CHAPTER 1

Over-the-Counter Markets

An over-the-counter (OTC) market does not use a centralized trading mechanism, such as an auction, specialist, or limit-order book, to aggregate bids and offers and to allocate trades. Instead, buyers and sellers negotiate terms privately, often in ignorance of the prices currently available from other potential counterparties and with limited knowledge of trades recently negotiated elsewhere in the market. OTC markets are thus said to be relatively opaque; investors are somewhat in the dark about the most attractive available terms and about whom to contact for attractive terms. Prices and allocations in OTC markets are to varying extents influenced by opaqueness and by the role of intermediating brokers and dealers. This chapter outlines some of the key institutional features of OTC markets that influence the formation of prices and allocations. Many details are omitted.

Some of the key research and policy issues regarding OTC markets include: (i) criteria that determine whether a financial product trades in an OTC market or on an exchange, (ii) the manner in which the price negotiated on a particular trade reflects the relative degrees of connectedness of the buyer and seller with the rest of the market, (iii) the formation and behavior of dealer franchises and oligopolies as well as interdealer brokers, (iv) the influence of market structure on the cross-sectional dispersion of prices negotiated at a particular time and on the time signature of price reactions to supply or information shocks, (v) the evolution over time of the distribution across investors of information learned from private trade negotiations, (vi) the effect of pre-trade and post-trade price transparency on market behavior, (vii) the impact of counterparty credit risk on pricing and financial stability, and (vii) the equilibrium strategies chosen by investors regarding their search for counterparties. Some of these issues form the main subject matter of this book. Others are left open for future research.

1.1 BILATERAL NEGOTIATION OF TERMS

Assets traded over the counter include most types of government and corporate bonds, asset-backed securities, securities lending and repurchase agreements, a wide range of derivatives, real estate, currencies, bulk commodities, shipping rights, and large blocks of equities. In most of these markets, trade is intermediated by dealers. Although the term "dealer" carries some legal distinctions in certain markets, the main difference between a dealer and other market participants is that, by convention, an OTC dealer is usually expected to make "two-way markets," in the following manner.

An OTC trade negotiation is typically initiated when an investor contacts a dealer and asks for terms of trade. Communication could be by phone, by e-mail, by screen-based query systems, or through a broker. A dealer making two-sided markets typically provides a take-it-or-leave-it pair of prices, a bid and an offer, to a customer. The customer may accept by immediately hitting the bid or lifting the offer. Dealer bids and offers are understood to be good for quantities up to an agreed conventional amount in standardized products. After agreeing on a price, a customer and dealer may further negotiate the quantity to be exchanged at the agreed price. Occasionally, if a customer declines to trade at the dealer's quotes, the dealer may offer a price concession rather than lose the opportunity to trade. In this case, the dealer is making a tradeoff between the value of maintaining a reputation for standing firm on its original quotes and the profit opportunity of the current trade.[1] A dealer trades for its own account and assumes the risk that it cannot offset a position at a profitable price.

An OTC bargaining game can be complex because of private information and the potentially rich sets of outside options of the respective counterparties. The counterparties may have different information regarding the common-value aspects of the asset (for example, the probability distribution of the asset's future cash flows), current market conditions, and their individual motives for trade. The counterparties may also be distinguished from each other with respect to their alternative trading opportunities, which depend on the manner and extent of their participation in the market. For

[1] A related reputation effect is explored by Atakan and Ekmekci (2010).

example, a dealer frequently negotiates trades and receives information, while the customer of a dealer usually has more limited opportunities to trade and thus relatively less information about recent transactions. Their different degrees of access to the market are in this case relatively common knowledge and tend to convey a negotiating advantage to the dealer. Beyond having more information than many of its customers, a dealer usually has less difficulty in adjusting inventories.[2] The negotiating advantage to a dealer may be increased when there are relatively few dealers with whom to negotiate. For example, as modeled by Zhu (2010), if an investor returns to a dealer from whom quotes have already been obtained, the dealer infers that the investor has had difficulty obtaining better terms from other dealers and is likely to offer even less attractive quotes to the investor.

Chapter 4 presents a simple model of the determination of prices in an OTC market with symmetric information. In this setting, prices reflect the fundamental values of the asset to the respective investors as well as the outside options of the respective counterparties to search for new trading partners. Chapter 5 focuses on the dynamics of asymmetric information across the population that is generated by OTC trading.

