between the realized variance and skewness.

Table 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 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 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.

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.

References

Adrian, T. and N. Boyarchenko (2012). Intermediary Leverage Cycles and Financial Stability.

Adrian, T. and H. S. Shin (2010). Liquidity and leverage. Journal of Financial Intermediation 19(3), 418–437.

Alvarez, F., A. Atkeson, and P. J. Kehoe (2002). Money, Interest Rates, and Exchange Rates with Endogenously Segmented Markets. Journal of Political Economy 110(1), 73–112.

Amihud, Y. and H. Mendelson (1980). Dealership Market. Market-Making with Inventory. Journal of Financial Economics 8(1), 31–53.

Atkeson, A. G., A. L. Eisfeldt, and P.-O. Weill (2015). Entry and Exit in OTC Derivatives Markets. Econometrica 83(6), 2231–2292.

Babus, A. and C. Parlatore (2016). Strategic Fragmented Markets.

Baker, S. and B. Routledge (2016). The Price of Oil Risk.

Bakshi, G., N. Kapadia, and D. Madan (2003). Stock Return Characteristics, Skew Laws, and the Differential Pricing of Individual Equity Options. Review of Financial Studies 16(1), 101–143.

Bakshi, G. and D. Madan (2000). Spanning and derivative-security valuation. Journal of Financial Economics 55(2), 205–238.

Basak, S. and D. Cuoco (1998). An Equilibrium Model with Restricted Stock Market Participation. Review of Financial Studies 11(2), 309–341.

Bates, D. S. (1996). Jumps and Stochastic Volatility: Exchange Rate Processes Implicit in Deutsche Mark Options. Review of Financial Studies 9(1), 69–107.

Bates, D. S. (2003). Empirical Option Pricing: A Retrospection.

Bates, D. S. (2008). The Market for Crash Risk. Journal of Economic Dynamics and Control 32(7), 2291–2321.

Bekaert, G. and E. Engstrom (2015). Asset Return Dynamics under Habits and Bad Environment Good Environment Fundamentals. Journal of Political Economy.

Bernanke, B. S. and M. Gertler (1989). Agency Costs, Net Worth, and Business Fluctuations. American Economic Review 79(1), pp. 14–31.

Bollen, N. P. B. and R. E. Whaley (2004). Does Net Buying Pressure Affect the Shape of Implied Volatility Functions? Journal of Finance 59(2), 711–753.

Bollerslev, T., G. Tauchen, and H. Zhou (2009). Expected Stock Returns and Variance Risk Premia. Review of Financial Studies 22(11), 4463–4492.

Brunnermeier, M. and Y. Sannikov (2014). A Macroeconomic Model with a Financial Sector. American Economic Review 104(2), 379–421.

Brunnermeier, M. K. and L. H. Pedersen (2009). Market Liquidity and Funding Liquidity. Review of Financial Studies 22(6), 2201–2238.

Buraschi, a. and J. C. Jackwerth (2001). The price of a smile: Hedging and spanning in option markets. Review of Financial Studies 14(2), 495–527.

Buraschi, A. and A. Jiltsov (2006). Model uncertainty and option markets with heterogeneous beliefs. Journal of Finance 61(6), 2841–2897.

Carr, P. and L. Wu (2009). Variance Risk Premiums. Review of Financial Studies 22(3), 1311–1341.

Chen, H., S. Joslin, and S. Ni. Demand for Crash Insurance, Intermediary Constraints, and Risk Premia in Financial Markets. Review of Financial Studies. forthcoming.

Cho, Y.-H. and R. Engle (1999). Modeling the Impacts of Market Activity on Bid-Ask Spreads in the Option Market. NBER Working Paper.

Christoffersen, P., R. Goyenko, K. Jacobs, and M. Karoui (2017). Illiquidity Premia in the Equity Options Market. The Review of Financial Studies.

Constantinides, G. M. and D. Duffie (1996). Asset Pricing with Heterogeneous Consumers. Journal of Political Economy 104(2), 219.

Constantinides, G. M. and L. Lian (2015). The Supply and Demand of S&P 500 Put Options.

Copeland, T. and D. Galai (1983). Information Effects on the Bid-Ask Spread. Journal of Finance 38(5), 1457–1469.

Cremers, M., A. Fodor, and D. Weinbaum (2015). Where Do Informed Traders Trade First? Option Trading Activity, News Releases, and Stock Return Predictability.

Danielsson, J., H. S. Shin, and J.-P. Zigrand (2012). Procyclical Leverage and Endogenous Risk.

