ISSN 2279-9362

Collegio Carlo Alberto

Equilibrium bid-ask spreads and the effect of competitive trading delays

Elisa Luciano
Antonella Tolomeo

No. 467
November 2016

Carlo Alberto Notebooks
www.carloalberto.org/research/working-papers

Equilibrium bid-ask spreads

and the effect of competitive trading delays

Elisa Luciano,[*] Antonella Tolomeo[†]

November 9, 2016

Abstract

The paper studies the equilibrium bid-ask spread and time-to-trade in a continuous-time, intermediated financial market. Market makers are monopolists, but random switches to a disintermediated market - interpreted as an option to wait and trade without bid-ask spread - occur. In equilibrium, this lowers spreads dramatically and generates higher fees for smaller investors, especially if their risk aversion is close to the market maker's one. Capital constraints on intermediaries easily lead to a second best. We analyze the effects of policy interventions in favour of disintermediation and deterring speculation on the part of market makers. We conclude for a positive effect of the Volcker rule on costs and liquidity.

Keywords: equilibrium with transaction costs, equilibrium with intermediaries, infrequent trading, trading delays, endogenous bid-ask spread, OTC markets.

JEL classification numbers: G12,G11

[*]Università di Torino and Collegio Carlo Alberto, Italy. Email: elisa.luciano@unito.it. Elisa Luciano thanks Suleyman Basak, Ricardo Caballero, Bernard Dumas, Alessandro Gavazza, Roberto Marfè, Marco Pagano, Larry Samuelson, Grigory Vilkov, participants to the International Capital Markets Conference at INSEAD, June 2016 and two anonymous referees for comments and suggestions on a companion paper. We are indebted with Riccardo Giacomelli for computational assistance. Any errors are ours.

[†]Università di Torino, Italy. Email:antonella.tolomeo@carloalberto.org

1 Introduction

Centralized trading through an intermediary or market maker, who does not simply match orders from buyers and sellers, but buys and sells assets on his own, providing the immediacy that decentralized trade and its search cost do not offer, is empirically very relevant. As Duffie (2012) reports, "the vast majority of transactions in OTC markets are with a market maker. The OTC market essentially covers all trade in bonds (corporate, municipal, US government, and foreign sovereign bonds), loans, mortage-related securities, currencies, and commodities, and about 60% of the outstanding notional amount of derivatives".

Since the market maker stands ready to absorb any order from the rest of the market, he will charge a bid-ask spread (a fee) for that. Faced with that spread, investors will trade infrequently. So*, an equilibrium model of intermediated market must explain both the endogenous level of spreads and the infrequency of intervention of investors*, as opposed to the standard continuous trading of frictionless models. That is the purpose of the present paper. We consider the case of a monopolistic market maker (a specialist), who can face the "competition" of a disintermediated market. Investors have the option to wait and trade in a decentralized marlet, where no bid-ask spread applies, instead of paying the specialist for the service of immediacy. The waiting time is random and models the trading delay that usually occurs when investors submit limit orders instead of using the intermediary service. We believe our model is important, not only because it is the first non-search-based model to provide the equilibrium, endogenous level of spreads, when all agents act rationally under symmetric information, but for two additional reasons. First, it is consistent with empirical observations on the tapering effect of the "competition" of disintermetiated trade on fees, and on the higher fees applied to small versus big investors. Second, it has neat policy implications. We indeed discuss recent policy interventions, in the spirit of the Volcker rule, meant to favour disintermediation, and we claim that they decrease fees.

The paper is expected to enhance comprehension of the price for intermediation and of the trade impact of strategic behavior by market makers. It aims at doing so with respect to the partial equilibrium models of investors' behavior in the presence of bid-ask spreads, such as Constantinides (1986) - which take those *spreads as exogenous* - and with respect to the traditional microstructure literature - such as Stoll (1978), Ho and Stoll (1981) - which takes the frequency of *arrival of customers as exogenous, and independent of the spread*. By creating an equilibrium model, we endogenize both spreads, or immediacy price, and trade.