Some OTC markets have special intermediaries known as brokers that assist in negotiations between buyers and sellers, conveying the terms of one investor to another, usually without revealing the identities of the counterparties to each other. As opposed to dealers, a broker need not trade on its own account. Some brokers specialize in intermediating between dealers. A broker's customers rely on the broker for anonymity and for the broker's information regarding the most likely interested counterparties. Negotiation in OTC markets through brokers is often based on a "workup" protocol. Once the initial price and quantity are agreed on, one party can offer to expand the quantity of the trade. Further expansions can continue until at least one of the parties declines, although other parties in contact with the broker can then "enter" the trade at the same price, until no further trading interest is expressed at the original price, closing the trade. Huang, Cai, and Wang (2002), Boni and Leach (2004), and Pancs (2010) provide further description and analysis of the workup procedure and expandable-order markets.

[2] Inventory management by dealers and its implications for pricing are analyzed by Lyons (1995), O'Hara (1995), and Massa and Simonov (2003).

Over the past decade, the majority of interdealer and customer-to-dealer brokerage has moved to electronic platforms such as eSpeed, BrokerTec, Bloomberg, MarketAxess, MTS, and TradeWeb. Electronic trading platforms are likely to become more popular in OTC derivatives markets because of language in the U.S. Dodd-Frank Act of 2010 that mandates the trade of standardized OTC derivative products on a "swap execution facility" (SEF) that allows some degree of competition for trades among multiple market participants. As of this writing, the Securities and Exchange Comission (SEC) and the Commodity Futures Trading Commission (CFTC) have yet to define what will constitute an acceptable norm for SEFs. Once that is completed, perhaps in late 2011, the design of SEFs and the migration of trade in standardized OTC derivatives to SEFs will occur over time.

1.2 OTC TRANSPARENCY

In some dealer-based OTC markets, especially those with active brokers, a selection of recently quoted or negotiated prices is revealed to a wide range of market participants, often through electronic media such as Reuters, Bloomberg, or MarkitPartners. For other OTC markets, such as those for U.S. corporate and municipal bonds, regulators have mandated post-trade price transparency through the publication of an almost complete record of transactions shortly after they occur. In the United States, post-trade price reporting in many bond markets is done through the Trade Reporting and Compliance Engine (TRACE), which provides the price for each transaction within minutes of the trade. If the trade size is less than a stipulated threshold, TRACE reports the traded quantity. Otherwise, TRACE reports that the quantity was above the threshold. The size of extremely large trades is not publicly disclosed so that liquidity providers such as dealers have the chance to reduce inventory imbalances stemming from large trades with less concern that the size of a trade or their reservation price will be used to the bargaining advantage of their next counterparties.

In some active OTC derivatives markets, such as the market for credit default swaps, clients of dealers can request "dealer runs," which are essentially lists of dealers' prospective bid and offer prices on a menu of potential

trades. Dealers risk a loss of reputation if they frequently decline the opportunity to trade near these indicative prices when contacted soon after providing quotes for a dealer run.

Despite these sources of information, price transparency is typically lower for OTC markets than for exchange-based markets, such as those based on electronic communication networks (ECNs), in which there is essentially complete and immediate transactions reporting as well as pre-trade transparency that includes at least the best executable bid and offer. The complete contents of a typical ECN limit-order book are not normally disclosed. Some limit-order-book markets allow "icebergs," components of a limit order that have undisclosed quantities. Those submitting icebergs may wish to reduce inference by market participants that could worsen the order submitter's ultimate average execution price.

Increasingly, investors can allocate orders to off-exchange "dark pools," where they are crossed at prices that were recently executed on exchanges. Individual trades in dark pools are not normally disclosed. There has been some concern that the increasing fraction of equity trades sent to dark pools may lead to less competition for priced (on-exchange) trades and therefore less market liquidity and price discovery. Zhu (2011) shows theoretical support, however, for a tendency of dark pools to be relatively heavily populated by less informed traders. Under conditions, the presence of dark pools leads to a higher concentration of more highly informed traders at "lit" exchanges, which can actually improve price discovery.