De Fontnouvelle, P., R. P. Fishe, and J. H. Harris (2003). The Behavior of Bid-Ask Spreads and Volume in Options Markets during the Competition for Listings in 1999.

Drechsler, I. (2013). Uncertainty, Time-Varying Fear, and Asset Prices. Journal of Finance 68(5), 1843–1889.

Drechsler, I. and A. Yaron (2011). What's Vol Got to Do with It. Review of Financial Studies 24(1), 1–45.

Duffie, D., N. Garleanu, and L. H. Pedersen (2005). Over-the-Counter Markets. Econometrica 73(6), 1815–1847.

Duffie, D., S. Malamud, and G. Manso (2015). Information Percolation in Segmented Markets. Journal of Economic Theory 158(PB), 838–869.

Dybvig, P. H. and S. A. Ross (2003). Chapter 10 Arbitrage, State Prices and Portfolio Theory.

Easley, D. and M. O'Hara (1987). Price, Trade Size, and Information in Securities Markets. Journal of Financial Economics 19, 69–90.

Edmond, C. and P. O. Weill (2012). Aggregate Implications of Micro Asset Market Segmentation. Journal of Monetary Economics 59(4), 319–335.

Etula, E. (2013). Broker-Dealer Risk Appetite and Commodity Returns. Journal of Financial Econometrics 11(3), 486–521.

Fournier, M. and K. Jacobs (2016). Inventory Risk , Market-Maker Wealth , and the Variance Risk Premium : Theory and Evidence.

Franke, G., R. C. Stapleton, and M. G. Subrahmanyam (1998). Who Buys and Who Sells Options: The Role of Options in an Economy with Background Risk. Journal of Economic Theory 82(1), 89–109.

Gârleanu, N., L. H. Pedersen, and A. M. Poteshman (2009). Demand-Based Option Pricing. Review of Financial Studies 22(10), 4259–4299.

Ge, L., T.-C. Lin, and N. D. Pearson (2015). Why does Option to Stock Volume Predict Stock Returns? Journal of Financial Economics 1(217).

Geanakoplos, J. (2010). The Leverage Cycle.

George, T. J. and F. A. Longstaff (1993). Bid-Ask Spreads and Trading Activity in the S&P 100 Index Options Market. Journal of Financial and Quantitative Analysis 28(3), 381–397.

Glosten, L. R. and P. R. Milgrom (1985). Bid, Ask and Transaction Prices in a Specialist Market with Heterogeneously Informed Traders. Journal of Financial Economics 14(1), 71–100.

Goldstein, I., Y. Li, and L. Yang (2014). Speculation and Hedging in Segmented Markets. Review of Financial Studies 27(3), 881–922.

Gromb, D. and D. Vayanos (2002). Equilibrium and Welfare in Markets with Financially Constrained Arbitrageurs. Journal of Financial Economics 66(2-3), 361–407.

Harald, H., H. Peter, L. Sam, and T. Yannick (2017). Discriminatory Pricing of Over-The-Counter FX Derivatives.

He, Z., B. Kelly, and A. Manela (2016). Intermediary Asset Pricing: New Evidence from Many Asset Classes. Journal of Financial Economics.

He, Z. and A. Krishnamurthy (2013). Intermediary Asset Pricing. American Economic Review 103(2), 732–70.

Heston, S. (1993). 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.

Ho, T. S. and H. R. Stoll (1983). The Dynamics of Dealer Markets Under Competition. The Journal of Finance 38(4), 1053–1074.

Hu, J. (2014). Does Option Trading Convey Stock Price Information. Journal of Financial Economics 111(3), 625–645.

Kyle, A. S. (1985). Continuous Auctions and Insider Trading. Econometrica 53(6), 1315–1335.

Kyle, A. S. (1989). Informed Speculation with Imperfect Competition. Review of Economic Studies 56(3), 317–355.

Liu, J., J. Pan, and T. Wang (2005). An Equilibrium Model of Rare-Event Premia and Its Implication for Option Smirks. Review of Financial Studies 18(1), 131–164.

Malamud, S. and A. Schrimpf (2017). Intermediation Markups and Monetary Policy Passthrough. Swiss Finance Institute Research Paper No. 16-75.

Malamud, S., M. Tseng, and Y. Zhang (2017). The Demand for Commodity Options.

Merton, R. C. (1976). Option Pricing when Underlying Stock Returns are Discontinuous. Journal of Financial Economics 3(1-2), 125–144.

Moore, J. and N. Kiyotaki (1997). Credit Cycles. Journal of Political Economy 105(2), 211–248.

