n° 2015-11
Agents' Behavior on Multi-Dealer-to-Client Bond Trading Platforms
J-D.Fermanian[1]
O.Guéant[2]
A.Rachez[3]

[1] ENSAE-CREST. Email: jean-david.fermanian@ensae.fr

[2] ENSAE-CREST. Email: olivier.gueant@ensae.fr

[3] Institut Europlace de Finance. Email: arnaud.rachez@gmail.com

Agents' Behavior on Multi-Dealer-to-Client Bond Trading Platforms*

Jean-David Fermanian,[†] Olivier Guéant,[‡] Arnaud Rachez[§]

Abstract

For the last two decades, most financial markets have undergone an evolution toward electronification. The market for corporate bonds is one of the last major financial markets to follow this unavoidable path. Traditionally quote-driven (that is, dealer-driven) rather than order-driven, the market for corporate bonds is still mainly dominated by voice trading, but a lot of electronic platforms have emerged that make it possible for buy-side agents to simultaneously request several dealers for quotes, or even directly trade with other buy-siders. The research presented in this article is based on a large proprietary database of requests for quotes (RFQ) sent, through the multi-dealer-to-client (MD2C) platforms operated by Bloomberg Fixed Income Trading and Tradeweb, to one of the major liquidity providers in European corporate bonds. Our goal is (i) to model the RFQ process on these platforms and the resulting competition between dealers, (ii) to use the RFQ database in order to implicit from our model the behavior of both dealers and clients on MD2C platforms, and (iii) to study the influence of several bond characteristics on the behavior of market participants.

1 Introduction

For many years, the trading of corporate bonds[1] on the secondary market only took place over the counter via private negotiations on the phone. The organization of the market corresponded to a classical quote-driven one, where market participants are divided into two groups: clients (the buy-side), *i.e.* institutional investors, wealth management companies,

†Ensae-Crest, 3 avenue Pierre Larousse, 92245 Malakoff Cedex, France. Email: `jean-david.fermanian@ensae.fr`

‡Ensae-Crest, 3 avenue Pierre Larousse, 92245 Malakoff Cedex, France. Email: `olivier.gueant@ensae.fr`

§Institut Europlace de Finance, 28 Place de la Bourse, 75002 Paris, France. Email: `arnaud.rachez@gmail.com`

[1]In this article, the term "corporate bonds" encompasses most non-government bonds, issued by financial institutions or other corporates.

and hedge funds in general, who call dealers (the sell-side) – mainly investment banks – to buy or sell bonds. In this economic environment, dealers act as market makers or liquidity providers, whereas clients are liquidity-taker agents. The market was also divided into two segments: the dealer-to-client (D2C) segment, where transactions occur between dealers and clients, and the interdealer-broker (IDB) segment.

This traditional description of the corporate bond market is still valid in many ways, but the way corporate bonds are traded is evolving constantly. Recent evolutions are due to at least two factors: technological innovation and financial regulation.

The electronification of financial markets has dramatically changed the way most securities are traded, and subsequently the organization of most financial markets. Electronic order books are now the norm for a lot of asset classes: stocks obviously, but also U.S. Treasuries and foreign exchanges. Although many historically dealer-driven markets eventually adopted the order-driven paradigm,[2] it is difficult to imagine a similar market organization for corporate bonds today. One major difference between bonds and stocks, to remain in the field of cash markets, has to do with heterogeneity and liquidity. There is usually one stock for a given company while there are often dozens of bonds for the same company, corresponding to different maturities, different coupons, and different seniorities. A natural consequence is that most corporate bonds are illiquid. The Securities Industry and Financial Markets Association (SIFMA) estimates indeed (see [16]) that the total market value of stocks is twice the market value of corporate bonds, but that there are more than six times more listed corporate bonds than listed stocks, and furthermore, the global average daily volume is assumed to be \$17.9 billion for corporate bonds vs. \$112.9 billion for stocks. Even worse, MarketAxess Research (see [11]) estimates that, in 2012, 38% of the 37,000 TRACE-eligible bonds did not trade, even once, and only 1% of these 37,000 bonds traded every day. The lack of standardization and the illiquidity of most corporate bonds, along with the buy-and-hold strategy of many investors, make it difficult the emergence of a match-based model with order books similar to those of equity markets, except perhaps for the most liquid corporate bonds. However, electronification occurs, and new models emerge that are different from those of equity markets.

Technological innovation occurs in the D2C market segment which is "electronifying" increasingly, whereas trades in the IDB market segment remain almost entirely executed via voice.[3] Several types of platforms have emerged. The most important and successful ones are multi-dealer-to-client (MD2C) trading platforms. Examples of such MD2C platforms are those proposed by Bloomberg Fixed Income Trading (FIT), Tradeweb and MarketAxess. They clearly dominate the landscape of electronic trading. With these platforms, there is still a distinction between dealers and clients, but clients can send simultaneously a request for quotes to several dealers who have streamed prices to the platforms. Another kind of

---

[2]In the case of the UK, the transition occurred in 1997, with the creation of the SETS.

[3]Although it is still high, the volume of the IDB market segment has been constantly decreasing for years. The IDB segment is not the purpose of this paper.

MD2C electronic platform that is used by many buy-siders (especially small ones) consists of executable quotes, but only for odd lots. Single-dealer electronic platforms have also emerged to replace the telephone. Crossing systems, and platforms proposing all-to-all central limit order books (CLOB) also exist. They try to revolutionized the classical distinction between dealers who provide liquidity and clients who take liquidity, but they are still very rarely used in practice.

An interesting study carried out in [11] shows that the current market structure is far from being stabilized, given the different viewpoints of market participants on the future organization of the market. If it is believed that the market will remain for years a dealer-driven one, with MD2C RFQ platforms holding the lion's share of electronic trading, the evolution of financial regulation will have a lot of influence on the market structure and the role of the different participants. It is already clear that Basel III capital requirement deters investment banks from holding large inventories. Therefore, their traditional role as dealers may change from being market makers to simply providing access to the market. This might encourage all-to-all platforms and give a new roles to large asset managers as specialists on these platforms. In addition to Basel III, the question of pre-trade transparency for corporate bonds has been raised in the debate on MiFID 2: the European landscape may change to apply new rules.

In this paper, we focus on the current state of the market, and on the dominant MD2C platforms, where buy-siders send quotes to a swath of dealers on a specific bond. More precisely, the process works as follows in the case of a client who wants to buy/sell a given bond:[4]

1. The client connects to a platform and sees the bid and offer prices streamed by the dealers for the security. These prices, streamed by dealers, correspond to prices for a given reference size. These streamed prices are not firm/binding prices.

2. The client selects dealers (up to 6 dealers on Bloomberg FIT for instance), and sends one RFQ through the platform to these dealers, with a precise volume (notional), and the side ("buy" order or "sell" order).

3. Requested dealers can answer a price to the client for the transaction (not necessarily the price he/she has streamed). Dealers know the identity of the client (contrary to what happens in the case of most of the systems organized around a CLOB) and the number of requested dealers (the degree of competition for this request). However, they do not see the prices that are streamed by other dealers. They only see a composite price at the bid and offer, based on some of the best streamed prices. These composite prices in the case of Bloomberg FIT are called the CBBT bid price and the CBBT offer price.

4. The client progressively receives the answers to the RFQ. He/She can deal at any time with the dealer who has proposed the best price, or decide not to trade.

---

[4]Sometimes, clients send RFQs with no intention to buy or sell bonds, but only to get information.

5. Each dealer knows whether a deal was done (with him/her, but also with another dealer – without knowing the identity of this dealer) or not. If a transaction occurred, the best dealer usually knows the cover price, if there is one.

The electronification of the request process between supply-siders and liquidity providers generates a lot of data. For each RFQ they receive, dealers can record information. Therefore, dealers now have a complete history of their interaction with their clients, in a very standardized fashion, and they even get notified of trades occurring with competitors when they participate to a RFQ process.

The work presented in this article is based on a database of RFQs received through Bloomberg FIT and Tradeweb. It has been supplied by one of the most important dealers in European corporate bonds (BNP Paribas), that will be called the "reference dealer" in this paper, even though it does not play a particular role in this market. The proprietary database we got access to represents a fraction of the RFQs received by this dealer over one year (part of 2013 and 2014). For each RFQ, we observe its characteristics (date, hour, id of the client, Isin of the bond, buy/sell side, notional, number of dealers requested, etc.), contextual information (prices streamed and answered by the reference dealer, CBBT bid and offer prices, etc.), and the outcomes of the RFQ (whether there was a deal or not, the cover price – if there is one – in the case of a trade with the reference dealer, etc.).

We build a parsimonious model for the RFQ process, in which the (unobserved) prices answered by other dealers follow an unknown distribution, and where clients decide to trade or not to trade depending on a (unobserved) reservation value following another unknown distribution. We apply a Markov chain Monte Carlo (MCMC) method on the database of RFQs to estimate these distributions for the quotes answered by the dealers, and for the reservation value of the clients. These distributions are parameterized to take account of the information available. Therefore, we address questions such as the dependence of the behavior of dealers on the number of dealers requested, the type of bonds, etc. Similarly, we analyze how buy-siders behave depending on the context.

Applications of our work are numerous. Assessing the behavior of competitors is indeed of the utmost importance for a dealer. By modelling the behavior of his competitors in a better way, a dealer can expect to better analyze/manage hit ratios, and better estimate the probability to trade at a given price. Competition can also be analyzed: the behavior of dealers answering a RFQ indeed depends on the degree of competition, *i.e.* the number of dealers requested. A good model to estimate the reservation value of clients with respect to a bond is also key for a dealer: it is one of the most important inputs for choosing the quote he/she answers to the client in the RFQ process. Our model leads to very general input functions for market making models. In a nutshell, the goal of market makers / dealers, is to quote bid and offer prices so as to make money out of the spread between these two prices, while mitigating the risk of price changes on the value of the inventory. Old market making models include those of Ho and Stoll [8, 9]. More recently, Avellaneda and Stoikov [1] (see also [6] for closed-form expressions for the optimal quotes) proposed a model that

could be applied to quote-driven markets such as the corporate bond market.[5] However, the exponential form of the execution intensities used in [1] and [6] is at odds with reality. The more general model presented in the appendix of Guéant and Lehalle [5] or in [7] goes beyond the case of exponential intensities, and it may be adapted to incorporate the findings of this paper in a dynamic market making model.

In Section 2, we describe our model for the RFQ process. In Section 3, we present the features of our dataset and some adapted estimation methods. Empirical results are shown in Section 4. In particular, we discuss the influence of several covariates to explain the behavior of both dealers and clients. In Section 5, we discuss possible extensions of our model.

2 The model for dealers' quotes and clients' behavior

We have described above the RFQ process on platforms such as Bloomberg FIT. Now, let us specify the model we propose for the behavior of dealers and clients, along with the associated notations.

To simplify, we consider in what follows the case of a RFQ i corresponding to a "buy" order[6] (the mechanism is similar for a "sell" order):

1. A (prospective) client has identified a specific corporate bond that may be of interest for him/her. He/she sees the prices streamed by dealers for this bond, and believes that it is worth sending a (buy) RFQ.[7] We denote by V_i the client's view on the value of the bond.

2. This client sends a RFQ to $n_i + 1$ dealers – $n_i \in \{0, \ldots, 5\}$ – to buy a number of bonds corresponding to a specific amount of cash.[8] Each dealer $k = 0, \ldots, n_i$ sees the CBBT bid and offer prices which are composite prices based on the prices streamed by dealers. Dealer k also has his/her own evaluation of the bond price, and we assume that he/she answers a price $W_{k,i}$ to the client,[9] without collusion between dealers. These prices are binding, in the sense that $W_{k,i}$ will be the transaction price if the dealer k is chosen by the client for this deal. Therefore, the $n_i + 1$ dealers compete with each other to be chosen by the client.

---

[5]The model was initially built for market making on stock markets, but it is better suited for market making on quote-driven markets.

[6]A "buy" (resp. "sell") order means that the client sends a request to buy (resp. sell) the bond. The dealer then answers a quote to sell (resp. buy) the bond.

[7]The client may believe that the value of this bond is likely to increase in the future, and he/she expects to sell it back. Alternatively, he may think that a buy-and-hold strategy is profitable, given the current evaluation of the market for that corporate bond.

[8]This amount is called the notional of the RFQ.

[9]This is a strong assumption because some dealers do not answer or do not have the time to answer. This assumption will be revisited in Section 5.

3. We assume that a deal occurs if and only if a dealer proposes a bond price that is lower than V_i.[10] In that case, the transaction occurs between the client and a dealer chosen by the client amongst the dealers who have proposed the lowest price. In particular, the deal price is $\min_{k=0,\ldots,n_i} W_{k,i}$.

Obviously, the whole previous process depends on the bond characteristics and the public market information at the time of the RFQ. We assume that this information is summarized into a σ-algebra Ω_i.

In full generality, the quoting processes of dealers for "buy" and "sell" orders should not be identical, because inventory management and business-related incentives differ. This should be the case on the client side too, although this is less intuitive. We model both the reservation value V_i of clients and the price answered by dealers $W_{k,i}$ by random variables. Throughout this paper, F (or in fact $F(\cdot|\Omega_i)$) refers to the cumulative distribution function of the variable $W_{k,i}$, and G (or in fact $G(\cdot|\Omega_i)$) refers to the cumulative distribution function of the variable V_i, in the case of a "buy" RFQ.[11] In the case of a "sell" RFQ, the same notations are used, with a star: F^*, G^*, etc. Note that the distributions of dealers' quotes are assumed to be the same across dealers. Moreover, for the sake of simplicity, we assume this is the case for clients too, even though it is possible – and easy – to associate different functions G and G^* to different clients.

To complete the model specifications, we need to state our assumptions concerning the functional form of the functions F and G, and similarly for F^* and G^*.

Because the CBBT bid and offer prices constitute reference prices, and because the streamed bid-to-mid Δ_i (*i.e.*, half the streamed bid-ask spread[12]) constitutes a proxy of liquidity – which is crucially linked to the level of risk aversion associated with bond quoting –, it is convenient to work with "reduced quotes": $(V_i - \mathrm{CBBT}_i)/\Delta_i$ for clients, and $(W_{k,i} - \mathrm{CBBT}_i)/\Delta_i$ for dealers, where CBBT_i is the CBBT offer price in the case of a "buy" RFQ, and the CBBT bid price in the case of a "sell" RFQ.