The profits of a dealer depend on the volume of trade it handles and on the average difference between bid and ask prices, which in turn depends on the degree to which the dealer's customers are likely to have information relevant to prices available elsewhere in the market. Dealers therefore prefer at least some degree of market opaqueness. As pre-trade and post-trade price transparency increases, dealers have incentives to narrow their bid-offer spreads in order to compete for trading opportunities. If price transparency is too great, however, some dealers may lose the incentive to intermediate, given their fixed costs and the risk of adverse selection by informed customers. Unless the potential demand to trade a financial product is sufficient to justify exchange trading, a sufficiently large increase in OTC market transparency could therefore potentially reduce trading opportunities for investors.

Empirical anlyses of the implications of post-trade price transparency in bond markets, through TRACE, include those of Bessembinder and Maxwell (2008), Edwards, Harris and Piwowar (2007), Goldstein, Hotchkiss, and Sirri (2007), Green, Hollifield, and Schürhoff (2007a, 2007b), and Green, Li Schürhoff (2011). At this point, the empirical evidence does not generally support prior concerns expressed by dealers that the introduction of TRACE would reduce market liquidity.

The U.S. Dodd-Frank Act of 2010 regulates the transparency and stability of OTC derivatives markets in the United States. The European Commission plans to follow with its own new regulations. In addition to the impact on market transparency of mandated trade of standardized derivatives on swap execution facilities, the transparency requirements of the Dodd-Frank Act are of several major types: (i) the disclosure to regulators of all trade information through data repositories, (ii) the public disclosure of some aggregate information on trading volumes, and (iii) the public disclosure of transaction prices of standardized derivatives. The system for transaction price disclosure could be modeled on TRACE.

1.3 WHY TRADE OVER THE COUNTER?

Some of the lack of transparency of OTC markets is inherent in the nature of the underlying products. For example, a wide range of collateralized debt obligations and other structured credit products are thinly traded and have complex contractual features that could be analyzed well by only a narrow range of specialized investors. Even if such instruments were traded on an exchange, liquidity and transparency would be lacking. In any case, many such instruments could rarely achieve the volume and breadth of participation that would justify exchange-based trade. These are natural candidates for OTC trading, where customization of financial products to the needs of investors is more easily arranged. On request, an investment bank can offer a customer a new financial instrument for which there is no active market. In effect, an OTC market for that instrument is thereby created and may over time become established on the basis of additional trading. Eventually, if the product becomes sufficiently standardized and popular, its trade may migrate to an exchange.

Tufano (1989) describes the motivation of banks to innovate financial products. The introducer of a new financial product benefits in the short run from a temporary monopoly in the distribution of the product, through its command of the necessary technical and legal knowledge. Over time, a product may become sufficiently popular to encourage entry by other dealers. At this point, the original dealer's profit margins are likely to decline, although it may have a somewhat persistent advantage in terms of volume handled.

Some financial products with high volumes of trade that are relatively standard, for example, recently issued U.S. government bonds and certain liquid OTC derivatives (such as simple interest rate swaps and credit derivative index products), seem like natural candidates for exchange-based trade but are normally traded over the counter. At this point, we lack convincing theories that explain why such simple and heavily traded instruments are traded over the counter. Dealers have a vested interest in maintaining trade in OTC markets, where the profitability of intermediation is enhanced by market opaqueness. Once liquidity is established in OTC markets, it may be difficult for exchanges to compete for a critical mass of order flow, as investors are naturally drawn to the most liquid available market.

The prevalence of OTC trading for some standard financial products may be based in part on the granularity of trade sizes. Transactions of on-the-run government bonds, for instance, are often in quantities that are a substantial fraction of the total issuance size or daily volume of trade. A large quantity might be bought or sold at a more favorable price through private negotiation with a specialized OTC counterparty than by exposing the new demand or supply to a limit-order-book market populated by diversely motivated order submitters, where more severe price concessions may be required to digest a large position. For example, large block trades of equities are often executed in the private "upstairs" market of the New York Stock Exchange.[3] Her Majesty's Treasury reports an estimate by CREST, the United Kingdom's securities settlement system, that of the cash equity transactions placed into its system, OTC executions account for approximately 4% of the share volume and 14% of the market value of trades.[4]

[3] Keim and Madhavan (1996) analyze price formation in the upstairs search-based market for large blocks of equities.

[4] This does not not include equity transactions that are internalized over custodians' books or clearance services.