Other recent models of centralized trading that provide endogenous bid-ask spreads explain it through search models and bargaining power (see Duffie, Gârleanu and Pedersen (2005, 2007)). As it will be the case for us, they assume that investors have the possibility of trading both in a decentralized and in a centralized way. We predict a spread behavior similar to Duffie *et al.* (2005) both when investors may more easily find other investors to trade with, consistently with empirical evidence, and when they have easier access to the market maker. As in Duffie *et al.* (2005), we interpret easiness of access as a difference between large and small investors and show that, in contrast with traditional models of market making under asymmetric information, we can explain why smaller spreads are offered by market makers to big investors. In comparison with Duffie *et al.*, we can also discuss the effect of risk aversion and the trade features (frequency and amount of trade). [1]

[1] Related literature also includes *decentralized* models of trade with symmetric information in general equilibrium, both when investors have the same risk aversion (Vayanos (1998), Lo, Mamaysky and Wang (2004)), and when they do not (Buss and Dumas (2013)). Since trading is competitive, agents simply share *exogenous trading costs*. Only the sharing rule is endogenous, while in our study the amount of the spread, that may include physical costs paid by the intermediary, is.

As in He and Krishnamurthy (2013), we develop an equilibrium model in which intermediaries are not a "veil", may be equity constrained, and we are able to characterize the cases in which a capital (or regulatory) constraint on intermediaries has a welfare effect on both them and investors.

In order to study equilibrium bid-ask spreads we go back to the simplest framework for investors' choices in continuous-time stochastic economies, characterized by a risky and a riskless asset, together with infinitely lived, power utility agents. We assume that a representative investor faces a single market maker, who sets the spreads and has no "outside option" to wait and trade without paying fees. We extend to the case in which investors have the option either to trade with the market maker at his bid-ask fee or to wait until another investor, with whom they can trade at no cost, submits an order. This reduces dramatically the bid-ask spread and increases trade. To model the random time needed to trade with another investor, we use a random switch to a disintermediated market.

The outline of the paper is as follows. Section 2 sets up the model and studies the optimization conditions for the two types of agents (investors and market maker) when the former cannot trade outside the intermediated market. Section 3 defines the corresponding equilibrium and studies its features. It also provides numerical examples, to study spreads, trading policy, transaction frequency in comparison with partial equilibrium models. Section 4 sets and solves the outside-option case, while Section 5 studies its implications for equilibrium spreads, trade, welfare and policy interventions. Section 6 summarizes and concludes. The Appendices contain proofs.

2 Model set up, no outside option

This section specifies the objective of the agents (investors and market maker) and solves their optimization problems, when there is no outside option. We consider the stationary equilibrium of a continuous-time stochastic economy in which two assets are traded: a riskless and a risky one. The interest rate r on the riskless asset is not determined endogenously, since the model features no consumption. The pre-bid, pre-ask price of the risky asset - the *fundamental value* of capital in the economy, as we explain below - is a geometric Brownian motion with parameters α and σ. Two agents populate our economy: a representative investor and a monopolistic market maker. Assuming that the intermediary selects the bid and ask fees means to assume that he determines the whole trade price processes.

The *investor* maximizes the expected utility of his terminal wealth, $\mathbb{E}\left[U(W(T))\right]$. He has an infinite-horizon power utility, $U(W) = W^{\gamma}/\gamma$. We assume that he is risk averse and non-myopic: $\gamma < 1, \gamma \neq 0$. His objective is

$$\lim_{T\to\infty} \sup \mathbb{E}\left[U(W(T))\right] = \lim_{T\to\infty} \sup \mathbb{E}\left[W(T)^{\gamma}/\gamma\right] \tag{1}$$

The *admissible, stationary transaction costs* are proportional to the value of trade. For each dollar of risky security he trades, the investor receives a bid price s and pays an ask price $1/q$, which will be constrained to be respectively smaller and greater than or at most equal to one: $s, q \in (0, 1]$. The differences $1 - s, 1/q - 1$ are the *transaction costs* for the investor. The difference $1/q - s$ is the *bid-ask spread*. The constants s and q will be determined in equilibrium and will therefore depend on the exogenous parameters. Since we search for a *stationary* equilibrium, s and q will be constant over time. The investor takes them as given. Let $x(t)$ and $y(t)$ be the fundamental values of his riskless and risky position.[2] His final wealth is their liquidation value, i.e. $W(T) = x(T) + sy(T)$.