Muravyev, D. (2016). Order Flow and Expected Option Returns. Journal of Finance 71(2), 673–708.

Neuberger, A. (2012). Realized Skewness. Review of Financial Studies 25(11), 3424–3455.

Pan, J. and A. M. Poteshman (2006). The Information in Option Volume for Future Stock Prices. Review of Financial Studies 19(3), 871–908.

Roll, R., E. Schwartz, and A. Subrahmanyam (2010). O/S: The Relative Trading Activity in Options and Stock. Journal of Financial Economics 96(1), 1–17.

Roll, R., E. Schwartz, and A. Subrahmanyam (2014). Trading Activity in the Equity Market and Its Contingent Claims: An Mmpirical Investigation. Journal of Empirical Finance 28, 13–35.

Ross, S. (2015). The Recovery Theorem. Journal of Finance 70(2), 615–648.

Rubinstein, M. (1976). The Valuation of Uncertain Income Streams and the Pricing of Options. The Bell Journal of Economics 7(2), 407–425.

Shleifer, A. and R. Vishny (1997). The Limits to Arbitrage. Journal of Finance 52, 35–55.

Trolle, A. B. and E. S. Schwartz (2010). Variance Risk Premia in Energy Commodities. Journal of Derivatives 17(3), 15–32.

Trolle, A. B. and E. S. Schwartz (2014). The Swaption Cube. Review of Financial Studies 27(8), 2307–2353.

A 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 (12), with $\pi_{(\text{i,j})}$ being endogenized as the Lagrange multiplier of the customer's participation constraint.

B Expansion

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

Definition B.1. *The linear operator on a nonlinear risk function F is defined as*

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

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

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, (6) and (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 4. *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 (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 (8), (13), (7):

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

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

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

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*

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

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* [35]

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

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

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 10.1. *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,*

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

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*

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

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 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* $X_1, X_2 \in [X_{\min}, X_{\max}]$; [36]
- *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* $X \in [X^*, X_{\max}]$;

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

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 $\mathrm{E}[\mathcal{M}[F]] = 0$, there are two solutions $X_1, X_2 \in [X_{\min}, X_{\max}]$ to the following equation

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

□

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

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

C Proof

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

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

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

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

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 (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,

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

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

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

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

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

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

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

□

Proof of Proposition 2. From Lemma 10, there exists a point K^*, such that $\int_0^{K^*}\mathcal{M}[F]\,\mathrm{dX}=0$. 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

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

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.

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

□

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

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

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 $\hat{X}\gg \mathrm{e}^r s^{(0)}$, 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

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

Aggregate over all customers to get,

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

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

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

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,

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

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

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

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

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

Plug-in the definitions to get

$$M^{(0)} + \epsilon_F \left(\frac{w_D^{(0)}}{(s^{(0)} + 2c\mathrm{e}^{-r})^2 - (w_D^{(0)})^2} \mathcal{M} \left[F_J^{(1)} \right] \right) .$$

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

$$\epsilon_\pi \frac{|\mathrm{E}[(M^{(1)} - \bar{M}^{(1)}) O(K)]|}{\mathrm{E}[M^{(0)} O(K)]} + \mathcal{O}(\epsilon_\pi^2) .$$

Plug-in the definitions to get the difference in the pricing kernel,

$$\left(-\frac{1}{w^{(0)}} \mathcal{M}[F] + \frac{1}{s^{(0)} + 2c\mathrm{e}^{-r} - w_D^{(0)}} \mathcal{M}[-F_J] \right) \pi^{(1)}$$

When the nonlinear risk is 'small' and dealers are monopolists, the spreads are given by

$$\epsilon_F \frac{|\mathrm{E}[(M^{(1)} - \bar{M}^{(1)}) O(K)]|}{\mathrm{E}[M^{(0)} O(K)]} + \mathcal{O}(\epsilon_F^2) .$$

Plug-in the definitions to get the difference in the pricing kernel,

$$\frac{1}{2}\left(-\frac{1}{w^{(0)}}\mathcal{M}\left[F^{(1)}\right]+\frac{1}{s^{(0)}+2c\mathrm{e}^{-r}-w_D^{(0)}}\mathcal{M}\left[-F_J^{(1)}\right]\right).$$

□

Proof of Proposition 9. The endogenous parameters are solved explicitly, [37]

$$\begin{aligned}
w^{(0)} &= \alpha s^{(0)}+c\mathrm{e}^{-r}\,,\\
\kappa^{(0)} &= 1\,,\\
\mu^{(0)} &= 0\,,\\
s^{(0)} &= \mathrm{E}[M^{(0)}X]\,,\\
w_D^{(0)} &= \alpha_D s^{(0)}+c\mathrm{e}^{-r}\,.
\end{aligned}$$