In other words, it makes sense to assume that

$$F(\xi|\Omega_i) = F_0\left(\frac{\xi - \mathrm{CBBT}_i}{\Delta_i}\right), \qquad F^*(\xi|\Omega_i) = F_0^*\left(\frac{\xi - \mathrm{CBBT}_i}{\Delta_i}\right), \qquad (2.1)$$

$$G(\xi|\Omega_i) = G_0\left(\frac{\xi - \mathrm{CBBT}_i}{\Delta_i}\right), \qquad G^*(\xi|\Omega_i) = G_0^*\left(\frac{\xi - \mathrm{CBBT}_i}{\Delta_i}\right), \qquad (2.2)$$

for some cumulative distribution functions F_0, F_0^*, G_0, and G_0^*.[13]

---

[10] For that reason, we refer to V_i as the reservation price or reservation value of the client.
[11] We denote by f and g the corresponding probability density functions.
[12] One could alternatively use the CBBT bid-ask spread.
[13] The CBBT bid and offer prices and the streamed bid-to-mid spread are both included into Ω_i.

Furthermore, we may want to include some bond characteristics, or evaluate the effect of some covariates (volume, investment grade/high-yield, maturity, issuer, inventory, etc.). These covariates can be stacked into a vector Z that contains two subgroups of variables, *i.e.* $Z := [Z_1, Z_2]'$: the variables indexed by 1 have a direct influence on the quotes/prices, whereas the variables indexed by 2 have to be compared to reduced quotes. This leads to:

$$F(\xi|\Omega_i) = F_0\left(\frac{\xi - \text{CBBT}_i - Z'_{i,1}b_D}{\Delta_i} - Z'_{i,2}c_D\right),$$

$$F^*(\xi|\Omega_i) = F_0^*\left(\frac{\xi - \text{CBBT}_i - Z'_{i,1}b_D^*}{\Delta_i} - Z'_{i,2}c_D^*\right), \tag{2.3}$$

$$G(\xi|\Omega_i) = G_0\left(\frac{\xi - \text{CBBT}_i - Z'_{i,1}b_C}{\Delta_i} - Z'_{i,2}c_C\right),$$

$$G^*(\xi|\Omega_i) = G_0^*\left(\frac{\xi - \text{CBBT}_i - Z'_{i,1}b_C^*}{\Delta_i} - Z'_{i,2}c_C^*\right), \tag{2.4}$$

for some vectors of parameters b_D, b_D^*, b_C, b_C^*, c_D, c_D^*, c_C, and c_C^* (to be estimated).

In (2.1) and (2.2), there is no dependency on the number of dealers. In what follows, we often consider instead the following specification:[14]

$$F(\xi|\Omega_i) = F_0\left(\frac{\xi - \text{CBBT}_i}{\Delta_i}; n_i\right), \qquad F^*(\xi|\Omega_i) = F_0^*\left(\frac{\xi - \text{CBBT}_i}{\Delta_i}; n_i\right), \tag{2.5}$$

$$G(\xi|\Omega_i) = G_0\left(\frac{\xi - \text{CBBT}_i}{\Delta_i}; n_i\right), \qquad G^*(\xi|\Omega_i) = G_0^*\left(\frac{\xi - \text{CBBT}_i}{\Delta_i}; n_i\right), \tag{2.6}$$

for some cumulative distribution functions $F_0(\cdot; n)$, $F_0^*(\cdot; n)$, $G_0(\cdot; n)$, and $G_0^*(\cdot; n)$, where $n \in \{0, \ldots, 5\}$, correspond to the numbers of dealers requested in addition to the reference dealer.

In this paper, we rely on a parametric specification of the above distributions. To estimate the model, we have used a large sample of RFQs obtained from one of the dealers (the "reference dealer"). All the prices answered by this dealer are denoted by the letter Y. In other words, the price answered by the reference dealer for the RFQ i is Y_i.

We have observed that the empirical distribution of $((Y_i - \text{CBBT}_i)/\Delta_i)_i$ is leptokurtic (fat tailed), very spiky around the composite price and asymmetric. We guess most dealers should behave similarly. Therefore, we would like to exhibit a flexible parametric family with such features. For that purpose, we promote the use of the skew exponential power distribution (SEP), for which the skewness and the kurtosis have been shown to belong to rather wide intervals.

---

[14]We can also add the same covariates as above, obviously.

As a preliminary, let us recall the exponential power distribution (see [17]) whose probability density function is

$$f_{EP}(x; \mu, \sigma, \alpha) = \frac{1}{c\sigma} \exp\left(-|z|^\alpha/\alpha\right), x \in \mathbb{R},$$

where $\alpha > 1$, $\mu \in \mathbb{R}$, $\sigma > 0$, $z := (x - \mu)/\sigma$ and $c := 2\alpha^{1/\alpha-1}\Gamma(1/\alpha)$.[15]

The density of the SEP distribution is deduced from the general methodology to accommodate asymmetry proposed by Azzalini in [2]: one introduces an additional parameter $\lambda \in \mathbb{R}$ that reflects asymmetry, and the density of the SEP distribution is given by

$$f_{SEP}(x) = 2\Phi(w) f_{EP}(x; \mu, \sigma, \alpha), \quad x \in \mathbb{R}, \tag{2.7}$$

where $w := \text{sign}(z)|z|^{\alpha/2}\lambda(2/\alpha)^{1/2}$, and Φ is the cumulative distribution function of the standard normal distribution. Such a distribution is denoted by $SEP(\mu, \sigma, \alpha, \lambda)$.

We stack all the SEP parameters into a vector $\theta := (\mu, \sigma, \alpha, \lambda)$. The SEP reduces to the exponential power when $\lambda = 0$, to the skew normal when $\alpha = 2$, and to the normal when $(\lambda, \alpha) = (0, 2)$. In what follows, we always consider $\alpha \leq 2$, because we want our distributions to be fat-tailed. We refer the interested reader to Azzalini [3] and DiCiccio and Monti [4] for detailed results concerning this family of distributions.

It must be noted that μ and σ are not the mean and standard deviation of a $SEP(\mu, \sigma, \alpha, \lambda)$ distribution. These parameters are called "location" and "scale" instead. The even moments of a random variable $Z \sim SEP(0, 1, \alpha, \lambda)$ are given by

$$E[Z^{2m}] = \alpha^{2m/\alpha}\Gamma((2m + 1)/\alpha)/\Gamma(1/\alpha), m \in \mathbb{N}, \tag{2.8}$$

and the odd moments by

$$E[Z^{2m+1}] = \frac{2\alpha^{(2m+1)/\alpha}\lambda}{\sqrt{\pi}\Gamma(1/\alpha)(1 + \lambda^2)^{s+1/2}} \sum_{n=0}^\infty \frac{\Gamma(s + n + 1/2)}{(2n + 1)!!} \left(\frac{2\lambda^2}{1 + \lambda^2}\right)^n, \tag{2.9}$$

where $s = 2(m + 1)/\alpha$ and $(2n + 1)!! := 1 \cdot 3 \cdots (2n - 1) \cdot (2n + 1)$ – see [4]. In particular, when $\lambda \geq 0$, $E[Z] \geq 0$.

Hereafter, we always assume that the cumulative distribution functions F_0, F_0^*, $F_0(\cdot; n)$ and $F_0^*(\cdot; n)$ are all of the skew exponential power type, $n = 0, \ldots, 5$.

As far as clients are concerned, we do not have any empirical intuition for the form of the distribution of reservation values. Therefore, by default, we propose to use a rather naive distribution. We assume that V_i is Gaussian conditionally to Ω_i: in the case of a "buy" (resp. "sell") order, $G_0(\cdot|\Omega_i)$ (resp. $G_0^*(\cdot|\Omega_i)$) is the cumulative distribution function of a Gaussian random variable $\mathcal{N}(\nu, \tau^2)$ (resp. $\mathcal{N}(\nu^*, (\tau^*)^2)$). When a client is interested in

---

[15]Despite the presence of an absolute value, this density is differentiable with respect to its parameters because of the constraint $\alpha > 1$. In other words, the maximum likelihood estimator satisfies the first order conditions.

buying (resp. selling) a particular bond, we expect that he/she thinks that the bond is underpriced (overpriced) by the market. In other words, we expect $\nu > 0$ and $\nu^* < 0$. That will be confirmed empirically (see Section 4).

3 Estimation of the model parameters

3.1 The dataset

If all the quotes V_i and $W_{k,i}$ and the final outcomes of (a subset of) RFQs were available, it would be easy to infer F and G. Unfortunately, this is not the case. Actually, every dealer faces a partial information problem, because the amount of information he/she can retrieve from the RFQs is strongly constrained and limited. For convenience, let us adopt the point of view of the "reference dealer".

To be specific, the reference dealer gets the following information with RFQ i:[16]

- The output I_i of the RFQ:[17]

  - $I_i = 1$ (Done), when the RFQ resulted in a trade with the reference dealer,
  - $I_i = 2$ (Traded Away), when the RFQ resulted in a trade, but with another dealer.[18]
  - $I_i = 3$ (Not Traded), when the RFQ resulted in no trade.

- The second best dealer price C_i, called the "cover price", when the reference dealer has made the deal, and when there was another answer.

- Y_i, the price/quote answered by the reference dealer for this RFQ. Note that this is the price of the deal when $I_i = 1$, *i.e.* when the reference dealer has been chosen by the client.

- n_i, the number of dealers requested during this RFQ, in addition to our reference dealer ($n_i \leq 5$ with this convention).[19]

---

[16] More information is available, in particular, the identity of the client is known by the dealer before he answers a price. We have not used this information in this paper.

[17] We only considered the RFQs for which the reference dealer has answered a quote.

[18] Additional information is actually available when a RFQ has been "traded away" ($I_i = 2$). Indeed, another discrete dummy variable $J_i \in \{1, 2, 3, 4\}$ is then available:

* $J_i = 1$ means "Tied" (traded away), *i.e.* the reference dealer has proposed exactly the same price as the winner, but has not been chosen.
* $J_i = 2$ means "Covered" (traded away), *i.e.* the reference dealer has proposed the second best price, and was the only dealer to propose this price.
* $J_i = 3$ means "Tied covered" (traded away), *i.e.* the reference dealer has proposed the second best price, but another dealer has been in exactly the same situation.
* $J_i = 4$ is the "Other" traded away case, that corresponds to the other configurations.

[19] We only considered the cases $n_i \geq 1$, because $n_i = 0$ is specific.

- Z_i, a vector of bond characteristics.

Our dataset contains a N-sample of such information. After filtering, the number N of "buy" (resp. "sell") RFQs is equal to 192,855 (resp. 228,903).

The available information Ω_i for all market participants, includes the variables n_i, Z_i and the standard market information (news, Bloomberg quotes, etc.), in addition to CBBT prices. The information of a given RFQ is assumed to be independent of other RFQs. In particular, we assume no "learning" effect, *i.e.* the clients do not learn from past RFQs. The dealers propose prices $(W_{k,i})_k$ drawn independently from the same conditional distribution $F(\cdot|\Omega_i)$ – or $F^*(\cdot|\Omega_i)$ –, that depend on the bond characteristics and on the number of dealers maybe. These quotes and V_i are chosen independently, conditionally on the whole market information.

Let us concatenate the information into an *i.i.d* sample $\mathcal{S}_N = (Y_i, n_i, C_i, I_i, J_i, Z_i)_{i=1}^N :=$ $(\mathbf{X}_i)_{i=1,\dots,N}$. Our goal is to estimate the conditional distributions $F(\cdot|\Omega_i)$, $G(\cdot|\Omega_i)$, $F^*(\cdot|\Omega_i)$ and $G^*(\cdot|\Omega_i)$.

3.2 Maximum likelihood inference

The usual maximum likelihood methodology can be invoked for evaluating all the unknown model parameters. However, the difficulties are twofold:

1. Computing the full log-likelihood $\mathcal{L}_N$ of the N-sample is cumbersome. In practice, it involves the numerical approximation of numerous univariate and bivariate integrals.

2. The maximization of $\mathcal{L}_N$ must be carried out over a large number of model parameters, even though the optimization has to be carried out for the "buy" case and the "sell" case separately. In the former (resp. latter) case, the parameters are stacked into a vector ζ (resp. ζ^*). When some usual conditions of regularity are satisfied, the likelihood function could be estimated by some classical algorithms (BFGS, simulated annealing,...), at least in theory.

Details of the full log-likelihood are given in Appendix A. When it is feasible in practice, the estimator $\hat{\zeta}$ is consistent and asymptotically normal: when N tends to the infinity,[20]

$$\sqrt{N}\left(\hat{\zeta} - \zeta\right) \xrightarrow{law} \mathcal{N}\left(0, J^{-1}IJ^{-1}\right), \text{ where} \tag{3.1}$$

$$I = E[\partial_\zeta \mathcal{L}_1(\theta)\partial_{\zeta'}\mathcal{L}_1(\theta)] \# \frac{1}{N}\sum_{i=1}^N \partial_\zeta \ln \mathcal{L}_i(\hat{\zeta})\partial_{\zeta'}\ln \mathcal{L}_i(\hat{\zeta}),$$

$$J = E[\partial_{\zeta,\zeta}^2 \mathcal{L}_1(\zeta)] \# \frac{1}{N}\sum_{i=1}^N \partial_{\zeta,\zeta}^2 \ln \mathcal{L}_i(\hat{\zeta}).$$

---

[20]Similar results apply to $\hat{\zeta}^*$.

3.3 Estimation by Markov chain Monte Carlo

Inference with partially observed data can be managed nicely with a Bayesian approach and Markov chain Monte Carlo techniques. See Appendix B for a quick reminder of the methodology.

In the case of a "buy" RFQ, the underlying vector of model parameters ζ can be decomposed as $\zeta := (\tilde{\alpha}, \tilde{\beta})$, where $\tilde{\alpha}$ (resp. $\tilde{\beta}$) is the vector of parameters defining the distribution of client's reservation values G (resp. dealers' quotes F). In particular, the previous SEP parameter θ is a sub-vector of $\tilde{\beta}$, and (ν, τ) are parts of $\tilde{\alpha}$.

Under a Bayesian point of view, all quotes are seen as unknown parameters, a situation that is in line with our data constraints. In the situation at hand, we have to estimate the following parameters: $\tilde{\alpha}, \tilde{\beta}$, as well as the latent variables $\mathbf{V}$ and $\mathbf{W}$ that concatenate all the non-observable quotes of the RFQs in the sample. We wish to simulate from the joint posterior $\pi(\tilde{\alpha}, \tilde{\beta}, \mathbf{V}, \mathbf{W} | \mathbf{X}, \mathbf{\Omega})$, where the main information coming from the data is included in $\mathbf{X}$, particularly the quotes $\mathbf{Y}$ of the reference dealer and the results $\mathbf{I}$ of the RFQ process. Estimation can be done through Gibbs sampling, *i.e.* sequential draws from the model conditional distributions. Some of these conditional distributions can be simulated easily; the others can be simulated by using accept-reject sampling.