The *market maker* issues both the riskless security, which is in zero-net supply, and provides access to the risky security, or *capital*, which is in infinite supply and whose value process is exogenous. He pays to the investor the returns on the risky asset and stands ready to absorb all the transactions required by the investor.[3] He charges a bid and an ask price for this, i.e. he sets s and q. Let x_s and y_s be the market maker's asset holdings: $x_s = -x$, $y_s = -y$. The ratio $\theta_s = y_s/x_s$ is therefore the same as the consumer one, namely $\theta_s = \theta$. The market maker is a power-utility agent which aims at maximizing the expected utility U_s of his final wealth, when the horizon becomes infinite:

$$\lim_{T\to\infty} \sup \mathbb{E}\left[U_s(W(T))\right] = \lim_{T\to\infty} \sup \mathbb{E}\left[W(T)^{\gamma'}/\gamma'\right] \tag{2}$$

We assume that he is risk-averse, $1-\gamma' > 0$. The dynamics of his assets is

$$\begin{cases} dx_s = rx_s dt - sdU + dL \\ dy_s = \alpha y_s dt + \sigma y_s dz - qdL + dU \end{cases} \tag{3}$$

while his final wealth is $W_s = x_s + y_s = -x - y$.

2.1 Optimization for the investor

The (partial equilibrium) investor's optimization problem has been solved by Dumas and Luciano (1991) for the case of non-infinitesimal spreads and by Gerhold, Guasoni, Muhle-Karbe and Schachermayer (2011) for the case of infinitesimal spreads. It is shown in both papers that - if z is the standard Brown-

[2] Later on y will be called also the pre-spread price: indeed, we do not need to distinguish prices and values.

[3] If capital were in inelastic supply, we should solve an intermediary's problem up to the first stopping time, in order to incorporate availability to trade up to the first intervention.

ian motion which drives y - there exist two increasing processes L and U which make the value of the investor's assets evolve according to

$$\begin{cases} dx(t) = rx(t)dt + sdU(t) - dL(t) \\ dy(t) = \alpha y(t)dt + \sigma y(t)dz(t) + qdL(t) - dU(t) \end{cases} \tag{4}$$

The processes L and U increase only when $\theta \doteq y/x$, the ratio of risky to riskless asset in portfolio, reaches respectively a lower and an upper barrier, which we denote as l and $u, l \leq u$. In what follows, for the sake of simplicity, we restrict our formulas to parameter combinations which make both barriers positive, i.e. $0 < l \leq u$. To this end, we restrict the parameters so that the optimal asset holdings would be positive in the absence of transaction costs by positing:

$$0 < \frac{\alpha - r}{(1-\gamma)\sigma^2} < 1 \tag{5}$$

We indeed know that asset holdings with bid-ask spreads include the optimal holdings in the corresponding frictionless market, $l < \theta^* < u$, and the optimal ratio θ^* is the standard Merton's one:

$$\left(\frac{y}{x}\right)^* = \theta^* = \frac{\alpha - r}{(1-\gamma)\sigma^2 - \alpha + r} \tag{6}$$

Condition (5) makes $\theta^* > 0$. With positive risk aversion, problem (1) under (4) reduces to solving for the function I the ODE

$$\underbrace{(r\gamma - \beta)}_{\doteq \delta} I(\theta) + (\alpha - r) I'(\theta)\,\theta + \sigma^2 I''(\theta)\,\frac{\theta^2}{2} = 0 \tag{7}$$