OTC derivatives and repurchase agreements, among certain other OTC bilateral contracts, expose each party of a trade to the risk that its counterparty will not perform as contracted, for example, because of bankruptcy. Even simple bank loans expose one of the two parties to the default risk of the other. If the default risk is not negligible, it is natural that the terms of these products are negotiated in the OTC market, so as to reflect the impact on each party of the default risk of the other. In OTC derivatives markets, the resulting price adjustment for counterparty default risk is called a credit value adjustment, or CVA.[5] It could be difficult to justify the cost of a distinct centralized market mechanism for derivatives that are materially exposed to the performance risk of an individual entity. A failure to price products according to the default risk of the specific counterparty invites adverse selection. For example, an offer to lend to a wide range of banks at a given interest rate leads to acceptance of the offer by borrowers with relatively high default risk. An attempt to correct for this adverse selection by raising the interest rate only worsens the credit quality of accepting borrowers. Anonymous exchange-based trading can therefore lead to inefficiently thin markets or even market failure.

1.4 MANAGING OTC CREDIT RISK

Techniques for mitigating OTC counterparty risk include the provision of collateral and, in some cases, central clearing. A financial contract is cleared when a special-purpose financial institution known as a central clearing party (CCP) becomes the buyer to the original seller, and the seller to the original buyer. By standing between the original parties to the trade, the CCP insulates each from the risk of default of the other, to the extent of the credit quality of the CCP itself.[6]

[5] Following the seminal paper of Sorensen and Bollier (1994), Duffie and Huang (1996) and Duffie, Schroder, and Skiadas (1996) modeled CVAs in settings in which the party at risk to default can change over time as the market value of the instrument to one of the two parties moves from positive to negative. This is typical of swap contracts.

[6] Duffie, Li, and Lubke (2009) review the central clearing of OTC derivatives among other policy issues regarding the infrastructure of OTC derivatives markets. Duffie and Zhu (2011) model the impact of clearing on counterparty credit risk and examine cases in which clearing is

Effectively designed and widely practiced central clearing improves the stability of the financial system by reducing the risk of chain-reaction failures of important financial institutions, and by lowering the incentives of investors to exit from positions with a financial institution whose solvency is suspected to be weak. These exits could contribute to a self-fulfilling "run" that reduces the liquidity of the weak financial institution to the point that it may indeed fail. CCPs can themselves be sources of inappropriately high systemic risk in the absence of sound CCP risk management, capital, collateral, regulatory supervision, and backstop sources of liquidity from lenders of last resort such as central banks.

The Dodd-Frank Act requires that standard derivatives traded by major market participants be cleared.[7] This legislation also mandates minimum collateral requirements, among other provisions.

1.5 PRICE BEHAVIOR WITH SEARCH FRICTIONS

There is extensive empirical evidence that supply or demand shocks in asset markets, in addition to causing an immediate price response, lead to adjustments over time in the distribution of capital across markets and adjustments over time in relative asset returns. In OTC markets, search delays could be responsible for some of these slow-recovery dynamics. Other causes of slow-recovery price dynamics are discussed by Duffie (2010). For modeling purposes, search could also proxy for other delays, such as those needed to establish trading relationships.

With delayed portfolio adjustments, there can be periods of time over which assets with essentially identical fundamental risks have different mean returns. More generally, there can be differences in mean returns across assets that are due not only to cross-sectional differences in "fundamental"

relatively ineffectual due to a potential loss of opportunities to net counterparty risk, which can happen with "too many" CCPs.

[7] The U.S. Treasury has proposed an exemption for foreign exchange derivatives that are settled by delivery of the two currencies. Duffie (2011) examines the foundation for this decision and provides a mechanism for obtaining the effect of clearing while leaving in place current processes for currency deliveries.

cash-flow risks but also to the degree to which the distribution of asset holdings across investors is inefficient (relative to a market without intermediation frictions). Empirical "factor" models of asset returns do not often account for factors related to the distribution of ownership of assets or related to likely changes in the distribution of ownership.

For example, in OTC corporate bond markets, one observes large price drops and delayed recovery in connection with major downgrades or defaults, as described by Hradsky and Long (1989) and Chen, Lookman, Schürhoff, and Seppi (2009), when certain classes of investors have an incentive or a contractual requirement to sell their holdings.