For sampling from the posterior distribution, we use Algorithm 1:

Algorithm 1 Specific Gibbs sampler

Initialize (*e.g.* at random) $(\tilde{\alpha}^0, \tilde{\beta}^0, \mathbf{V}^0, \mathbf{W}^0)$
for iteration t=1 to T **do**
    Simulate $\tilde{\alpha}^t$ from the conditional distribution $\pi(\tilde{\alpha} | \tilde{\beta}^{t-1}, \mathbf{V}^{t-1}, \mathbf{W}^{t-1}, \mathbf{X}, \mathbf{\Omega})$
    Simulate $\tilde{\beta}^t$ from the conditional distribution $\pi(\tilde{\beta} | \tilde{\alpha}^t, \mathbf{V}^{t-1}, \mathbf{W}^{t-1}, \mathbf{X}, \mathbf{\Omega})$
    Simulate $\mathbf{V}^t$ from the conditional distribution $\pi(\mathbf{V} | \tilde{\alpha}^t, \tilde{\beta}^t, \mathbf{W}^{t-1}, \mathbf{X}, \mathbf{\Omega})$
    Simulate $\mathbf{W}^t$ from the conditional distribution $\pi(\mathbf{W} | \tilde{\alpha}^t, \tilde{\beta}^t, \mathbf{V}^t, \mathbf{X}, \mathbf{\Omega})$
end for

It is well-known that the Markov chain $(\tilde{\alpha}^t, \tilde{\beta}^t, \mathbf{V}^t, \mathbf{W}^t)_{t=1,\dots,T}$ converges in law towards its stationary distribution given the data. We deduce the law of $(\tilde{\alpha}, \tilde{\beta})$ given $(\mathbf{X}, \mathbf{\Omega})$ (the "posterior"). Note that when simulating a (sub)vector, one can either simulate all the values of the (sub)vector jointly when it is possible, or simulate them sequentially, as in the other steps of the algorithm.

A particular and non-standard stage is the random draw of a random variable $X \sim SEP(\mu, \sigma, \alpha, \lambda)$. In DiCiccio and Monti [4], Section 4, a simulation procedure is proposed:

1. Draw $Y := R_1(-\alpha B \ln U)^{1/\alpha}$, where U is uniformly distributed on $(0,1)$, B is Beta distributed with parameters $(1/\alpha, 1-1/\alpha)$ and R_1 is uniformly distributed on $\{-1, 1\}$. All the three latter random variables are mutually independent.

2. Let $Z := R_2 Y$, where $R_2 = 1$ with probability $\Phi(W)$, and $R_2 = -1$ with probability $1 - \Phi(W)$, $W := \text{sign}(Y)|Y|^{\alpha/2}\lambda(2/\alpha)^{1/2}$. Then Z follows a $SEP(0, 1, \alpha, \lambda)$.

3. Set $X = \mu + \sigma Z$.

Choice of priors

Since we do not have strong a priori information about the model parameters, two principles were used for the choice of their prior distributions:

1. Whenever possible, conjugate priors were preferred for computational ease. For the Gaussian model, this corresponds to a Gaussian prior for the mean parameter, and an Inverse Gamma distribution for the standard deviation one. Several different reasonable parameter values have been tried in the prior, without any strong impact on the posterior.

2. Otherwise, we used flat improper priors (see Appendix B). But improper priors are not an issue here, since the posterior distribution is still proper.

Simulating from the conditional distributions

Several conditional distributions are truncated. For instance, knowing $I_i, \mathbf{W}_i, \tilde{\alpha}, \tilde{\beta}, Y_i, \Omega_i$, the reduced quote (reservation price) $(V_i - \text{CBBT}_i)/\Delta_i$ of the client i is drawn from a truncated Gaussian distribution. For sampling from such truncated distributions, we used a simple rejection method.

Some of the conditional distributions are not easily available, especially those for the regression parameters in the model with covariates. When it was necessary, we used a Metropolis-within-Gibbs approach: instead of updating using the conditional distribution, we updated the relevant parameter using an iteration of the Metropolis-Hastings algorithm. For example, in our Gibbs algorithm, if the conditional distribution of $\tilde{\beta}$ is not available easily – *i.e.* cannot be simulated directly –, then the following algorithm also samples from the same posterior:

Algorithm 2 Metropolis-within-Gibbs sampler

Initialize (*e.g.* at random) $(\tilde{\alpha}^0, \tilde{\beta}^0)$
for iteration t=1 to T **do**
    Simulate $\tilde{\alpha}^t$ from the conditional distribution $\pi(\tilde{\alpha}|\tilde{\beta}^{t-1}, \mathbf{X})$
    Propose $\tilde{\beta}^* \sim \mathcal{N}(\tilde{\beta}^{t-1}, \varepsilon^2)$ (ε chosen by the user)
    Compute $r = \min\left(1, \frac{\pi(\tilde{\beta}^*|\tilde{\alpha}^t, \mathbf{X})}{\pi(\tilde{\beta}^{t-1}|\tilde{\alpha}^t, \mathbf{X})}\right)$
    Set $\tilde{\beta}^t = \tilde{\beta}^*$ with probability r, and $\tilde{\beta}^t = \tilde{\beta}^{t-1}$ with probability $1 - r$.
end for

---

4 Empirical results

4.1 Model without covariates

4.1.1 General results

In the following paragraphs, we analyze the estimates of the parameters obtained with the MCMC algorithm described previously. We start with the model with no covariates (Equations (2.1) and (2.2)). Visually, the MCMC algorithm leads to stable realizations of the random parameters $(\tilde{\alpha}, \tilde{\beta})$ relatively quickly, for sure after at most 1000 iterations.

We exhibit in Table 1 the means, standard deviations and quartiles[21] of the realizations of the parameters which characterize the SEP distributions of the dealers' quotes.[22] We also exhibit the same statistics for the realizations of the parameters which characterize the Gaussian distributions of the clients' reservation values.[23]

		dealers				clients	
		α	λ	μ	σ	ν	τ
"buy" RFQs	mean	1.41	1.35	-0.61	2.90	1.10	2.41
	$q_{25\%}$	1.39	1.32	-0.63	2.88	1.08	2.39
	$q_{50\%}$	1.40	1.35	-0.61	2.90	1.10	2.42
	$q_{75\%}$	1.42	1.38	-0.60	2.93	1.11	2.44
	std dev.	0.03	0.05	0.03	0.04	0.02	0.03
		α^*	λ^*	μ^*	σ^*	ν^*	τ^*
"sell" RFQs	mean	1.06	-0.73	0.35	2.23	-1.09	2.15
	$q_{25\%}$	1.05	-0.74	0.34	2.21	-1.10	2.13
	$q_{50\%}$	1.07	-0.72	0.35	2.22	-1.09	2.15
	$q_{75\%}$	1.08	-0.71	0.37	2.24	-1.07	2.17
	std dev.	0.02	0.02	0.02	0.02	0.02	0.03

Table 1: Estimation of the model parameters for the "buy" and "sell" RFQs, by MCMC. Statistics computed over the last 1500 iterations among 3000.

All these random parameters have very tight distributions, as indicated through closed quartiles and small standard deviations. Moreover, the means and medians of the parameters are close to one another: this is a nice feature for stating their most likely value.

The probability density functions for dealers' quotes and clients' reservation prices are plotted in Figures 1 and 2 – the parameters being fixed to their empirical means (as provided in Table 1).

---

[21]The statistics are estimated empirically over a sufficiently large number of the last iterations of the Markov process.

[22]We have assumed that all dealers have the same behavior, *i.e.* they draw their quotes from the same SEP distribution. This assumption is questionable in practice – see below.

[23]All clients are assumed to behave similarly here. We have observed empirically significant differences across clients, but such results cannot be detailed in this document for compliance reasons.
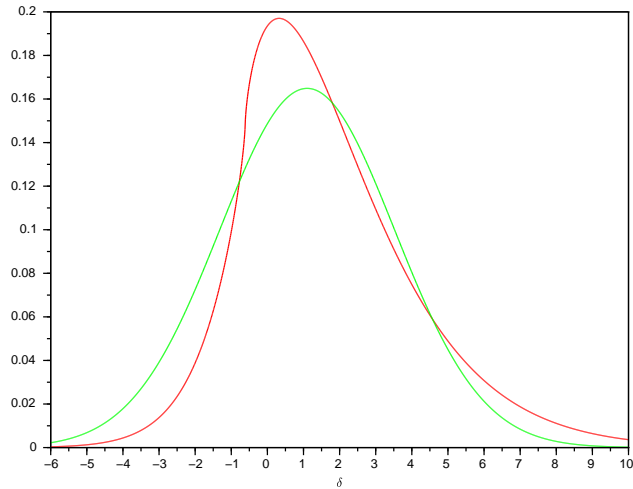
Figure 1: Distributions F_0 and G_0 estimated using the "buy" RFQs. Red line: SEP distribution for the dealers' quotes. Green line: Gaussian distribution for the clients' reservation prices.
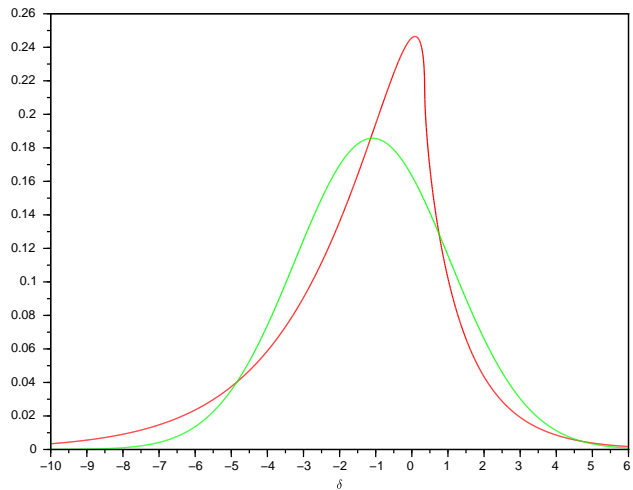

Figure 2: Distributions F_0^* and G_0^* estimated using the "sell" RFQs. Red line: SEP distribution for the dealers' quotes. Green line: Gaussian distribution for the clients' reservation prices.

We see in Figures 1 and 2 that the distributions of the dealers' quotes are clearly asymmetric. To understand the rationale of this empirical result, let us consider the case of a "buy" RFQ. For the dealers, there is almost no difference between answering a high price and a very high price: in both cases, the price will be too high to be accepted by the dealer, and

there will be no trade. However, there is a significant difference between answering a low price and a very low price: in both cases a trade may occur, and the trade price is the price answered by the dealer. The same reasoning applies in the case of a "sell" RFQ, *mutatis mutandis*. This explains the skewness of the distributions of the dealers' quotes.

These distributions are also heavy-tailed (see the value of α and α^* in Table 1). This is an important feature, but it has to be considered with care. A reason why there is a fat tail on the right-hand side (resp. left-hand side) for "buy" (resp. "sell") RFQs is indeed linked to one of the assumptions of our model. In practice, contrary to what we have assumed in the model, some of the requested dealers do not answer effectively: some dealers do not answer fast enough, some other dealers simply do not want to answer, because they are not interested in trading the requested bond, or not interested in dealing with the client who has sent the RFQ. In our model, that is equivalent to answering very conservative prices, hence an effect on the right-hand side (resp. left-hand side) tail.

A visual comparison between the buy and sell cases is made possible by changing the sign, in the "sell" case, of the location and asymmetry parameters μ^* and λ^* of the SEP distribution of dealers' quotes, and the mean ν^* of the Gaussian distribution of clients' reservation values – see Figure 3.[24] It is noteworthy that the probability density function for the dealers' quotes is relatively more "spiky" in the case of "sell" RFQs.
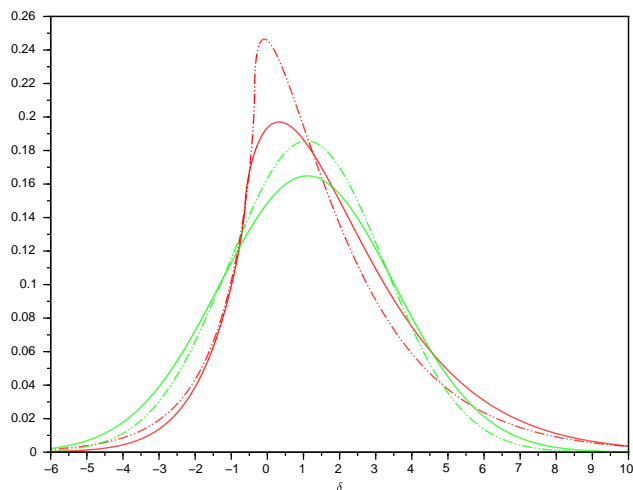

Figure 3: Comparison of the distributions of dealers' quotes and clients' reservation prices for "buy" and "sell" RFQs. Red: SEP distributions for the dealers. Green: Gaussian distributions for the clients. Solid lines represent the case of "buy" RFQs. Dotted lines represent the case of "sell" RFQs, after symmetrization.

---

[24]In other words, in Figure 3, the parameters of the dotted red curve are $(\alpha, \lambda, \mu, \sigma) = (1.06, 0.73, -0.35, 2.23)$, and those of the dotted green curve are $(\nu, \tau) = (1.09, 2.15)$.

The parameters $(\alpha, \lambda, \mu, \sigma)$ and $(\alpha^*, \lambda^*, \mu^*, \sigma^*)$ characterize the SEP distributions of the dealers' quotes. However, as highlighted in Section 2, these parameters are related in a complex way to the moments of the distributions. In particular, μ (resp. μ^*) and σ (resp. σ^*) do not correspond to the mean and the standard deviation of the SEP distributions, because of asymmetry and fat tails. We exhibit the first four moments of F_0 and F_0^* – computed by applying Equations (2.8) and (2.9) – in Table 2.

	mean	variance	skewness	kurtosis
"buy" RFQs	1.660	6.188	0.550	2.471
"sell" RFQs	-1.304	6.360	-1.395	9.851

Table 2: First four moments of the distributions of the dealers' quotes (when the model parameters are fixed to their means – see Table 1).

In the case of "buy" RFQs, dealers tend to answer bond prices above the market price (CBBT): the average bond price answered by dealers, as estimated in our model – see Table 2 –, is $\text{CBBT}_{\text{offer}} + 1.66 \times$ bid-to-mid spread. As far as clients are concerned, the average reservation price is estimated to be $\text{CBBT}_{\text{offer}} + 1.10 \times$ bid-to-mid spread. For "sell" RFQs, the analysis is similar: the average bond price answered by dealers is $\text{CBBT}_{\text{bid}} - 1.304 \times$ bid-to-mid spread. As far as clients are concerned, the average reservation price is estimated to be $\text{CBBT}_{\text{bid}} - 1.09 \times$ bid-to-mid spread. In particular, this confirms our intuition that clients willing to buy (resp. sell) bonds tend to think that they are underpriced (resp. overpriced). It is also noteworthy that the agreement between both sides should be reached more easily for "sell" RFQs than for "buy" RFQs, on average.