- with $\beta \in \mathbb{R}$, to be specified below - under the value-matching and smooth-pasting BCs, namely

$$\begin{cases} lI'(l) = \gamma I(l)\varepsilon_l \\ uI'(u) = \gamma I(u)\varepsilon_u \\ lI''(l) = (\gamma - 1)\, I'(l)\varepsilon_l \\ uI''(u) = (\gamma - 1)\, I'(u)\varepsilon_u \end{cases} \tag{8}$$

where we have used the shortcut notation

$$\varepsilon_l \doteq \frac{l}{l+q} \tag{9}$$

$$\varepsilon_u \doteq \frac{us}{1+us} \tag{10}$$

The function I provides us with the value function K of the problem,

$$\lim_{T\to\infty} K(x,y,t;T) \doteq \lim_{T\to\infty} \sup \mathbb{E}\left[W(T)^\gamma/\gamma\right]$$

if there exists a constant β - an artificial discount rate - which makes K itself, once discounted, finite and stationary. The discount rate β is defined by the requirement that, once K is discounted, the resulting value function

$$J(x,y,t;T) = e^{-\beta(T-t)}K(x,y,t;T)$$

is stationary:

$$\lim_{T\to\infty} J(x,y,t;T) = J(x,y)$$

and would be

$$r\gamma - \frac{(\alpha - r)^2}{2\sigma^2}\frac{\gamma}{1-\gamma} \tag{11}$$

in the absence of bid-ask spread. Given that the utility function is power, we can safely assume that $J(x,y) = x^\gamma I(\theta)$ and solve for I. The I function that solves (7) is

$$I(\theta) = \begin{cases} \theta^{-\mu}\left[Asi(\nu \ln(\theta)) + Bco(\nu \ln(\theta))\right] \\ \mathcal{A}\theta^{x_1} + \mathcal{B}\theta^{x_2} \end{cases} \tag{12}$$

where $A, B, \mathcal{A}, \mathcal{B}$ are constant, $x_{1,2} = \mu \pm \nu$,

$$\mu \doteq (\alpha - r)/\sigma^2 - 1/2, \tag{13}$$

$$\nu \doteq \frac{\sqrt{\left|\left(\alpha - r - \sigma^2/2\right)^2 - 2\delta\sigma^2\right|}}{\sigma^2} \tag{14}$$

The first (second) expression in (12) applies when the discriminant to the algebraic equation which corresponds to (7), namely

$$\sigma^2 x^2/2 + \left(\alpha - r - \sigma^2/2\right)x + \delta = 0 \tag{15}$$

is negative (positive or null). This happens when $\delta > (\leq)\delta_c$,where

$$\delta_{c.} \doteq \frac{\left(\alpha - r - \sigma^2/2\right)^2}{2\sigma^2}, \tag{16}$$

The second expression is equivalent to writing the first with hyperbolic sines and cosines.

It has also been shown that a solution technique for the above problem consists of three steps. The steps - which are described in Appendix A - turn the investor's problem into a single algebraic equation in the unknown δ, that is

$$a(l, q)b(u, s) - c(u, s)d(l, q) = 0 \tag{17}$$

where the expressions for a, b, c, d are

$$\begin{aligned} a(l,q) &= (-\mu - \gamma\epsilon_l)\, si(\nu \ln(l)) + \nu co(\nu \ln(l)) \\ b(u,s) &= (-\mu - \gamma\epsilon_u)\, co(\nu \ln(u)) - (+)\nu si(\nu \ln(u)) \\ c(u,s) &= (-\mu - \gamma\epsilon_u)\, si(\nu \ln(u)) + \nu co(\nu \ln(u)) \\ d(l,q) &= (-\mu - \gamma\epsilon_l)\, co(\nu \ln(l)) - (+)\nu si(\nu \ln(l)) \end{aligned} \tag{18}$$

according to whether the si and co are trigonometric or hyperbolic sines and cosines (the signs into brackets being the hyperbolic ones).

In addition, the solutions for δ which are acceptable are the ones which make ε_l and ε_u real. For the case of negative (positive) γ, a straightforward computation shows that this is the case as long as $\delta \leq (\geq)\, \delta^*$, where

$$\delta^* \doteq \frac{\gamma(\alpha - r)^2}{2(\gamma - 1)\sigma^2}. \tag{19}$$

So, a solution to the investor's problem exists as long as $\delta \leq (\geq)\, \delta^*$ - depending on the sign of γ - and entails the first or second representation for I depending on whether $\delta > (\leq)\delta_c$.