□

Proof of Proposition 10. First-order expand on the first-order condition (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 $\pi^{(0)}=\pi^{(1)}=0$.

$$\begin{aligned}
0 = {} & \frac{1}{(M^{(0)})^2}\left(M^{(0)}w^{(1)}-2M^{(1)}w^{(0)}+N^{(1)}w^{(0)}\right)\\
& -F^{(0)}\kappa^{(1)}-F^{(1)}-X\mu^{(1)}-\kappa^{(1)}c+\mathrm{e}^r\mu^{(1)}s^{(0)}\,.
\end{aligned}$$

[37] I define $\alpha_D \equiv 1-\iint_{[0,1]^2}\alpha\, didj$.

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

$$w^{(1)} \;=\; \mathrm{E}[F^{(0)} M^{(1)} + F^{(1)} M^{(0)} + M^{(1)} c] \,.$$

From the no-arbitrage conditions, we know that $\mathrm{E}[M^{(1)}] = 0$ and $\mathrm{E}[M^{(1)} X] = s^{(1)}$. Hence

$$w^{(1)} \;=\; \alpha s^{(1)} + \mathrm{E}[F^{(1)} M^{(0)}] \,.$$

The definition for the endogenous parameter κ implies

$$\kappa^{(1)} \;=\; -\mathrm{E}\left[\frac{1}{M^{(0)}} \left(M^{(1)} - N^{(1)} \right) \right] .$$

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 $\mathrm{E}[M^{(1)}] = \mathrm{E}[N^{(1)}] = 0$, I get

$$\mu^{(1)} \;=\; \frac{\mathrm{E}[M^{(0)} w^{(1)} - (M^{(0)})^2 F^{(1)}]}{\mathrm{E}[(M^{(0)})^2 (X - \mathrm{e}^r s^{(0)})]} \,.$$

Under the D2D market clearing condition

$$0 \;=\; \left(M^{(0)} w_D^{(1)} - N^{(1)} w_D^{(0)} \right) + \iint_{[0,1]^2} \left(M^{(0)} w^{(1)} - M^{(1)} w^{(0)} \right) didj \,.$$

Take expectation to get

$$w_D^{(1)} \;=\; -\iint_{[0,1]^2} w^{(1)} \, didj \,.$$

Multiply by X and take expectation on the D2D market clearing condition to get

$$s^{(1)} \;=\; 0 \,.$$

Then solve for the D2D pricing kernel,

$$\begin{aligned} N^{(1)} = \; & \left((1 + \alpha_D) \, s^{(0)} + 3c\mathrm{e}^{-r} \right)^{-1} \\ & \times \iint_{[0,1]^2} (M^{(0)})^2 (F^{(1)} + \mu^{(1)} (X - \mathrm{e}^r s^{(0)})) - M^{(0)} w^{(1)} \, didj \,. \end{aligned}$$

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

$$\begin{aligned}
w^{(1)} &= \mathrm{E}\left[F^{(1)} M^{(0)} \right] , \\
\kappa^{(1)} &= 0 , \\
\mu^{(1)} &= \frac{1}{\mathrm{e}^r s^{(0)} + 2c} \frac{\mathrm{Cov}\left[M^{(0)}, M^{(0)} F^{(1)} \right]}{\mathrm{Var}\left[M^{(0)} \right]} , \\
s^{(1)} &= 0 . \\
w_D^{(1)} &= - \iint_{[0,1]^2} w^{(1)} \, didj .
\end{aligned}$$

□

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

$$w^{(1)} \;=\; \mathrm{E}[F M^{(1)}] .$$

From the no-arbitrage condition, I get

$$\mu^{(1)} \;=\; \pi^{(1)} \frac{\mathrm{E}[(M^{(0)})^2 (F + c)] - \mathrm{E}[M^{(0)}] w^{(0)}}{\mathrm{E}[(M^{(0)})^2 X] - \mathrm{E}[(M^{(0)})^2] \mathrm{e}^r s^{(0)}} .$$

From the D2D market-clearing condition

$$\begin{aligned}
N^{(1)} = & (s^{(0)} + 2c\mathrm{e}^{-r})^{-1} \\
& \times \left(- \iint_{[0,1]^2} (M^{(0)})^2 \left((F + c)\pi^{(1)} - \mu^{(1)} (X - \mathrm{e}^r s^{(0)}) \right) - w^{(0)} M^{(0)} \pi^{(1)} \, didj \right) .
\end{aligned}$$