The signs of the skewness are not surprising, given the asymmetry observed in Figures 1 and 2. The larger kurtosis in the case F_0^* is linked to the spike around the CBBT, and to what happens in the tail of the distributions. A dealer who really wants to trade with a client tends to quote more aggressively in the case of a "sell" RFQ than in the case of a "buy" RFQ: in Figure 3, more mass appears on the l.h.s. for the dotted red curve than for the solid red curve.[25]

The difference between "buy" and "sell" RFQs is less significant as far as clients' reservation prices are concerned. The average reservation prices $\nu > 0$ and $\nu^* < 0$ have almost the same absolute value, and, more generally, the distributions of ν and $-\nu^*$ are very similar in the MCMC simulations – see Table 1. Moreover, the distributions of the standard deviation parameters τ and τ^* are very similar in the MCMC simulations, even though τ^* is slightly lower on average. Overall, this means that there is a symmetry property between the (optimistic) views of the clients who send "buy" RFQs and the (pessimistic) views of the clients who send "sell" RFQs.

---

[25]Differences in the tails on the other side are less meaningful.

4.1.2 The influence of competition

We now turn to the results of our MCMC method on the different subsamples of RFQs corresponding to different numbers of dealers requested. In other words, instead of estimating F_0, F_0^*, G_0 and G_0^*, we estimate the distributions $F_0(\cdot; n)$, $F_0^*(\cdot; n)$, $G_0(\cdot; n)$, and $G_0^*(\cdot; n)$, defined in Equations (2.5) and (2.6) $- n \in \{1, \ldots, 5\}$.

It is likely that the behavior of dealers and clients depends on the level of competition. In what follows, we aim at answering questions such as: (i) how does the distribution of answered quotes depend on the number of dealers requested?, (ii) are the distributions of clients' reservation prices similar for clients requesting a few dealers and clients requesting a lot of dealers?, (iii) do clients obtain better prices by requesting more dealers?, etc.

We exhibit in Tables 3 and 4 the means, standard deviations and quartiles of the realizations of the parameters which characterize the SEP distributions of the dealers' quotes, and the Gaussian distributions of the clients' reservation prices, for "buy" and "sell" RFQs with different numbers of requested dealers.

		dealers				clients	
		α	λ	μ	σ	ν	τ
$n_i = 1$	mean	1.33	0.22	-0.04	0.91	0.27	3.46
	$q_{25\%}$	1.24	0.18	-0.06	0.89	0.23	3.14
	$q_{50\%}$	1.32	0.21	-0.03	0.91	0.27	3.31
	$q_{75\%}$	1.40	0.24	-0.01	0.93	0.32	3.68
	std dev.	0.12	0.06	0.04	0.04	0.07	0.47
$n_i = 2$	mean	1.42	0.68	-0.27	1.59	0.87	3.40
	$q_{25\%}$	1.36	0.64	-0.30	1.56	0.79	3.23
	$q_{50\%}$	1.42	0.68	-0.27	1.59	0.86	3.38
	$q_{75\%}$	1.48	0.72	-0.24	1.62	0.93	3.54
	std dev.	0.09	0.07	0.05	0.05	0.10	0.24
$n_i = 3$	mean	1.51	1.23	-0.56	2.37	1.11	2.82
	$q_{25\%}$	1.45	1.15	-0.60	2.33	1.05	2.71
	$q_{50\%}$	1.50	1.22	-0.56	2.37	1.10	2.81
	$q_{75\%}$	1.56	1.29	-0.52	2.42	1.17	2.92
	std dev.	0.08	0.11	0.06	0.07	0.09	0.15
$n_i = 4$	mean	1.31	1.18	-0.56	2.72	1.92	3.39
	$q_{25\%}$	1.28	1.14	-0.58	2.70	1.85	3.29
	$q_{50\%}$	1.31	1.17	-0.56	2.72	1.91	3.37
	$q_{75\%}$	1.34	1.21	-0.54	2.75	1.98	3.46
	std dev.	0.05	0.06	0.03	0.04	0.10	0.13
$n_i = 5$	mean	1.46	1.85	-0.85	3.58	1.72	2.90
	$q_{25\%}$	1.43	1.76	-0.88	3.54	1.66	2.81
	$q_{50\%}$	1.46	1.83	-0.85	3.58	1.71	2.88
	$q_{75\%}$	1.49	1.92	-0.83	3.62	1.79	2.97
	std dev.	0.05	0.11	0.04	0.05	0.11	0.13

Table 3: Estimation of the model parameters for the "buy" RFQs, by MCMC. The subsamples correspond to RFQs with different (fixed) numbers of dealers ($n_i = 1$ to 5, *i.e.* from two to six dealers). The statistics are computed over the last 5000 iterations among 10000.

		dealers				clients	
		α^*	λ^*	μ^*	σ^*	ν^*	τ^*
$n_i = 1$	mean	1.28	0.10	0.07	0.89	0.28	3.23
	$q_{25\%}$	1.20	0.04	0.02	0.87	0.25	3.03
	$q_{50\%}$	1.27	0.10	0.06	0.89	0.29	3.19
	$q_{75\%}$	1.36	0.15	0.11	0.91	0.32	3.41
	std dev.	0.12	0.08	0.08	0.03	0.06	0.31
$n_i = 2$	mean	1.16	-0.41	0.26	1.40	-1.42	3.94
	$q_{25\%}$	1.12	-0.43	0.23	1.38	-1.54	3.71
	$q_{50\%}$	1.16	-0.41	0.26	1.40	-1.39	3.89
	$q_{75\%}$	1.21	-0.38	0.28	1.42	-1.29	4.15
	std dev.	0.07	0.04	0.04	0.03	0.21	0.39
$n_i = 3$	mean	1.10	-0.59	0.29	1.77	-1.58	3.12
	$q_{25\%}$	1.05	-0.61	0.26	1.75	-1.67	2.96
	$q_{50\%}$	1.09	-0.58	0.28	1.77	-1.57	3.11
	$q_{75\%}$	1.14	-0.55	0.31	1.80	-1.46	3.24
	std dev.	0.07	0.05	0.04	0.04	0.17	0.24
$n_i = 4$	mean	1.05	-0.71	0.37	2.21	-1.94	3.04
	$q_{25\%}$	1.03	-0.73	0.35	2.19	-2.00	2.95
	$q_{50\%}$	1.04	-0.71	0.37	2.21	-1.92	3.02
	$q_{75\%}$	1.07	-0.69	0.39	2.24	-1.86	3.11
	std dev.	0.03	0.03	0.02	0.03	0.13	0.14
$n_i = 5$	mean	1.23	-1.20	0.70	2.84	-2.00	2.77
	$q_{25\%}$	1.22	-1.22	0.69	2.82	-2.07	2.71
	$q_{50\%}$	1.24	-1.20	0.70	2.84	-2.00	2.77
	$q_{75\%}$	1.25	-1.17	0.71	2.87	-1.94	2.84
	std dev.	0.03	0.04	0.02	0.03	0.09	0.09

Table 4: Estimation of the model parameters for the "sell" RFQs, by MCMC. The subsamples correspond to RFQs with different (fixed) numbers of dealers ($n_i = 1$ to 5, *i.e.* from two to six dealers). The statistics are computed over the last 5000 iterations among 10000.

		mean	variance	skewness	kurtosis
"buy"	$n_i = 1$	0.046	1.381	0.004	4.395
	$n_i = 2$	0.650	2.560	0.401	4.188
	$n_i = 3$	1.169	4.027	0.395	2.127
	$n_i = 4$	1.571	6.133	0.683	3.323
	$n_i = 5$	2.082	8.193	0.497	1.962
"sell"	$n_i = 1$	0.169	1.154	0.109	9.341
	$n_i = 2$	-0.402	2.758	-0.656	9.714
	$n_i = 3$	-0.856	4.320	-1.146	10.436
	$n_i = 4$	-1.285	6.491	-1.490	10.672
	$n_i = 5$	-1.628	7.117	-0.878	4.114

Table 5: First four moments of the distributions $F_0(\cdot; n)$ and $F_0^*(\cdot; n)$ of the dealers' quotes (when the model parameters are fixed to their means – see Tables 3 and 4).
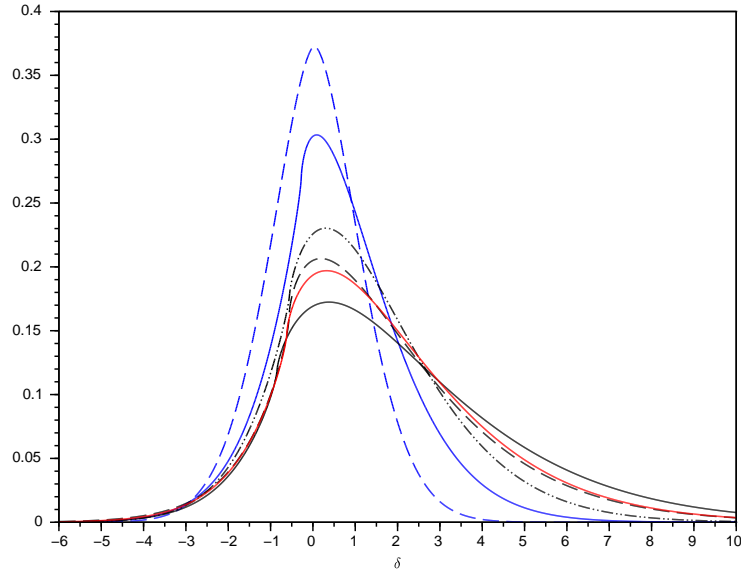
Figure 4: SEP distributions $(F_0(\cdot; n))_n$ for the dealers' quotes on the subsamples of "buy" RFQs. Blue dashed line: $n = 1$. Blue solid line: $n = 2$. Black dash-dotted line: $n = 3$. Black dotted line: $n = 4$. Black solid line: $n = 5$. Red line: all "buy" RFQs.
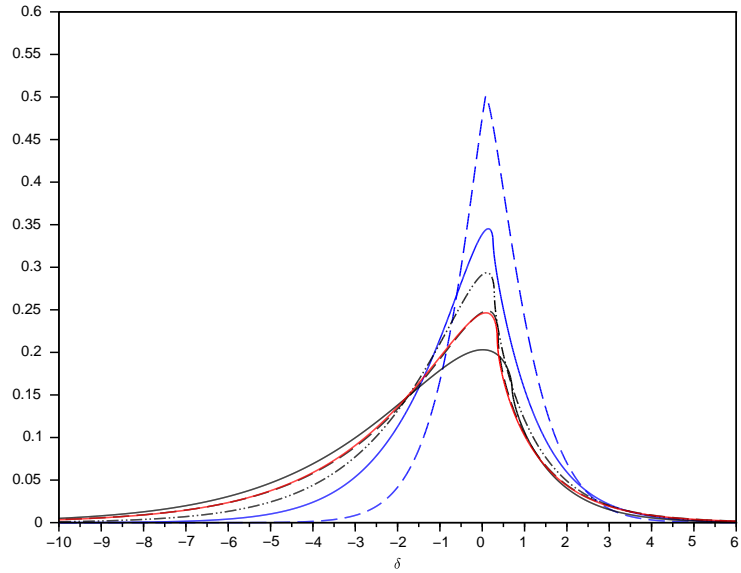

Figure 5: SEP distributions $(F_0^*(\cdot; n))_n$ for the dealers' quotes on the subsamples of "sell" RFQs. Blue dashed line: $n = 1$. Blue solid line: $n = 2$. Black dash-dotted line: $n = 3$. Black dotted line: $n = 4$. Black solid line: $n = 5$. Red line: all "sell" RFQs.

As in the global case, we see in Tables 3 and 4 that the random parameters have very tight distributions: quartiles are very close and standard deviations are small. Moreover, the means and medians of the parameters are very close.

For both "buy" and "sell" RFQs, the estimates of the location (μ, μ^*), scale (σ, σ^*), and asymmetry (λ, λ^*) parameters of the distributions of dealers' quotes tend to behave monotonically with the number of requested dealers.

In Figures 4 and 5, we clearly see that the distribution of dealers' quotes depends on the number of dealers requested in a very specific and ordered way. In particular, in the case of "buy" (resp. "sell") RFQs, the more dealers in competition, the higher (resp. lower) their answered quotes. In other words, the more dealers in competition, the more conservative their answered quotes. The evolution of the first moment (the mean) of the dealers' quotes distributions corroborates this finding – see Table 5.

"Discouragement" is a way to explain this phenomenon. When a dealer answers to a RFQ sent to a few dealers only, he may think that his/her effort to propose a good price will lead to a deal, because competition is not strong. Conversely, in the case of a RFQ sent to many dealers, he may think that there is little chance for him to be chosen, and therefore no reason to spend time choosing a relevant non-conservative price.

Another explanation is related to clients. A client requesting only a few dealers may be a client who has a close/preferential relationship with one or all of the dealers requested. Therefore, the effect under scrutiny may be due to dealers answering better prices to "their" important clients, than to other clients.

An alternative explanation is related to a selection bias: when a client requests $n+1$ dealers, he may simply pick the $n+1$ dealers who have streamed the best $n+1$ prices. Therefore, if we assume that there is a positive correlation between streamed prices and answered prices – a natural assumption –, then dealers' quotes should be more aggressive (by construction) when only a few dealers are requested than when a lot of dealers are requested. The latter explanation requires another model, because dealers' quotes are assumed to be *i.i.d.* in our model.

Another "off-model" possible explanation is related to dealers who do not answer: the larger the number of requested dealers, the higher the probability that a dealer does not have time to answer, for instance because the client has traded with an early-answerer. In our model, this effect artificially leads to an increase in the estimated probability of conservative answered quotes.

In Tables 3 and 4, we also see that the Gaussian distributions of the clients' reservation prices depend on the number of dealers requested – see also Figures 6 and 7. It is noteworthy that, in the case of "buy" RFQs, the mean parameter ν tends to increase with n_i.