2.2 Optimization for the market maker

The market maker's problem is novel in the literature. It is subject to the standard value-matching conditions, when the processes L and U are different from zero. His instruments are not the trading barriers l and u, but the trading price processes, which obtain by modifying to the fundamental value (geometric process) of y by s and q. The FOCs with respect to l and u which provide the smooth-pasting conditions for the investor must be substituted by optimality conditions with respect to s and q. It can be shown that the value function cannot - and need not - be maximized with respect to s, q on the whole domain, but at most for specific choices of θ.[4] The natural choices are $\theta = l$ and $\theta = u$, since trade occurs at those levels only. Using the traditional approach to smooth pasting (see for instance Peskir and Shyriaev, 2006), we set the derivatives of the value function equal to zero with respect to s and q respectively at $\theta = l$ and $\theta = u$. Let K^s be the market maker's value function, i.e.

$$\lim_{T\to\infty} K^s(x_s, y_s, t; T) = \lim_{T\to\infty} \sup \mathbb{E}\left[W_s(T)^{\gamma'}/\gamma'\right]$$

It is easy to show, as in the investor's case, that, if we aim at a *stationary* value function, we must discount K^s at a rate $\beta' \doteq r\gamma' - \delta'$. We can define the discounted value function

$$J^s(x_s, y_s, t; T) = e^{-\beta'(T-t)} K^s(x_s, y_s, t; T)$$

and assume that it has a stationary limit:

$$\lim_{T\to\infty} J^s(x_s, y_s, t; T) = J^s(x_s, y_s) = (-x)^{\gamma'} I^s(\theta).$$

We end up with the following differential equation for I^s :

$$(r\gamma' - \beta')\, I^s(\theta) + (\alpha - r) I^{s\prime}(\theta)\,\theta + \sigma^2 I^{s\prime\prime}(\theta)\,\frac{\theta^2}{2} = 0 \tag{20}$$

whose solution is of the type

$$I^s(\theta) = \begin{cases} \theta^{-\mu}\left[A' si(\nu' \ln(\theta)) + B' co(\nu' \ln(\theta))\right] \\ \mathcal{A}'\theta^{x_1'} + \mathcal{B}'\theta^{x_2'} \end{cases} \tag{21}$$

[4]The proof can be obtained from the Authors upon request.

with $x'_{1,2} = \mu \pm \nu'$

$$\nu' \doteq \frac{\sqrt{\left|(\alpha - r - \sigma^2/2)^2 - 2\delta'\sigma^2\right|}}{\sigma^2}$$

and trigonometric (hyperbolic) sines and cosines applying when $\delta' > (\leq)\delta_c$. The value-matching conditions impose continuity of the value function at the trading points. Indeed, the investor chooses a trading policy which requires his counterpart to trade so as to stay at the boundary of the trading region too. We have:

$$\begin{cases} lI^{s\prime}(l) = \gamma' I^s(l)\varepsilon_l \\ uI^{s\prime}(u) = \gamma' I^s(u)\varepsilon_u \end{cases} \tag{22}$$

where the ε are the ones defined above (and decided by the investor). As in the investor's case, these value matching conditions are consistent with a non-degenerate value function if the constant δ' satisfies

$$a'(l,q)b'(u,s) - c'(u,s)d'(l,q) = 0 \tag{23}$$

where

$$a'(l,q) = (-\mu - \gamma'\epsilon_l)\, si(\nu' \ln(l)) + \nu' co(\nu' \ln(l))$$

$$b'(u,s) = (-\mu - \gamma'\epsilon_u)\, co(\nu' \ln(u)) - (+)\nu' si(\nu' \ln(u))$$

$$c'(u,s) = (-\mu - \gamma'\epsilon_u)\, si(\nu' \ln(u)) + \nu' co(\nu' \ln(u))$$

$$d'(l,q) = (-\mu - \gamma'\epsilon_l)\, co(\nu' \ln(l)) - (+)\nu' si(\nu' \ln(l))$$