The endogenous parameters are given by

$$\begin{aligned}
w^{(1)} &= \mathrm{E}[FM^{(1)}]\,,\\
\kappa^{(1)} &= 0\,,\\
\mu^{(1)} &= \frac{\pi^{(1)}}{\mathrm{e}^r s^{(0)} + 2c}\frac{\mathrm{Cov}[M^{(0)}, M^{(0)}(F+c)]}{\mathrm{Var}[M^{(0)}]}\,,\\
s^{(1)} &= 0\,.\\
w_D^{(1)} &= -\iint_{[0,1]^2} w^{(1)}\, didj\,.
\end{aligned}$$

□

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*

$$\begin{aligned}
\frac{\delta \nu_{(i,j)}[M^{\text{D2C}}_{(i,j)}]}{\delta M^{\text{D2C}}_{(i,j)}} &= -P\lambda_{(i,j)}(J(\lambda_{(i,j)}M^{\text{D2C}}_{(i,j)}) - F_i^{\text{C}}),\\
\frac{\delta \nu_{(j,i)}[M^{\text{D2C}}_{(i,j)}]}{\delta M^{\text{D2C}}_{(i,j)}} &= P\lambda_{(j,i)}\kappa_{(i,j)}(J(\lambda_{(i,j)}M^{\text{D2C}}_{(i,j)}) - F_i^{\text{C}})\\
&\quad + P\lambda_{(j,i)}\lambda_{(i,j)}J'(\lambda_{(i,j)}M^{\text{D2C}}_{(i,j)})(\kappa_{(i,j)}M^{\text{D2C}}_{(i,j)} - M^{\text{D2D}}).
\end{aligned}$$

where $\kappa_{(i,j)}$ is defined in (13).

Proof. Direct calculation yields the results. □

Lemma 13. *Dealer j's profit from D2C trading is $-\mathrm{E}[M^{\text{D2D}}G^*_{(i,j)}(M^{\text{D2C}}_{(i,j)})]$, and it is decreasing in the inverse market power $\pi_{(i,j)}$.*

Proof. From the Nash bargaining first-order condition, I get

$$G^*_{(\mathrm{i,j})} \;=\; -\frac{\lambda_{(\mathrm{i,j})}J'(\lambda_{(\mathrm{i,j})}M^{\text{D2C}}_{(\mathrm{i,j})})(\kappa_{(\mathrm{i,j})}M^{\text{D2C}}_{(\mathrm{i,j})} - M^{\text{D2D}})}{\kappa_{(\mathrm{i,j})} - \pi_{(\mathrm{i,j})}}\,.$$

Taking expectations and plug-in the definition to get

$$-\mathrm{E}[M^{\text{D2D}}G^*_{(\text{i,j})}] \;=\; \lambda_{(\text{i,j})}\frac{\mathrm{E}[M^{\text{D2C}}_{(\text{i,j})}M^{\text{D2D}}J'(\lambda_{(\text{i,j})}M^{\text{D2C}}_{(\text{i,j})})]^2-\mathrm{E}[(M^{\text{D2D}})^2J'(\lambda_{(\text{i,j})}M^{\text{D2C}}_{(\text{i,j})})]\mathrm{E}[(M^{\text{D2C}}_{(\text{i,j})})^2J'(\lambda_{(\text{i,j})}M^{\text{D2C}}_{(\text{i,j})})]}{(\kappa_{(\text{i,j})}-\pi_{(\text{i,j})})\mathrm{E}[M^{\text{D2C}}_{(\text{i,j})}J'(\lambda_{(\text{i,j})}M^{\text{D2C}}_{(\text{i,j})})]} \;>\; 0$$

The inequality comes from Holder inequality. Also, dealer will agree to trade only when this term is positive, hence, I need that $\kappa_{(\text{i,j})} > \pi_{(\text{i,j})}$. □

D 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

$$\text{Variance}^{\mathbb{Q}}_t(T) \;=\; \frac{\mathrm{e}^{r(T-t)}V_t(T)-\mu_t(T)^2}{T-t}$$

The risk-neutral skewness is given by

$$\text{Skew}^{\mathbb{Q}}_t(T) \;=\; \frac{\mathrm{e}^{r(T-t)}W_t(T)-3\mu_t(T)\mathrm{e}^{r(T-t)}V_t(T)+2\mu_t(T)^3}{\left(\mathrm{e}^{r(T-t)}V_t(T)-\mu_t(T)^2\right)^{3/2}}\,.$$