Conversely, in the case of "sell" RFQs, the mean parameter ν^* tends to decrease with n_i.[26]
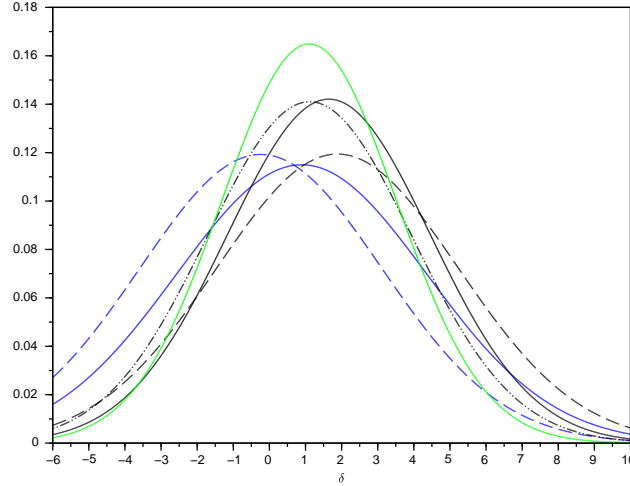

Figure 6: Gaussian distributions $(G_0(\cdot; n))_n$ for the clients' reservation values on the sub-samples of "buy" RFQs. Blue dashed line: $n = 1$. Blue solid line: $n = 2$. Black dash-dotted line: $n = 3$. Black dotted line: $n = 4$. Black solid line: $n = 5$. Green line: all "buy" RFQs.
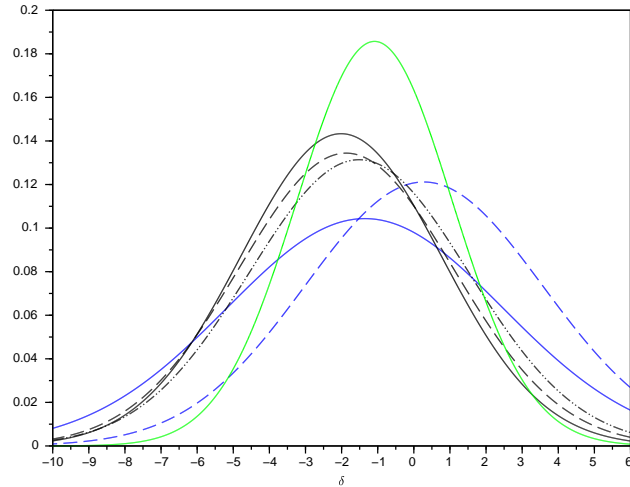

Figure 7: Gaussian distributions $(G_0^*(\cdot; n))_n$ for the clients' reservation values on the sub-samples of "sell" RFQs. Blue dashed line: $n = 1$. Blue solid line: $n = 2$. Black dash-dotted line: $n = 3$. Black dotted line: $n = 4$. Black solid line: $n = 5$. Green line: all "sell" RFQs.

---

[26]As far as variances τ and τ^* are concerned, there is no significant pattern: the values are similar for different n_i, and for both "buy" and "sell" RFQs.

A possible interpretation of this monotonicity property is that clients who are very confident in their views about a bond they want to buy or sell – *i.e.*, based on the streamed prices and the information they have, they strongly believe that the bond is undervalued or overvalued – tend to contact more dealers to be sure to quickly obtain an acceptable price. That may explain part of the effect observed, but it does not explain why there is such a difference in the clients' behavior between RFQs involving 2 or 3 dealers and RFQs involving 4, 5, or 6 dealers – see Figures 6 and 7.

Another explanation may be that some (very demanding) clients are sending RFQs to the 2 or 3 dealers who have streamed the most interesting prices, because they do not expect others to propose interesting prices. Similarly, opportunistic clients may only send RFQs when they notice interesting prices streamed by one or several dealers. In that case, they may contact these specific dealer(s), or if there is only one, the specific dealer and another one, for proving they receive best execution. These explanations are coherent with the dependence of the distributions of dealers' quotes on the number of dealers. However, once again, these explanation are "off-model" ones, because we have assumed identically distributed answered quotes across dealers.

An alternative "off-model" rationale may also be that some (informal) agreements between clients and dealers are reached outside of the MD2C platform: an interesting price may be proposed by a dealer to a client by telephone or chat, and then the formal agreement reached on the platform after the client had requested a few dealers – for proving they receive best execution.

Using the distributions of dealers' quotes, it is also possible to compute an estimate of the distribution of the best price proposed to clients, for the different values of the number of dealers requested.

In the case of a "buy" RFQ, the probability density function of the minimum of the dealers' quotes – including the "reference dealer", who is assumed to behave as a similar additional dealer – is given by:

$$\delta \mapsto (n_i + 1) f_0(\delta; n_i) \left(1 - F_0(\delta; n_i)\right)^{n_i},$$

where $n_i + 1$ is the total number of dealers requested (including the reference dealer). This probability density function is plotted in Figure 8.

For "sell" RFQs, the best price proposed to a client is the maximum of the dealers' quotes. The associated probability density function is

$$\delta \mapsto (n_i + 1) f_0^*(\delta; n_i) F_0^*(\delta; n_i)^{n_i},$$

when n_i dealers are requested in addition to the reference dealer. This probability density function is plotted in Figure 9.
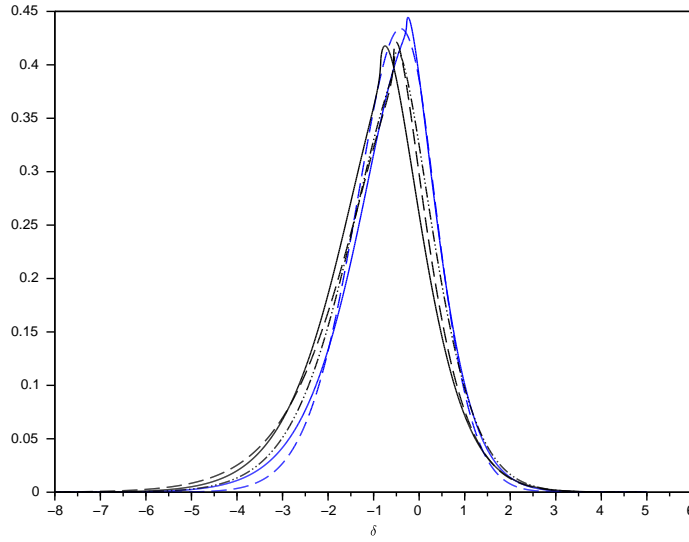
Figure 8: Distribution of the best (reduced) quote $(\min_k W_{k,i} - CBBT_i)/\Delta_i$ proposed to clients, calculated on each subsample of "buy" RFQs. Blue dashed line: $n = 1$. Blue solid line: $n = 2$. Black dash-dotted line: $n = 3$. Black dotted line: $n = 4$. Black solid line: $n = 5$.

Figure 9: Distribution of the best (reduced) quote $(\max_k W_{k,i} - CBBT_i)/\Delta_i$ proposed to clients, calculated on each subsample of "sell" RFQs. Blue dashed line: $n = 1$. Blue solid line: $n = 2$. Black dash-dotted line: $n = 3$. Black dotted line: $n = 4$. Black solid line: $n = 5$.
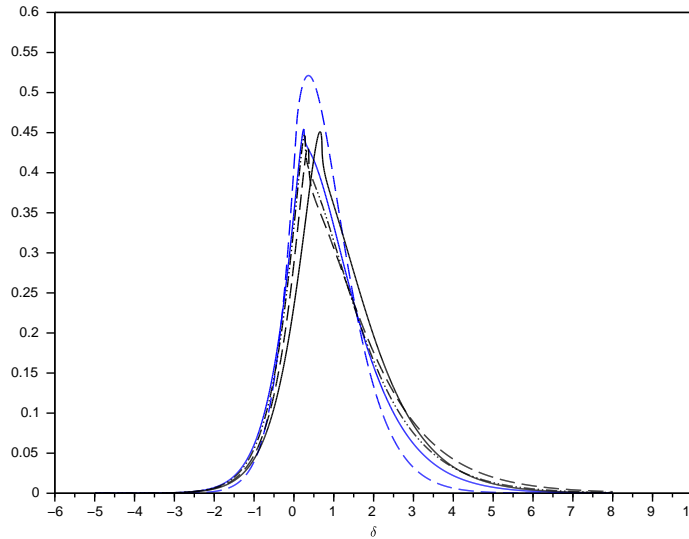
We have seen previously that, on average, dealers are more prone to answer interesting prices to clients in the case of RFQs involving a few dealers only. Nevertheless, we see in Figures 8 and 9 that clients should expect to get better prices if they send RFQs to a lot of dealers. This is not contradictory, because requesting more dealers increases the range of prices obtained, and clients are eventually only interested in the best price. For the sake of completeness, we have provided in Table 6 the means of the distributions in Figures 8 and 9. The average of the best price proposed to the clients tends to behaves monotonically with the number of dealers requested, as expected through the visual inspection of the previous figures.

total number of dealers	"buy" RFQs	"sell" RFQs
2	−0.611	0.757
3	−0.641	0.917
4	−0.735	1.065
5	−0.940	1.257
6	−0.951	1.197

Table 6: Average of the best (reduced) quotes proposed to clients, for different levels of competition.

Practitioners like to measure hit ratios. Hit ratios correspond to the probability to trade, given that the RFQ led to a trade with one of the dealers.

For "buy" RFQs, if n_i dealers are requested in addition to the reference dealer, the hit ratio is the following function of the (reduced) quote answered by the dealer:

$$\delta \mapsto (1 - F_0(\delta; n_i))^{n_i}.$$

For "sell" RFQs, if n_i dealers are requested in addition to the reference dealer, the hit ratio is the following function of the (reduced) quote answered by the dealer:

$$\delta \mapsto F_0^*(\delta; n_i)^{n_i}.$$

These hit ratios are plotted in Figures 10 and 11. Unsurprisingly, the hit ratios are monotonic functions of the price answered by the dealer: in the case of a "buy" (resp. "sell") RFQ, the probability to propose the best price decreases (resp. increases) with the price. Furthermore, hit ratios are decreasing functions of the number of requested dealers: the more dealers, the lower the probability to be the one who offers the best price.
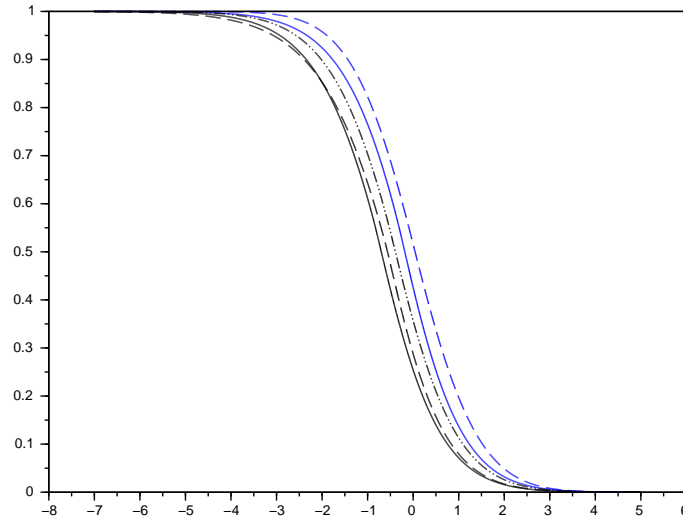
Figure 10: Hit ratios for a "buy" RFQ, as a function of the answered (reduced) quote, for each possible value of the number of competing dealers. Blue dashed line: $n = 1$. Blue solid line: $n = 2$. Black dash-dotted line: $n = 3$. Black dotted line: $n = 4$. Black solid line: $n = 5$.

Figure 11: Hit ratios for a "sell" RFQ, as a function of the answered (reduced) quote, for each possible value of the number of competing dealers. Blue dashed line: $n = 1$. Blue solid line: $n = 2$. Black dash-dotted line: $n = 3$. Black dotted line: $n = 4$. Black solid line: $n = 5$.
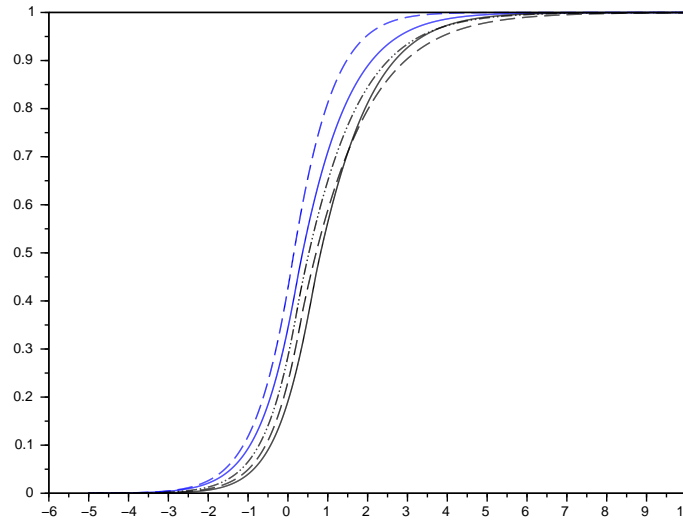
In addition to hit ratios, we can also compute the probability to trade, given the price proposed and the number of dealers requested. For "buy" RFQs, if n_i dealers are requested in addition to the reference dealer, this probability is given, as a function of the dealer's (reduced) quote, by:

$$\delta \mapsto (1 - F_0(\delta; n_i))^{n_i} (1 - G_0(\delta; n_i)).$$

For "sell" RFQs, if n_i dealers are requested in addition to the reference dealer, this probability is given, as a function of the dealer's (reduced) quote, by:

$$\delta \mapsto F_0^*(\delta; n_i)^{n_i} G_0^*(\delta; n_i).$$

The probability to trade is a monotonic function of the price answered by the dealer: in the case of a "buy" (resp. "sell") RFQ, both the probability to propose the best price and to propose a price that the client is going to accept decrease (resp. increase) with the price. However, unlike what happens with hit ratios, the probability to trade is an increasing function of the number of requested dealers. For a given price answered by the dealer, two effects are present: (i) the probability to propose the best price decreases with the number of competing dealers, and (ii) the probability to propose a price that reaches the client's reservation value increases with the number of competing dealers – because clients requesting only a few dealers tend to be more demanding (see the previous discussions and Figures 6 and 7). We see in Figures 12 and 13 that the latter effect dominates.[27]
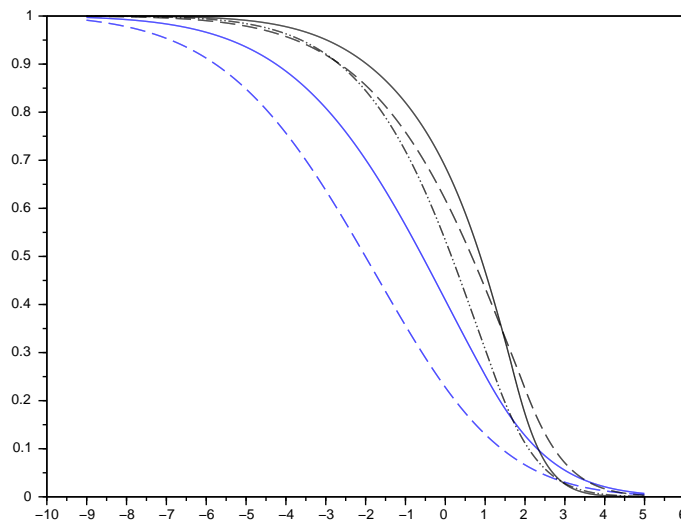

Figure 12: Model probability of closing a "buy" deal, as a function of the answered (reduced) quote, for each possible value of the number of competing dealers. Blue dashed line: $n = 1$. Blue solid line: $n = 2$. Black dash-dotted line: $n = 3$. Black dotted line: $n = 4$. Black solid line: $n = 5$.