The optimality conditions of the market maker are obtained from (22), differentiating with respect to q and s, i.e. computing[5][6]

$$\begin{cases} \dfrac{d}{dq}\left[-\gamma' I_s\ (l) + (q+l)\, I_s\ '(l)\right] = 0 \\ \dfrac{d}{ds}\left[(1+us)\, I_s\ '(u) - \gamma' I_s(u)s\right] = 0 \end{cases} \tag{24}$$

which gives the "modified" smooth-pasting conditions

$$\begin{cases} \dfrac{\partial l}{\partial q}\left[(1-\gamma')\, I_s\ '(l) + (q+l)\, I_s\ ''(l)\right] + I_s\ '(l) = 0 \\ \dfrac{\partial u}{\partial s}\left[(1+us)\, I_s\ ''(u) + (1-\gamma')\, sI_s'(u)\right] - \gamma' I_s(u) + uI\prime_s(u) = 0 \end{cases} \tag{25}$$

In the last system we have the derivatives of the boundaries with respect to the costs, $\frac{\partial l}{\partial q}, \frac{\partial u}{\partial s}$, which must be obtained from the investor's problem solution, as in Appendix B, expressions (61) and (62). Substituting for (22) and (25) into the ODE, we get the following algebraic equations, which synthesize the value-matching and "modified" smooth-pasting conditions for the market maker:

$$\delta' + \varepsilon_l(\alpha - r)\gamma^{'} - \frac{\sigma^2}{2}\gamma' \varepsilon_l^2 \left[\frac{1}{\frac{\partial l}{\partial q}} + 1 - \gamma'\right] = 0 \tag{26}$$

$$\delta' + \varepsilon_u(\alpha - r)\gamma^{'} + \frac{\sigma^2}{2}\gamma^{'} \varepsilon_u^2 \left[\frac{u}{s}\frac{1-\varepsilon_u}{\varepsilon_u \frac{\partial u}{\partial s}} - 1 + \gamma^{'}\right] = 0 \tag{27}$$

where the derivatives $\frac{\partial l}{\partial q}, \frac{\partial u}{\partial s}$ are given by (61) and (62). The market maker's problem would be solved by I^s, for given δ, if δ', s, q solved (23), (26), (27), with

[5] In order to take the derivatives of the value function with respect to the specialist' s choice variables, recognize that the bid price s applies at the upper barrier u only, while the ask price $1/q$ applies at the lower barrier l only. As a consequence, the derivatives to be equated to zero are with respect to q at l and with respect to s at u. In taking these derivatives, he considers the investor's reaction to his choice of the spreads.

[6] It can be demonstrated that an equilibrium in which specialists do not take the reaction of their counterpart into consideration does not exist. The reaction is evaluated in terms of barriers, not in terms of traded quantities, since we know that the investor trades so as to stay along the barriers of the no-transaction cone. The only investor's reaction is in terms of the level, or barrier, not in terms of quantity of intervention, or amount of trade. The proof can be obtained from the Authors upon request.

$s, q \in (0,1]^2$.[7]

3 Equilibrium

This section defines an equilibrium for the no-outside-option economy, comments on the properties of its prices and quantities and provides a numerical example.

An *equilibrium* in the previous market is a quadruple (δ, δ', s, q), with $s, q \in (0,1]^2$, such that, assuming for simplicity that $\theta(0) = l$ or $\theta(0) = u$[8]

- the investor's maximization problem is solved
- the market maker's one is solved too
- and the barriers l and u are real: $\delta \leq (\geq)\delta^*$ if $\gamma < (>)0$.[9]

Since, by definition, the market maker absorbs any trading need of the investor, we do not need to worry about matching demand and supply of the risky and riskless asset. No further market clearing condition is needed, since $\theta = \theta_s$. Overall, an equilibrium requires that the four algebraic equations (17), (23), (26), (27) - which we report here for the sake of convenience - be solved at the same time with $s, q \in (0,1]^2, \delta \leq (\geq)\delta^*$ if $\gamma < (>)0$, and, with the expression for the derivatives given in Appendix, they are in quasi-closed form and unique up to the constants (δ, δ'). The barriers depend on the exogenous and endogenous

[7]The model can be extended to incorporate costs on the part of the intermediary (technology, execution etc.). It can be shown that these costs do not affect the qualitative results in equilbrium.

[8]We could also start from a position for the investor at time 0 different from the intervention one. In that case, as observed in Dumas and Luciano (1991), the conditions should be modified to take the initial adustment to the barrier into consideration.