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

$$\mu_t(T)\;\approx\;\mathrm{e}^{r(T-t)}-1-\frac{\mathrm{e}^{r(T-t)}}{2}V_t(T)-\frac{\mathrm{e}^{r(T-t)}}{6}W_t(T)-\frac{\mathrm{e}^{r(T-t)}}{24}X_t(T)\,.$$

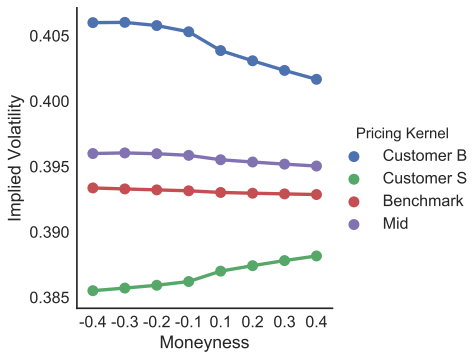

Figure 1: **Implied volatility for D2C exchanges.** Moneyness is defined as the log $\frac{K}{e^r s}$. This example uses parameters in Table 1 and does not include any shocks.

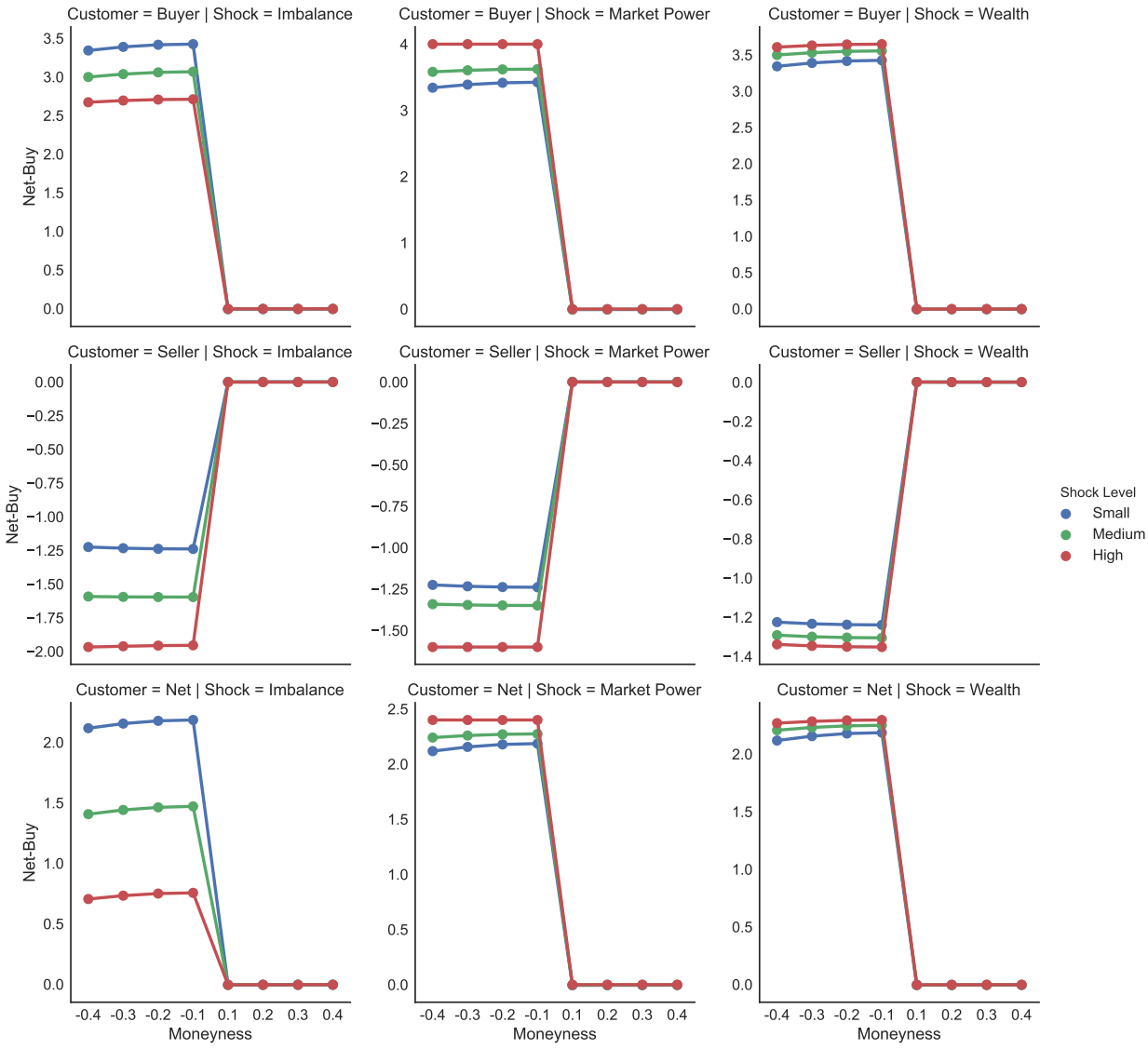

Figure 2: **Effects of 'macro' shocks on option demand.**

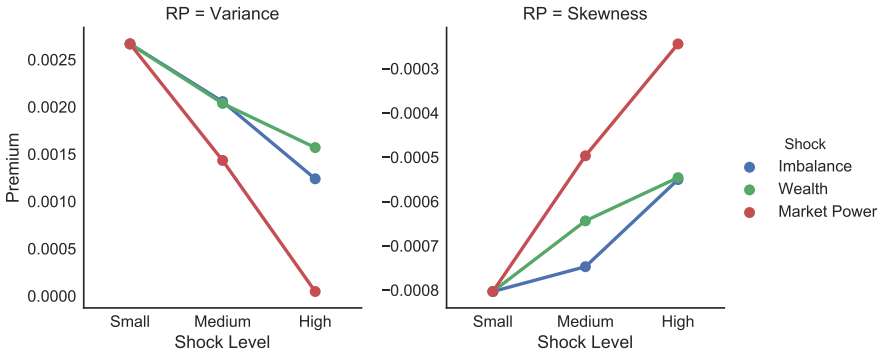

Figure 3: **Effects of 'macro' shocks on option risk premium.**

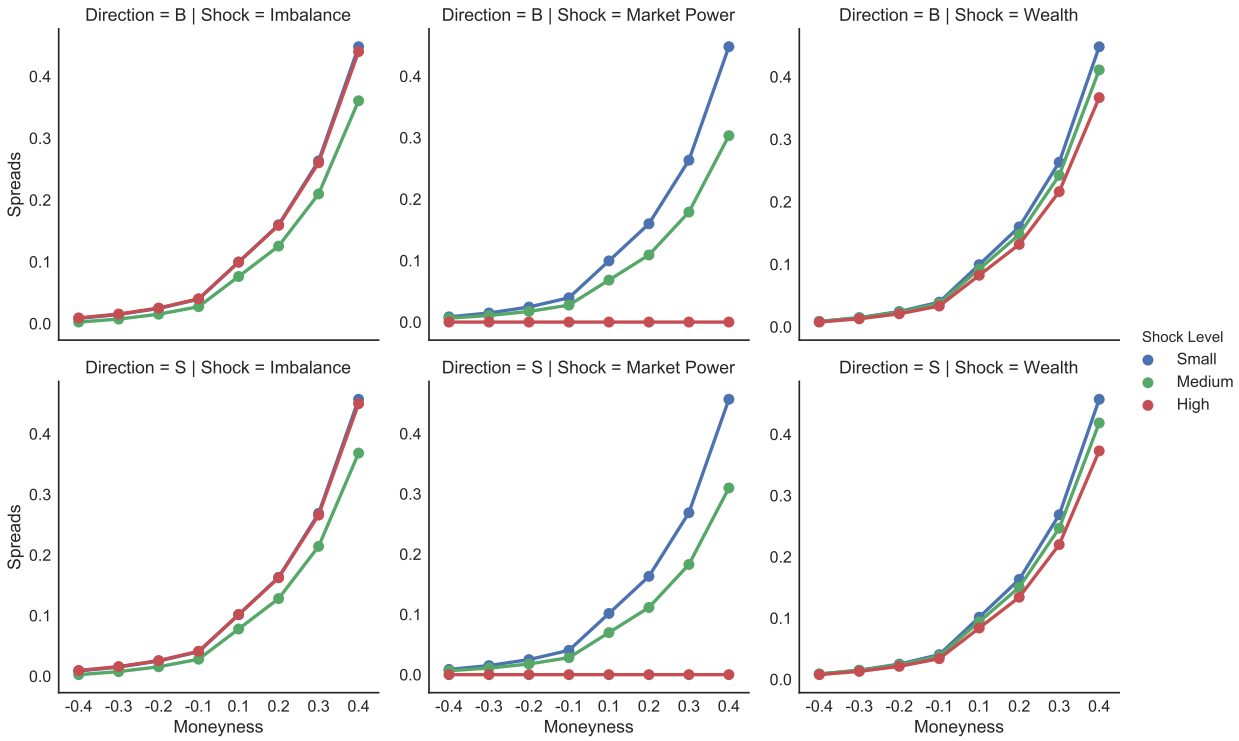

Figure 4: **Effects of 'macro' shocks on effective percentage bid-ask spreads for call options.**