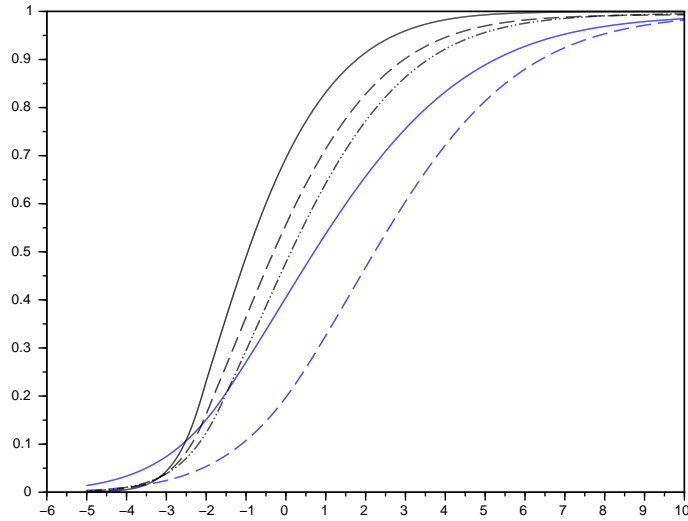[27]Except in the tail, on the irrelevant side.

Figure 13: Model probability of closing a "sell" RFQ deal, as a function of the answered (reduced) quote, for each possible value of the number of competing dealers. Blue dashed line: $n = 1$. Blue solid line: $n = 2$. Black dash-dotted line: $n = 3$. Black dotted line: $n = 4$. Black solid line: $n = 5$.

4.2 Model with covariates

Now, let us introduce in the model some covariates related to the RFQs and the underlying bonds. We choose the specifications of Equations (2.3) and (2.4), with the following variables:

- a single dummy indicator related to bond seniority: it distinguishes "Senior" corporate bonds from all the other types of bonds (the reference category);

- a continuous variable called "Centered Price", defined as the difference between the bond CBBT and the median of corporate bond prices in the database. The idea is to measure very roughly the riskiness associated with the bond (*e.g.* to capture effects specific to "junk bonds"), or to capture its atypical features;

- two dummy variables "Investment Grade" and "High-Yield", directly linked to the usual corporate bond ratings. The reference category consists of bonds issued by financial institutions (excluded from the classification "Investment Grade"/"High-Yield" in our database);

- a continuous variable called "LogNotional", defined as the logarithm of the notional of the RFQ (in euros);

- eight dummy variables that are related to the sector of the issuer: "Bank", "Financial", "Manufacturing", "Retail", "Sovereign", "Telecommunication/Media", "Transportation"

27

and "Utilities". Bonds issued by firms from other sectors are merged in the reference category "Others".

In Equations (2.3) and (2.4), we separated the variables – stacked in Z_1 – that have a direct influence on prices, from the variables – stacked in Z_2 – that have an influence on the reduced quotes. In the model of this section, all the variables belong to the sub-vector Z_2, except the variable "Centered Price", which is part of Z_1.

By using the MCMC method presented previously (see also Appendix B), we have estimated the "full" model, *i.e.* including all these variables simultaneously. The results are gathered in Tables 7 and 8.

To save space, we have not detailed the new estimates of the parameters characterizing the SEP distributions of the dealers' quotes and the Gaussian distributions of the clients' reservation prices. They are consistent with the figures we obtained in the previous subsection, and can be provided upon request. In particular, the convergence of the MCMC method is clear for these parameters. Moreover, for each of these parameters, the quartiles of the distribution are close, and the average value is close to the median.

		Senior	Centered Price	Bond type		LogNotional
				Investment Grade	High-Yield	
"Buy" RFQ, dealers	mean	0.060	-0.005	0.009	-0.064	0.035
	$q_{25\%}$	0.012	-0.005	0.007	-0.051	0.030
	$q_{50\%}$	0.028	-0.004	0.010	-0.041	0.035
	$q_{75\%}$	0.039	-0.003	0.013	-0.030	0.039
	std dev.	0.050	0.000	0.000	0.004	0.011
"Buy" RFQ, clients	mean	0.089	0.010	0.109	0.064	-0.344
	$q_{25\%}$	-0.017	0.006	0.113	0.101	-0.375
	$q_{50\%}$	0.038	0.008	0.122	0.121	-0.360
	$q_{75\%}$	0.074	0.010	0.129	0.144	-0.339
	std dev.	0.093	0.005	0.000	0.036	0.005
"Sell" RFQ, dealers	mean	-0.072	0.004	-0.004	0.001	-0.020
	$q_{25\%}$	-0.087	0.004	-0.003	-0.021	-0.022
	$q_{50\%}$	-0.073	0.005	-0.002	-0.014	-0.021
	$q_{75\%}$	-0.066	0.006	0.001	-0.007	-0.020
	std dev.	0.002	0.000	0.002	0.000	0.000
"Sell" RFQ, clients	mean	-0.295	-0.016	-0.121	0.035	0.350
	$q_{25\%}$	-0.277	-0.018	-0.127	0.007	0.331
	$q_{50\%}$	-0.238	-0.016	-0.119	0.031	0.344
	$q_{75\%}$	-0.204	-0.014	-0.113	0.051	0.360
	std dev.	0.038	0.001	0.021	0.052	0.021

Table 7: Estimation of the effect of the covariates, by MCMC. Statistics computed over the last 5000 iterations among 10000.

		Bank	Financial	Manufact.	Retail	Sovereign	Tel./Media	Transport.	Utilities
"Buy" RFQ, dealers	mean	-0.094	-0.049	0.014	-0.077	-0.072	0.222	-0.091	0.079
	$q_{25\%}$	-0.157	-0.069	-0.031	-0.081	-0.158	0.207	-0.125	0.033
	$q_{50\%}$	-0.135	-0.058	-0.022	-0.070	-0.129	0.247	-0.113	0.055
	$q_{75\%}$	-0.101	-0.043	-0.008	-0.051	-0.096	0.277	-0.098	0.083
	std dev.	0.024	0.018	0.020	0.010	0.011	0.071	0.009	0.001
"Buy" RFQ, clients	mean	0.016	-0.098	0.081	0.273	-0.113	-0.019	-0.369	0.061
	$q_{25\%}$	0.002	-0.160	0.072	0.067	-0.188	-0.115	-0.461	0.034
	$q_{50\%}$	0.078	-0.108	0.115	0.150	-0.138	-0.052	-0.359	0.095
	$q_{75\%}$	0.128	-0.052	0.181	0.191	-0.087	0.014	-0.291	0.221
	std dev.	0.033	0.111	0.091	0.117	0.078	0.121	0.164	0.142
"Sell" RFQ, dealers	mean	0.049	0.066	0.156	0.075	-0.068	-0.031	-0.117	-0.036
	$q_{25\%}$	0.048	0.070	0.138	0.049	-0.076	-0.053	-0.189	-0.103
	$q_{50\%}$	0.059	0.075	0.148	0.054	-0.071	-0.033	-0.167	-0.093
	$q_{75\%}$	0.065	0.100	0.154	0.080	-0.065	-0.010	-0.137	-0.064
	std dev.	0.015	0.007	0.033	0.010	0.004	0.052	0.000	0.040
"Sell" RFQ, clients	mean	-0.425	-0.129	-0.049	-0.324	0.202	0.015	0.060	-0.187
	$q_{25\%}$	-0.466	-0.106	-0.244	-0.377	0.087	0.032	0.122	-0.242
	$q_{50\%}$	-0.429	-0.070	-0.172	-0.237	0.140	0.059	0.154	-0.207
	$q_{75\%}$	-0.403	-0.040	-0.111	-0.140	0.181	0.085	0.195	-0.138
	std dev.	0.138	0.141	0.080	0.235	0.171	0.041	0.099	0.036

Table 8: Estimation of the influence of the sector of the issuer, by MCMC. Statistics computed over the last 5000 iterations among 10000.

As far as the parameters associated with the covariates are concerned, the picture is less rosy. For some of the parameters, we indeed observe in Tables 7 and 8 a wide spread between the first quartile $q_{25\%}$ and the third quartile $q_{75\%}$, and an average value of the parameter outside of the interval $[q_{25\%}, q_{75\%}]$. This invites to regard the median as a more meaningful estimate than the average. Nevertheless, interesting facts can be deduced from our estimations.

We see in Table 7 that clients are, in general, more sensitive than dealers to the bond characteristics and the requested amount. In other words, clients tend to take them more into account when they evaluate the price at which they are ready to buy or sell a bond, than dealers take them into account when choosing the quote they answer.

The most interesting phenomenon captured by our model is certainly the influence of the notional. The clients who want to buy a large quantity of bonds require from dealers a large discount to accept to trade. In other words, a dealer willing to trade with a client who has sent a "buy" RFQ with a large nominal should be ready to propose a low price. A similar phenomenon can be observed for "sell" RFQs too, with the same amplitude: the larger the notional, the higher the reservation price of clients when they want to sell bonds. Dealers, however, do not seem to be ready to follow these clients' expectations: in the case of "buy" RFQs with large notional, they tend to shift upward the quotes they answer, while, in the case of "sell" RFQs with large notional, they tend to shift downward the quotes they answer. Dealers' behavior[28] reflects two risks associated with large transactions: the classical inventory risk faced by market makers, and a risk related to adverse selection. On average, reaching agreement between dealers and clients is more complex in the case of RFQs with large notional. However, in practice, specific dealers might be willing to accept the discount/premium expected by the clients, when the side and the notional of the RFQ is such that making the transaction would reduce significantly their inventory.

---

[28]It is noteworthy that the amplitude of the quote adjustment made by dealers is minor, compared to the discount/premium expected by clients.

All other things being equal, clients are ready to accept lower prices for senior corporate bonds in a "sell" RFQ, and dealers indeed propose lower quotes. For a "buy" RFQ, we observe the symmetrical effect, but a lot weakened.[29]

The effect of the variable "Centered price" is rather weak. This is a bit similar for bond ratings: the only interesting effect is that clients tend to accept slightly larger (resp. smaller) prices for buying (resp. selling) investment grade bonds.

Even though convergence of the parameters are not always insured (see the relatively large standard deviations), we can analyze the effect of the industrial sector. In terms of magnitude, the sector can induce a shift of half the bid-to-mid price at most. In our database,[30] all other things being equal, clients seem to show an appetite for buying and selling bonds from the "Retail" sector (which includes large retailers, luxury, clothing, tobacco, packaging, etc.): they accept to pay more to buy and accept to sell at lower prices. Conversely, they show an aversion for the "Transportation" sector. Interestingly, when clients want to sell bonds issued by banks or utilities, they accept a significant discount. As far as dealers are concerned, they tend to overvalue bonds issued by telecommunication and/or media firms (for "buy" RFQs), and those issued by manufacturing (for "sell" RFQs). These conclusions related to the influence of the industrial sectors are fragile and may change over time.

5   Extensions

The model we have introduced in this paper makes it possible to analyze the behavior of both dealers and clients. However, we have seen in Section 4 that some of the results obtained could be related to phenomenons that are incompatible with our initial (simple) assumptions. In this last section, we propose several ways to extend our model, for it to be more realistic.

5.1   The "true" number of competing dealers

For a given RFQ i, we suspect that the number of requested dealers $n_i + 1$ will not always be the number of dealers ready to answer a price. Some of the requested dealers may not be on their desk when the RFQ is sent. Some may not be interested in dealing with the client who has sent the RFQ – $e.g.$ because the client is not an important one for them, or because they fear adverse selection. Some others may not be interested in buying or selling the bond requested, because of their inventory. This means that the "true" number of competing dealers, denoted by $\tilde{n}_i + 1$, has many reasons to be less than $n_i + 1$.

In order to take this phenomenon into account, our model and the associated likelihood equations have to be modified. If $\tilde{n}_i$ were observable, these equations could be written by

---

[29]Figures are however arguable in the case of "buy" RFQs, given the difference between the average and the median of the parameters, and the standard deviations in the MCMC simulation.

[30]Our database covers part of 2013 and 2014.

replacing the theoretical quantity n_i by the true value $\tilde{n}_i$. Unfortunately, the latter numbers are not observable.

In order to enrich our model, a possibility is to assume that the dealers have a given probability p_i (resp. p_i^*) of answering a "buy" (resp. "sell") RFQ i. More precisely, we can model the participation of each requested dealer (in addition to the reference dealer[31]) by a Bernoulli random variable with parameter p_i (resp. p_i^*). If we assume that the n_i Bernoulli random variables are independent, with the same value of p_i (resp. p_i^*) across dealers, then the number of dealers who indeed answer to the RFQ i (in addition to the reference dealer) is a random variable distributed according to a binomial distribution $\mathcal{B}(n_i, p_i)$ (resp. $\mathcal{B}(n_i, p_i^*)$).

Under this assumption, the likelihood associated with a "buy" RFQ i is given by a mixture distribution, *i.e.* it is a weighted average of several likelihoods given $\tilde{n}_i$:

$$\mathcal{L}_{i|\text{buy}}(X_i) = \sum_{k=0}^{n_i} C_{n_i}^k p_i^k (1 - p_i)^{n_i - k} \mathcal{L}_{i|\text{buy}}(X_i | \tilde{n}_i = k).$$

Similarly, in the case of a "sell" RFQ i, it is

$$\mathcal{L}_{i|\text{sell}}(X_i) = \sum_{k=0}^{n_i} C_{n_i}^k p_i^{*k} (1 - p_i^*)^{n_i - k} \mathcal{L}_{i|\text{sell}}(X_i | \tilde{n}_i = k).$$

An important point is then to choose / estimate the parameters p_i and p_i^*. They may depend on the identity of the client (some clients are not seen as "strategic" for most dealers, some clients may be associated with a risk of being adversely selected, etc.), the type of bond (some bonds may be too illiquid, regarded as too risky, or they may not belong to the family of securities most dealers are interested in), the notional of the RFQ (RFQs with large nominal may cause risk management problems), etc.