Mitchell, Pedersen, and Pulvino (2007) document the effect on convertible bond hedge funds of large capital redemptions in 2005. Convertible bond prices dropped and rebounded over several months. A similar drop-and-rebound pattern was observed in connection was the LTCM collapse in 1998. Newman and Rierson (2003) show that large issuances of credit-risky bonds temporarily raise credit spreads throughout the issuer's sector because providers of liquidity such as underwriters, dealers, and hedge funds bear extra risk as they search for long-term investors. They provide empirical evidence of temporary bulges in credit spreads across the European Telecom debt market during 1999–2002 in response to large issues by individual firms in this sector. Keim and Madhavan (1996) document analogous effects in the upstairs search-based market for large blocks of equities.

The market for catastrophe risk insurance is characterized by price surges, and then multiyear price declines, following sudden and large losses of capital by insurers, for example, at times of major natural disasters, as explained by Froot and O'Connell (1999). Periods of high insurance rates are typically accompanied by new entrants into the market, including hedge funds and other new insurers, whose capital has been mobilized by the high risk premiums, but not immediately. It takes time to set up a viable new provider of catastrophe risk insurance.

The pricing of securities lending is relatively strongly connected to search frictions, as explained by D'Avolio (2002), Duffie, Gârleanu, and Pedersen (2002), Geczy, Musto, and Reed (2002), Porras Prado (2010), and Blocher, Reed, and Van Wesep (2010).

In all of these examples, the time pattern of returns or prices after supply or demand shocks reveals that the friction at work is not a transaction cost

for trade. If that were the nature of the friction, then all investors would immediately adjust their portfolios, or not, optimally, and the new market price and expected return would be immediately established and remain constant until the next change in fundamentals. In all of the above examples, however, after the immediate price response, whose magnitude reflects the size of the shock and the degree of short-term competition among investors who are immediately available to absorb the sudden supply or demand, there is a relatively lengthy period of time over which the price recovers somewhat (up after an adverse shock, down after a positive shock), reverting toward its new fundamental level. In the meantime, additional shocks can occur with overlapping consequences. The typical pattern suggests that the initial price response is larger than would occur with perfect capital mobility and reflects the demand curve of the limited pool of investors immediately available to absorb the shock. The speed of adjustment after the initial price response is a reflection, at least in part, of the time that it takes to contact and negotiate with additional investors.

A significant body of theory treats the implications of search frictions for asset pricing. Early search-based models of intermediation include those of Rubinstein and Wolinsky (1987), Bhattacharya and Hagerty (1987), Moresi (1991), Gehrig (1993), and Yavaş (1996). Differences in search frictions across different asset markets are treated by Vayanos and Wang (2007), Weill (2008), and Vayanos and Weill (2008). Duffie, Gârleanu, and Pedersen (2005) study search frictions in a single asset market with market making. Their work is extended by Lagos and Rocheteau (2009) to model heterogeneity in asset positions. Duffie, Gârleanu, and Pedersen (2007) and Lagos, Rocheteau, and Weill (2009) model the dynamics of prices in an asset market that are caused by a shock to the preferences of investors. Search frictions imply a gradual reallocation of the asset to the most suitable investors, as captured by a search-based asset-pricing model in chapter 4.

Duffie and Strulovici (2007) model the flow of capital from one asset market to another, seeking higher risk premia for the same risk, with intermediation at a search intensity that is endogenously chosen based on the rents that can be negotiated for intermediation, extending work by Weill (2007). In some parameterizations, search efforts go from minimal to maximal whenever the total amount of capital in one market is sufficiently low relative to that in another. Alternative approaches to modeling the implications for price

dynamics of institutional frictions in capital movements are taken by Krishnamurthy and He (2010) and Gromb and Vayanos (2007).

In a relatively opaque OTC market, different investors may pay quite different prices for the same asset at essentially the same time. The investors may vary in terms of their relative bargaining power, their access to alternative trading opportunities, the quality of their information both about the fundamentals of the asset and about recent transactions.

For example, Green, Hollifield, and Schürhoff (2007a) document dramatic variation across investors in the prices paid for the same municipal bond. Massa and Simonov (2003) report dispersion in the prices at which different dealers trade the same Italian government bonds. In the next chapter, based on Ashcraft and Duffie (2007), we will see that the rate at which a pair of banks negotiate a loan of federal funds in the overnight interbank market at a given time of day, relative to the average rate negotiated by other pairs of banks at the same time of day, varies according to the relative cash holdings of the two banks at that moment of the day, the degree to which the two banks are active in the interbank lending market, and the extent to which the banks have had a prior borrower-lender relationship, among other factors.