[9]The value functions will have a representation in terms of trigonometric (hyperbolic) sines and cosines when $\delta, \delta' > (\leq)\delta_c$. As a whole, four combinations may occur.

parameters, so that we should write $l(\gamma, \gamma', \alpha, r, \sigma, s, q)$, and similarly for u.

$$\begin{cases} ab - cd = 0 \\ a'b' - c'd' = 0 \\ \delta' + \varepsilon_l(\alpha - r)\gamma' - \frac{\sigma^2}{2}\gamma'\varepsilon_l^2\left[\frac{1}{\frac{\partial l}{\partial q}} + 1 - \gamma'\right] = 0 \\ \delta' + \varepsilon_u(\alpha - r)\gamma' + \frac{\sigma^2}{2}\gamma'\varepsilon_u^2\left[\frac{u}{s}\frac{1-\varepsilon_u}{\varepsilon_u \frac{\partial u}{\partial s}} - 1 + \gamma'\right] = 0 \end{cases} \tag{28}$$

Equilibrium prices, quantities and trade are as follows.

The procedure we follow for *prices* consists in verifying that when the pre-bid, pre-ask geometric Brownian motion[10] price that obtains by setting $dl = dU = 0$ in the SDE for y in (4) is a *fundamental value,* the ensuing bid-ask spread $1/q - s$ is the equilibrium one.The fundamental value of capital is never observed as a trading price, while sy and y/q are. They can be observed only when trade occurs, though. There are two different trading prices. When trade occurs because the investor reaches his upper barrier, and needs to sell the risky asset, the cum-bid price sy is the *observed trading price*; when trade occurs at the lower investor's barrier, the cum-ask price y/q is the observed trading price. Both prices are reduced (substantially reduced, as we will see in numerical examples) because of transaction costs s and q. This is in the spirit of Amihud and Mendelson (1986).[11]

As for *quantities*, it is known that L and U are the local times of the process

[10] In the inventory-based microstructure literature there is a constant fundamental value of the asset, to which cum-spread prices tend to revert. This mean reversion does not exist in our model, since pre-spread prices are geometric Brownian motions, while spreads are constant and time-independent. This makes our model consistent with the lack of mean reversion on specialists' prices, as empirically detected, for instance, by Madhavan and Smidt (1991) in equity markets.

[11] A main difference between our model and traditional inventory ones is the lack of mean reversion in the inventory level of the specialist (l, u). Some inventory-based models do indeed determine a preferred inventory position, to which he aims at reverting. In our model the inventory, measured by the ratio $\theta_s = \theta$, fluctuates between l and u, and is kept within those barriers because of the optimal policies of investors. There is no optimal portfolio for the specialist itself. As a consequence, we do not have problems in matching the lack of empirical mean reversion in inventories.

θ at l, u respectively: trade per unit of time is infinitesimal, with infinite total and finite quadratic variation. In this sense, there is no "order size" in the traditional sense of the microstructure literature.[12]

The *trading policy* behind our equilibrium is such that observed trade is not continuous in time, but *infrequent. The frequency of trade will depend on the distance between the barriers* l and u. The closer the barriers, the more frequent trade will be. This is an empirical feature of markets with market-makers. Duffie (2012) and references therein report for instance that trades in individual corporate US bonds or CDS, which are already quite active markets, happen in a small number per day, even considering all market makers. Knowing that the portfolio ratio stays between the barriers and using the properties of local times of regulated Brownian motion, we will compute the trade frequency, namely the expected time to next trade.

It is clear from the equilibrium conditions that spreads and trade will depend on the *risk aversion* of market participants. It is quite intuitive that an equilibrium will exist if the market maker is less risk averse than the investor. Pagano and Roell (1989) already proved, although in a different context, that market makers trade only with customers more risk averse than themselves.[13] We investigate this circumstance, represented by the case $1-\gamma > 1-\gamma'$, or $\gamma' - \gamma > 0$, as well as the spread and trade dependence on the difference in risk aversion between market participants, in the next section. before doing that,

[12]The price and quantity features just listed are consistent with the findings in Buss and Dumas (2012) for a competitive market. The intuition is that their bid-ask spread is exogenous and time is discrete, but their endowment evolves as a binomial tree, and our risky asset's fundamental value evolves as a geometric Brownian motion. So, in both cases transaction prices and trades have infinite total and finite quadratic variation.