More importantly, p_i or p_i^* may depend on n_i itself. Our analysis of Figures 4 and 5 suggests that the larger the number of requested dealers, the less likely each dealer really participates to the RFQ. A simple approach consists in setting $p_i = p(n_i)$ (resp. $p_i^* = p^*(n_i)$). That leads to 10 new parameters, because $n_i \in \{1, \ldots, 5\}$, and because we separate "buy" and "sell" RFQs.

For estimating these additional parameters, the best suited way is through an adaptation of our previous Bayesian framework. In that case, the variables $\tilde{n}_i$ are additional unobservable latent variables which play a similar role as the latent variables $W_{k,i}$ and V_i. Given a value of $\tilde{n}_i$, the underlying distributions of the observations are similar to the ones in our initial model. Therefore, the MCMC algorithm should be adapted to tackle such an extended framework.

---

[31] We only consider RFQs with a price answered by the reference dealer.

5.2 A mixture distribution for the dealers' quotes

In the previous paragraphs, we have assumed that, for a given RFQ, some dealers answer a price and some do not. Another – less parsimonious – way of introducing heterogenous behaviors among dealers is to assume that the answer of dealers can be of two types:

(i) If a dealer wants to trade with the client and seeks to propose the best price, then we assume that his/her (aggressive) answered quote is distributed according to F_1 (resp. F_1^*).

(ii) If a dealer is not really interested in trading the requested bond with the client, but still willing to answer a (conservative) price, then his/her answered quote is distributed according to F_2 (resp. F_2^*).

For a given "buy" (resp. "sell") RFQ i and a given dealer, the probability to be in case (i) is denoted by q_i (resp. q_i^*). In other words, in the case of a "buy" RFQ, the quote is distributed according to F_1 with probability q_i, and distributed according to F_2 with probability $1 - q_i$. Similarly, in the case of a "sell" RFQ, the quote is distributed according to F_1^* with probability q_i^*, and distributed according to F_2^* with probability $1 - q_i^*$. As in Subsection 5.1, the probabilities q_i and q_i^* can depend on the client identity, the bond characteristics, etc. They can also depend on the number of requested dealers $n_i + 1$. In particular, by assuming that the probability to quote conservatively increases with the number of requested dealers, it may be possible to replicate the phenomenon observed in Figures 4 and 5.

It is noteworthy that the extension proposed in Subsection 5.1 can be regarded as a special case of this one, because the former boils down to the latter in the limit case where F_2 (resp. F_2^*) is a Dirac mass at $+\infty$ (resp. $-\infty$).

This extended framework leads to a more complex model: it may capture and explain additional effects, but the statistical inference is more challenging. If we choose F_1 and F_2 (resp. F_1^* and F_2^*) inside the same parametric family of distributions (with different parameters), then we multiply by two the number of parameters to estimate on the dealer side – and we also have to estimate q_i (resp. q_i^*). F_2 (resp. F_2^*) can also be chosen in a simpler family of distributions in order to reduce the number of parameters to estimate.

In all cases, the MCMC algorithm used throughout this paper can be generalized to tackle this extended framework.

Conclusion

In this paper, we have introduced a model to infer the behavior of dealers and clients on the corporate bond market from a hitherto unexploited[32] database of RFQs.

---

[32]Although practitioners have been using this kind of dataset for several years, it is the first time an academic research work is based on such a type of dataset.

We have modelled in a simple way the RFQ process on MD2C platforms by assuming skewed exponential power distributions for dealers' quotes and Gaussian distributions for clients' reservation prices. The model has been estimated by Markov chain Monte Carlo techniques (Gibbs sampling and Metropolis-Hastings random draws).

We have studied the influence of the number of competing dealers, the notional of the RFQ, and some bond characteristics. Our model sheds light on the dependence of the behaviors of both clients and dealers on the number of dealers requested.

Some extensions of the model have been proposed to tackle the potential problem of an unknown number of responding dealers and/or the heterogeneity of their behaviors. These extensions will be the topic of future research works.

References

[1] Avellaneda, M., Stoikov, S. (2008). High-frequency trading in a limit order book. Quantitative Finance, 8(3), 217-224.

[2] Azzalini, A. (1985). A Class of Distributions Which Includes the Normal Ones. *Scandinavian Journal of Statistics* **12**, 171-178.

[3] Azzalini, A. (1986). Further Results on a Class of Distributions Which Includes the Normal Ones. *Statistics* **46**, 199-208.

[4] DiCiccio, T., Monti, A. C. (2004). Inferential aspects of the skew exponential power distribution. *JASA*, **99** 439-450.

[5] Guéant, O., Lehalle, C.-A. (2013). General intensity shapes in optimal liquidation. Mathematical Finance, to appear.

[6] Guéant, O., Lehalle, C.-A., Fernandez-Tapia, J. (2013). Dealing with the inventory risk: a solution to the market making problem. Mathematics and financial economics, 7(4), 477-507.

[7] Guéant, O. The Financial Mathematics of Market Liquidity: From Optimal Execution to Market Making. CRC Press, to appear.

[8] Ho, T., Stoll, H. R. (1981). Optimal dealer pricing under transactions and return uncertainty. Journal of Financial economics, 9(1), 47-73.

[9] Ho, T. S., Stoll, H. R. (1983). The dynamics of dealer markets under competition. The Journal of finance, 38(4), 1053-1074.

[10] Kozubowski, T. J., Podgorski, K. (2001). Asymmetric Laplace Laws and Modeling Financial Data. *Mathematical and Computer Modeling*, 34(9):1003-1021.

[11] Mc Kinsey (2013), Corporate bond e-trading: same game, new playing field.

[12] Pelletier, F. (2014). Modélisation des rendements financiers à l'aide de la distribution de Laplace asymétrique généralisée. Univ. Laval, Québec.

[13] Robert, C. P. (1995). Simulation of truncated normal variables. *Statistics and Computing* **5**, 121-125.

[14] Robert, C. P., Casella, G. (2004). *Monte Carlo Statistical Methods*, Springer.

[15] Sanchez, L. B., Lachos, V. H., Labra, F. V. (2013). Likelihood Based Inference for Quantile Regression Using the Asymmetric Laplace Distribution. Working paper.

[16] Scott-Quinn, B., Cano, D. (2013), European Corporate Bond Trading – the role of the buy-side in pricing and liquidity provision

[17] Subbotin, M. T. (1923). On the Law of Frequency of Error. *Mathematicheskii Sbornik* 296-300.

A The likelihood

Let us detail the likelihood associated with the sample $\mathcal{S}_N$.

a. If $I_i = 1$ (Done), then the likelihood of X_i in the case of a "buy" order is

$$\mathcal{L}_{i|\text{buy}}^{(1)} = P\left(\min_{k=1,\dots,n_i} W_{k,i} = C_i, V_i \geq Y_i | \Omega_i\right) = n_i f(C_i|\Omega)(1 - F(C_i|\Omega_i))^{n_i-1}(1 - G(Y_i|\Omega_i)),$$

if we know the cover price.

In the case of a "sell" order, it is

$$\mathcal{L}_{i|\text{sell}}^{(1)} = P\left(\max_{k=1,\dots,n_i} W_{k,i} = C_i, V_i \leq Y_i | \Omega_i\right) = n_i f^*(C_i|\Omega) F^*(C_i|\Omega_i)^{n_i-1} G^*(Y_i|\Omega_i),$$

if we know the cover price.

When the cover price is unknown, the likelihood of X_i can be written as

$$\mathcal{L}_{i|\text{buy}}^{(1)} = P\left(\min_{k=1,\dots,n_i} W_{k,i} \geq Y_i, V_i \geq Y_i | \Omega_i\right) = (1 - F(Y_i|\Omega_i))^{n_i}(1 - G(Y_i|\Omega_i)),$$

for a "buy" RFQ, and

$$\mathcal{L}_{i|\text{sell}}^{(1)} = P\left(\max_{k=1,\dots,n_i} W_{k,i} \leq Y_i, V_i \leq Y_i | \Omega_i\right) = F^*(Y_i|\Omega_i)^{n_i} G^*(Y_i|\Omega_i),$$

for a "sell" RFQ.

b. If $I_i = 2$ (Traded Away), then the likelihood of X_i in the "buy" case is

$$
\begin{aligned}
\mathcal{L}_{i|\text{buy}}^{(2)} &= P\left(\min_{k=1,\dots,n_i} W_{k,i} \leq \min(V_i, Y_i) | \Omega_i\right) \\
&= E\left[\left(1 - E\left[\mathbf{1}_{\min_{k=1,\dots,n_i} W_{k,i} \geq \min(V_i,Y_i)} | V_i, \Omega_i\right]\right) | \Omega_i\right] \\
&= E\left[(1 - (1 - F(\min(V_i, Y_i)|\Omega_i))^{n_i}) | \Omega_i\right] \\
&= \int (1 - (1 - F(\min(v, Y_i)|\Omega_i))^{n_i}) \, g(v|\Omega_i) \, dv.
\end{aligned}
$$

In the "sell" case, we get similarly

$$
\begin{aligned}
\mathcal{L}_{i|\text{sell}}^{(2)} &= P\left(\max_{k=1,\dots,n_i} W_{k,i} \geq \max(V_i, Y_i) | \Omega_i\right) \\
&= E\left[\left(1 - E\left[\mathbf{1}_{\max_{k=1,\dots,n_i} W_{k,i} \leq \max(V_i,Y_i)} | V_i, \Omega_i\right]\right) | \Omega_i\right] \\
&= E\left[(1 - F^*(\max(V_i, Y_i)|\Omega_i)^{n_i}) | \Omega_i\right] \\
&= \int (1 - F^*(\max(v, Y_i)|\Omega_i)^{n_i}) \, g^*(v|\Omega_i) \, dv.
\end{aligned}
$$

c. If $I_i = 3$ (Not Traded), then the likelihood of X_i in the case of a "buy" order is

$$\mathcal{L}_{i|\text{buy}}^{(3)} = P\left(\min_{k=1,\ldots,n_i} W_{k,i} \geq V_i, Y_i \geq V_i | \Omega_i\right) = \int \mathbf{1}_{Y_i \geq v}(1 - F(v | \Omega_i))^{n_i} g(v | \Omega_i) \, dv.$$

For a "sell" order, it is

$$\mathcal{L}_{i|\text{sell}}^{(3)} = P\left(\max_{k=1,\ldots,n_i} W_{k,i} \leq V_i, Y_i \leq V_i | \Omega_i\right) = \int \mathbf{1}_{Y_i \leq v} F^*(v | \Omega_i)^{n_i} g^*(v | \Omega_i) \, dv.$$

If we want to take into account the additional information ("Tied", "Covered", "Tied Covered", "Other"), the part of the likelihood that is related to "Traded Away" RFQs becomes more complicated.

This part of the total likelihood is now

$$\mathcal{L}^{(2)} := \prod_{i=1, I_i=2}^{N} \mathcal{L}_i^{(I_i, J_i)} = \prod_{i=1, I_i=2}^{N} \sum_{l=1}^{4} \mathbf{1}_{J_i=l} \mathcal{L}_i^{(2,l)},$$

To write the different terms, let us consider the dealers' quotes $(W_{k,i})_k$ corresponding to the RFQ i in ascending order:[33] $W_{(1)} \leq W_{(2)} \leq \cdots \leq W_{(n_i)}$.

Let us start with "buy" RFQs. We recall that the joint probability density function of $(W_{(1)}, W_{(2)})$ is

$$f_{(1),(2)}(w_{(1)}, w_{(2)} | \Omega_i) =$$

$$\frac{n_i(n_i - 1)}{2} \left(1 - F(w_{(2)} | \Omega_i)\right)^{n_i-2} f(w_{(1)} | \Omega_i) \, f(w_{(2)} | \Omega_i) \, \mathbf{1}_{w_{(1)} \leq w_{(2)}}.$$

We can write:

- In the "Tied" case:

$$\begin{aligned}
\mathcal{L}_{i|\text{buy}}^{(2,1)} &= P\left(W_{(1)} = Y_i \leq V_i | \Omega_i\right) \\
&= P\left(W_{(1)} = Y_i | \Omega_i\right) P\left(Y_i \leq V_i | \Omega_i\right) \\
&= n_i(1 - F(Y_i | \Omega_i))^{n_i-1} f(Y_i | \Omega_i) \, (1 - G(Y_i | \Omega_i)).
\end{aligned}$$

- In the "Covered" case:

$$\begin{aligned}
\mathcal{L}_{i|\text{buy}}^{(2,2)} &= P\left(W_{(1)} < Y_i < W_{(2)}, W_{(1)} \leq V_i | \Omega_i\right) \\
&= E\left[\mathbf{1}_{W_{(1)} < Y_i < W_{(2)}}(1 - G(W_{(1)} | \Omega_i)) | \Omega_i\right] \\
&= \int \mathbf{1}_{w_{(1)} < Y_i < w_{(2)}}(1 - G(w_{(1)} | \Omega_i)) f_{(1),(2)}(w_{(1)}, w_{(2)} | \Omega_i) \, dw_{(1)} dw_{(2)},
\end{aligned}$$

if $n_i > 1$. When $n_i = 1$, we have:

$$\begin{aligned}
\mathcal{L}_{i|\text{buy}}^{(2,2)} &= P\left(W_{(1)} < Y_i, W_{(1)} \leq V_i | \Omega_i\right) \\
&= \int \mathbf{1}_{w_{(1)} < Y_i}(1 - G(w_{(1)} | \Omega_i)) f(w_{(1)} | \Omega_i) \, dw_{(1)},
\end{aligned}$$

---

[33]We skip here the index i of the RFQ.