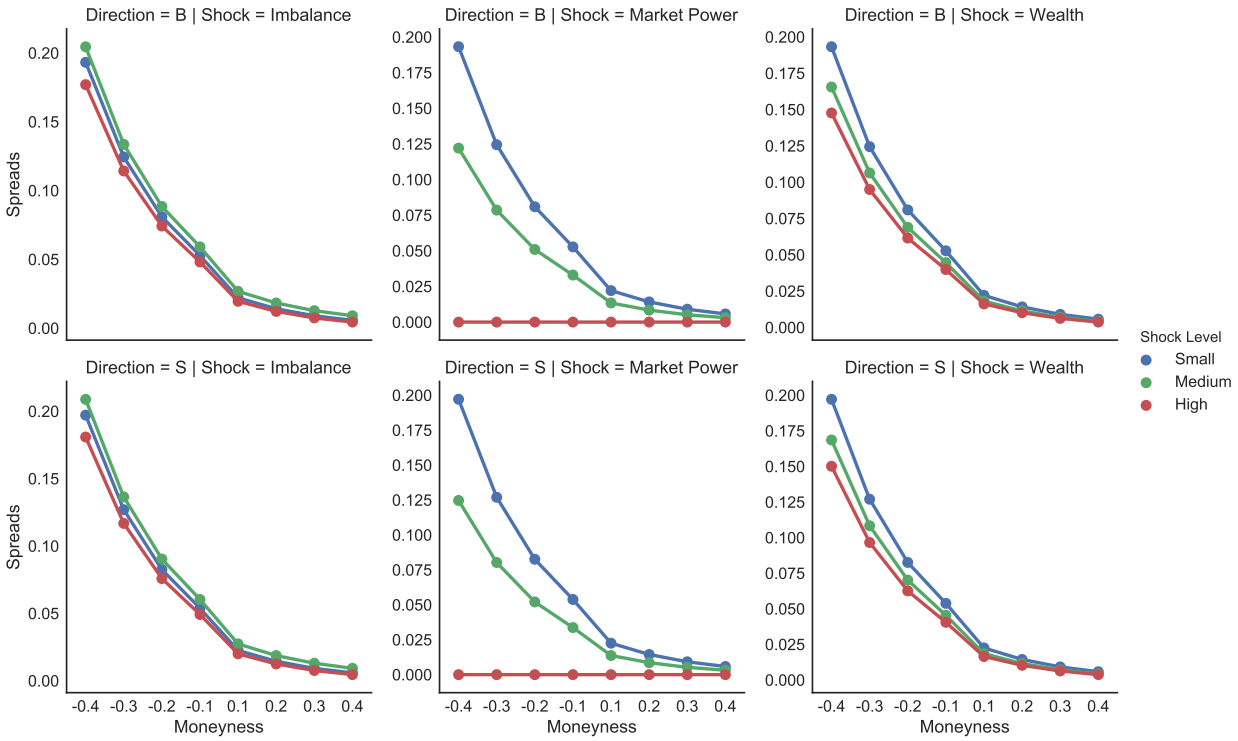

Figure 5: **Effects of 'macro' shocks on effective percentage bid-ask spreads for put options.**

Table 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	$F_{\mathrm{S}}^{\mathrm{C}}(0) = 0.6X + 0.8F_J$
Endowment of customer B	$F_{\mathrm{B}}^{\mathrm{C}}(0) = 0.8X - 2.0F_J$
Endowment of dealer j	$F_j^{\mathrm{D}}(0) = 0.3X + 0.6F_J$
Market Power	$\theta_i(0) = 1$ for $i = S, B$

Table 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]

Table 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 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	$\text{IMB}_t^{\text{LEVEL}}$	$\text{Variance}^{\mathbb{P}}_{t,t+30}$	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 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	$\text{IMB}_t^{\text{SKEW}}$	$\text{Skew}^{\mathbb{P}}_{t,t+30}$	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 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 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 8: **Correlation between the realized variance and realized skewness.** The quantile is based on the shape measure, $\text{IMB}_t^{\text{SHAPE}}$. 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 9: **Correlation between the variance and skewness risk premia.** The quantile is based on the shape measure, $\text{IMB}_t^{\text{SHAPE}}$. 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 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 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 $R^2 = 0.0088$	