36

- In the "Tied Covered" case:

$$\mathcal{L}_{i|\text{buy}}^{(2,3)} = P\left(W_{(1)} < Y_i = W_{(2)}, W_{(1)} \le V_i | \Omega_i\right)$$
$$= E\left[\mathbf{1}_{W_{(1)} < Y_i = W_{(2)}}(1 - G(W_{(1)}|\Omega_i))|\Omega_i\right]$$
$$= \int \mathbf{1}_{w_{(1)} < Y_i}(1 - G(w_{(1)}|\Omega_i)) f_{(1),(2)}(w_{(1)}, Y_i | \Omega_i)\, dw_{(1)}.$$

- In the "Other" case:

$$\mathcal{L}_{i|\text{buy}}^{(2,4)} = P\left(W_{(2)} < Y_i, W_{(1)} \le V_i | \Omega_i\right)$$
$$= E\left[\mathbf{1}_{W_{(2)} < Y_i}(1 - G(W_{(1)}|\Omega_i))|\Omega_i\right]$$
$$= \int \mathbf{1}_{w_{(2)} < Y_i}(1 - G(w_{(1)}|\Omega_i)) f_{(1),(2)}(w_{(1)}, w_{(2)} | \Omega_i)\, dw_{(1)} dw_{(2)}.$$

For dealing with "sell" orders, we recall the joint density of $(W_{(n_i)}, W_{(n_i-1)})$:

$$f_{(n_i),(n_i-1)}^*(w_{(n_i)}, w_{(n_i-1)} | \Omega_i) =$$

$$\frac{n_i(n_i - 1)}{2} F^*(w_{(n_i-1)} | \Omega_i)^{n_i-2} f^*(w_{(n_i-1)} | \Omega_i)\, f^*(w_{(n_i)} | \Omega_i)\, \mathbf{1}_{w_{(n_i-1)} \le w_{(n_i)}}.$$

We can write:

- In the "Tied" case:

$$\mathcal{L}_{i|\text{sell}}^{(2,1)} = P\left(W_{(n_i)} = Y_i \ge V_i | \Omega_i\right)$$
$$= P\left(W_{(n_i)} = Y_i | \Omega_i\right) P\left(Y_i \ge V_i | \Omega_i\right)$$
$$= n_i F^*(Y_i|\Omega_i)^{n_i-1} f^*(Y_i|\Omega_i) G^*(Y_i|\Omega_i).$$

- In the "Covered" case:

$$\mathcal{L}_{i|\text{sell}}^{(2,2)} = P\left(W_{(n_i)} > Y_i > W_{(n_i-1)}, W_{(n_i)} \ge V_i | \Omega_i\right)$$
$$= E\left[\mathbf{1}_{W_{(n_i)} > Y_i > W_{(n_i-1)}} G^*(W_{(n_i)}|\Omega_i)|\Omega_i\right]$$
$$= \int \mathbf{1}_{w_{(n_i)} > Y_i > w_{(n_i-1)}} G^*(w_{(n_i)}|\Omega_i) f_{(n_i),(n_i-1)}^*(w_{(n_i)}, w_{(n_i-1)} | \Omega_i)\, dw_{(n_i)} dw_{(n_i-1)}.$$

if $n_i > 1$. When $n_i = 1$, we have

$$\mathcal{L}_{i|\text{sell}}^{(2,2)} = P\left(W_{(n_i)} > Y_i, W_{(n_i)} \ge V_i | \Omega_i\right)$$
$$= \int \mathbf{1}_{w_{(n_i)} > Y_i} G^*(w_{(n_i)}|\Omega_i) f^*(w_{(n_i)} | \Omega_i)\, dw_{(n_i)}.$$

- In the "Tied Covered" case:

$$\mathcal{L}_{i|\text{sell}}^{(2,3)} = P\left(W_{(n_i)} > Y_i = W_{(n_i-1)}, W_{(n_i)} \ge V_i | \Omega_i\right)$$
$$= E\left[\mathbf{1}_{W_{(n_i)} > Y_i = W_{(n_i-1)}} G^*(W_{(n_i)}|\Omega_i)|\Omega_i\right]$$
$$= \int \mathbf{1}_{w_{(n_i)} > Y_i} G^*(w_{(n_i)}|\Omega_i) f_{(n_i),(n_i-1)}^*(w_{(n_i)}, Y_i | \Omega_i)\, dw_{(n_i)}.$$

- In the "Other" case:

$$\mathcal{L}_{i|\text{sell}}^{(2,4)} = P\left(W_{(n_i-1)} > Y_i, W_{(n_i)} \geq V_i | \Omega_i\right)$$

$$= E\left[\mathbf{1}_{W_{(n_i-1)}>Y_i} G^*(W_{(n_i)} | \Omega_i) | \Omega_i\right]$$

$$= \int \mathbf{1}_{w_{(n_i-1)}>Y_i} G^*(w_{(n_i)} | \Omega_i) f^*_{(n_i),(n_i-1)}(w_{(n_i)}, w_{(n_i-1)} | \Omega_i) \, dw_{(n_i)} dw_{(n_i-1)}.$$

B Markov chain Monte Carlo and Gibbs sampling

Gibbs sampling is a Markov chain Monte Carlo method. It is used to sample from a joint posterior distribution of the form $\pi(\tilde{\alpha}, \tilde{\beta}|y)$ when one has an algorithm for sampling from the conditional distributions $\pi(\tilde{\alpha}|\tilde{\beta}, y)$ and $\pi(\tilde{\beta}|\tilde{\alpha}, y)$. The aim is to construct a Markov chain $(\tilde{\alpha}^t, \tilde{\beta}^t)$ whose stationary distribution is $\pi(\tilde{\alpha}, \tilde{\beta}|y)$. If we run the Markov chain for long enough, then the values of the chain will eventually follow the posterior distribution.

Algorithm 3 describes the basic Gibbs sampler (with two parameters).

Algorithm 3 Basic Gibbs sampler

Initialize (*e.g.* at random) $(\tilde{\alpha}^0, \tilde{\beta}^0)$
for iteration t=1 to T **do**
    Simulate $\tilde{\alpha}^t$ from the conditional distribution $\pi(\tilde{\alpha}|\tilde{\beta}^{t-1}, y)$
    Simulate $\tilde{\beta}^t$ from the conditional distribution $\pi(\tilde{\beta}|\tilde{\alpha}^t, y)$
end for

Convergence of the method is true under very general conditions. In practice, it is necessary to check that the algorithm has converged. There are numerous procedures for doing so, which are out of scope here – the interested reader is referred to [14]. The run of the Markov chain can then be split into two parts: the initial part of the run, until it reaches stationarity, is called the *burn-in* and is usually discarded. The remainder of the run corresponds to the chain exploring the posterior distribution, and is used for inference, possibly after sub-sampling to reduce the correlation between iterations.

A complete implementation of the MCMC procedure in Python and Cython with its documentation is available from our website at `http://squad.compmath.fr/doc/`

The conditional distributions are derived from the complete likelihood:[34]

$$\pi(\mathbf{I}, \boldsymbol{W}, \mathbf{V}, \tilde{\alpha}, \tilde{\beta}|\mathbf{X}, \boldsymbol{\Omega}) = \prod_i \prod_k g(V_i|\tilde{\alpha}, \Omega_i) f(W_{k,i}|\tilde{\beta}, \Omega_i) \pi(I_i|\boldsymbol{W}_i, V_i, X_i) \cdot \pi(\tilde{\alpha}, \tilde{\beta}).$$

The "complete conditional" distributions for the latent variables depend only on per-quote information:

$$\pi(W_{k,i}|I_i, \boldsymbol{W}_{l \neq k,i}, V_i, \tilde{\alpha}, \tilde{\beta}, X_i, \Omega_i) \propto f(W_{k,i}|\tilde{\beta}, \Omega_i) \pi(I_i|\boldsymbol{W}_i, V_i, X_i), \quad \text{(B.1)}$$

$$\pi(V_i|I_i, \boldsymbol{W}_i, \tilde{\alpha}, \tilde{\beta}, X_i, \Omega_i) \propto g(V_i|\tilde{\alpha}, \Omega_i) \pi(I_i|\boldsymbol{W}_i, V_i, X_i). \quad \text{(B.2)}$$

As for the parameters, we have

$$\pi(\tilde{\beta}|\boldsymbol{I}, \boldsymbol{W}, \boldsymbol{V}, \tilde{\alpha}, \boldsymbol{X}, \boldsymbol{\Omega}) \propto \pi(\tilde{\beta}) \prod_i \left(\prod_k f(W_{k,i}|\tilde{\beta}, \Omega_i) \right), \quad \text{(B.3)}$$

---
[34]We only consider here the case of "buy" RFQs.

$$\pi(\tilde{\alpha}|\boldsymbol{I},\boldsymbol{W},\boldsymbol{V},\tilde{\beta},\boldsymbol{X},\boldsymbol{\Omega}) \propto \pi(\tilde{\alpha})\prod_i g(V_i|\tilde{\alpha},\Omega_i). \tag{B.4}$$

The latter distributions require the total information about all quotes, and sampling from them is thus more demanding computationally.

Sampling latent variables

In our model, G_0 is a normal distribution parameterized by $\tilde{\alpha} = (\nu_{\tilde{\alpha}}, \tau_{\tilde{\alpha}})$. Sampling the latent variables V_i amounts to sampling from the truncated Gaussian distributions:

$$\pi\left(\frac{V_i - CBBT_i}{\Delta_i}|I_i,\boldsymbol{W}_i,\tilde{\alpha},\tilde{\beta},X_i,\Omega_i\right) = \mathcal{N}(\nu_{\tilde{\alpha}},\tau_{\tilde{\alpha}}^2)\mathbf{1}_{I_i|\boldsymbol{W}_i,V_i,X_i}.$$

F_0 is a SEP distribution with parameters $\tilde{\beta} = (\mu_{\tilde{\beta}}, \sigma_{\tilde{\beta}}, \alpha_{\tilde{\beta}}, \lambda_{\tilde{\beta}})$. Sampling the latent variables $W_{k,i}$ amounts to sampling from the truncated SEP distributions:

$$\pi\left(\frac{W_{k,i} - CBBT_i}{\Delta_i}|I_i,\boldsymbol{W}_{l\neq k,i},V_i,\tilde{\alpha},\tilde{\beta},X_i,\Omega_i\right) = SEP(\mu_{\tilde{\beta}},\sigma_{\tilde{\beta}},\alpha_{\tilde{\beta}},\lambda_{\tilde{\beta}})\mathbf{1}_{I_i|\boldsymbol{W}_i,V_i,X_i}.$$

Both $W_{k,i}$ and V_i can be sampled by a simple rejection algorithm. We have verified empirically that the rejection rates remain acceptable with our dataset under this basic sampling technique.

Sampling parameters

Using conjugate priors for $\nu_{\tilde{\alpha}}$ and $\tau_{\tilde{\alpha}}$ leads to a simple form for the posteriors.

As far as the mean is concerned, we have used a Gaussian prior $\pi(\nu_{\tilde{\alpha}}) = \mathcal{N}(\nu_0, \epsilon_0^2)$. This implies

$$\pi(\nu_{\tilde{\alpha}}|\boldsymbol{I},\boldsymbol{W},\boldsymbol{V},\tau_{\tilde{\alpha}},\boldsymbol{X},\boldsymbol{\Omega}) = \mathcal{N}(\bar{\nu},\bar{\epsilon}^2),$$

where

$$\bar{\nu} = \left(\frac{\nu_0}{\epsilon_0^2} + \frac{1}{\tau_{\tilde{\alpha}}^2}\sum_{i=1}^{N}\frac{V_i - CBBT_i}{\Delta_i}\right)\Big/\left(\frac{1}{\epsilon_0^2} + \frac{N}{\tau_{\tilde{\alpha}}^2}\right)$$

and

$$\frac{1}{\bar{\epsilon}^2} = \left(\frac{1}{\epsilon_0^2} + \frac{N}{\tau_{\tilde{\alpha}}^2}\right).$$

In our experiments, we used $\nu_0 = 0$ and $\epsilon_0 = 10$.

Regarding $\tau_{\tilde{\alpha}}$, the conjugate prior is an Inverse Gamma distribution: $\pi(\tau_{\tilde{\alpha}}^2) = \mathcal{IG}(a,b)$.

With this prior, we have:

$$\pi(\tau_{\tilde{\alpha}}^2|\boldsymbol{I},\boldsymbol{W},\boldsymbol{V},\nu_{\tilde{\alpha}},\boldsymbol{X},\boldsymbol{\Omega}) = \mathcal{IG}\left(\bar{a},\bar{b}\right),$$

where $\bar{a} = a + \frac{N}{2}$ and $\bar{b} = b + \frac{1}{2}\sum_{i=1}^{N}\left(\frac{V_i - CBBT_i}{\Delta_i} - \nu_{\tilde{\alpha}}\right)^2$.

We used $a = 3$ and $b = 1$ in the experiments presented in this paper.

Sampling the parameters $\tilde{\beta} = (\mu_{\tilde{\beta}}, \sigma_{\tilde{\beta}}, \alpha_{\tilde{\beta}}, \lambda_{\tilde{\beta}})$ of the SEP is a bit more involved. To palliate the absence of analytical solution, we resorted to using a Metropolis algorithm to sample from the posterior. Using a Metropolis step inside a Gibbs sampler is known as the Metropolis-within-Gibbs algorithm, and retains the convergence properties of Gibbs sampling alone.

Algorithm 2 in Section 3.3 gives a procedural description of the Metropolis-within-Gibbs algorithm. The distribution

$$\pi(\tilde{\beta}|\tilde{\alpha}^t, \boldsymbol{X}, \mathbf{V}, \mathbf{W}, \boldsymbol{\Omega}) = \frac{\pi(\mathbf{V}, \mathbf{W}|\tilde{\beta}, \tilde{\alpha}^t, \mathbf{X}, \boldsymbol{\Omega})\pi(\tilde{\beta})}{\pi(\mathbf{V}, \mathbf{W}|\tilde{\alpha}^t, \mathbf{X}, \boldsymbol{\Omega})}$$

is computed from the likelihood of the data and a prior distribution[35] for $\tilde{\beta} = (\mu_{\tilde{\beta}}, \sigma_{\tilde{\beta}}, \alpha_{\tilde{\beta}}, \lambda_{\tilde{\beta}})$. The likelihood term $\pi(\mathbf{V}, \mathbf{W}|\tilde{\beta}, \tilde{\alpha}^t, \mathbf{X}, \boldsymbol{\Omega})$ was explicited in (B.1) and (B.2). As for the priors, we initially left $\tilde{\beta}$ unconstrained using an improper uniform prior on $\mathbb{R} \times \mathbb{R} \times \mathbb{R}^+ \times \mathbb{R}$. Then, we chose to add the constraints $\mu_{\tilde{\beta}} \leq 0$ and $\lambda_{\tilde{\beta}} > 0$ (resp. $\mu_{\tilde{\beta}} > 0$ and $\lambda_{\tilde{\beta}} < 0$) for the experiments on the "buy" (resp. "sell") RFQs. This induced more robust inference procedures and prevented the algorithm from being trapped into a local minimum of the likelihood.

Sampling covariate coefficients

Finally, the coefficients of the linear regressions, *i.e.* (b_D, c_D, b_C, c_C) for the "buy" RFQs, and $(b_D^*, c_D^*, b_C^*, c_C^*)$ for the "sell" RFQs (as described in Equations (2.3) and (2.4)) must be estimated using a Metropolis-within-Gibbs algorithm as well. We estimated each group of coefficients from its own complete conditional distribution (which is equivalent to the procedure that was used for sampling parameters $\tilde{\beta}$), and used an (improper) uniform prior for all the experiments.

---

[35] $\pi(\tilde{\alpha}^t)$ appears both on the numerator and denominator, and need not be computed.