[13]In Pagano and Roell's set up, intermediaries set the bid-ask price competitively, by equating the utility they get with and without operating as intermediaries. Investors equate the utility they get when selling (buying) in an intermediated market with the one they get when selling (buying) in a competitive market, i.e. an auction or limit-order one. When customers are more risk averse than market makers, the possibility of trading depends also on the spread which would prevail on a competitive market and on the probability of finding a counterpart in it. When trade occurs in the intermediated market, the spread magnitude depends on the difference between the risk attitudes of participants, exactly as in our setting.

we compare with partial equilibrium.

3.1 Numerical illustration

The equilibrium conditions provided above cannot be solved explicitly. We discuss them starting from a base-case, which is calibrated to the pioneering literature in single investor's – or partial equilibrium – optimality with transaction costs (Constantinides (1986)). We first obtain the equilibrium quadruple in the base-case and discuss the resulting bid-ask spread, transaction policy, expected time to next trade and rate of growth of derived utility, in comparison with their partial equilibrium values. We then discuss the impact of the difference in risk aversions on the results.

A proviso is in order: there is no attempt to calibrate our spreads to a specific market. As a consequence, the results should be interpreted as those of Buss and Dumas (2013) or Lo *et al.* (2004): we care about the relative magnitude of partial versus general equilibrium results, as well as the behavior in terms of risk aversion, not about the absolute level of spreads.

3.2 Base case

Starting from the fundamental risk-return base-case in Constantinides (1986), i.e. $\alpha = 7\%, r = 2\%$ so that $\alpha - r = 5\%, \sigma^2 = 4\%$, we assume a coefficient of risk aversion for the investor equal to $1 - \gamma = 4$, which is within Constantinides' range, and a market maker's risk aversion equal to: $1 - \gamma' = 3.85$, so that $\gamma' - \gamma = 0.15$.

The investor, monopolistic-market-maker equilibrium is indeed characterized by the quadruple[14]

$$(\delta, \delta', s, q) = (1.999\%, 1.999\%, 84.2\%, 75.7\%),$$

with barriers equal to $l = 0.1809 < \theta^* < u = 0.9041$, since the corresponding no-cost problem has optimal portfolio mix $\theta^* = 0.4545$. Let us denote with an index p the corresponding partial-equilibrium solutions. Keeping costs at the level provided in general equilibrium, for the sake of comparison, barriers become equal to $l_p = 0.10 < \theta^* < u_p = 1.2279$, while the discount rate δ_p becomes 1.4%.

Let us comment on the equilibrium *bid/ask spread* first. The bid-ask spread - or round-trip cost - is as high as $1/q - s = 47.9\%$. With a unique market maker and no outside-option, even if the risk aversion of the two counterparts is quite close, we may have huge costs and a huge *spread* in equilibrium. The spread level in equilibrium may well be unrealistic. However, recall that, by using the parameters of the previous transaction-cost, partial-equilibrium literature, we simply aim at stressing how important a monopolistic position of the market maker can be in terms of spread (which was exogenous in partial equilibrium). In a later section we weaken the monopoly power by introducing outside options for investors. We still do not calibrate to a specific market, but maintain the interest in the sensitivity of spreads to external determinants in equilibrium.

Let us see the effects on the *no-trade region*. By keeping costs the same between the general and partial equilibrium, as we did above, we find that the intervention barriers are slightly further apart in the partial-equilibrium than in the equilibrium case: $l - l_p = 0.0809, u_p - u = 0.3238$. The no-transaction cone in partial equilibrium incorporates the general equilibrium one:

[14]For the given parametrization, $\delta_c = 0.01125$, $\delta^* = 0.0234375$. Since both δ and δ' are greater than δ_c, the roots of the algebraic equation corresponding to (7), which is equation (15) in Appendix A - and its equivalent for the intermediary - are imaginary. The transaction boundaries are real, since $\delta < \delta^*$.

[insert here Figure 1]

This means that partial-equilibrium models are likely to have slightly *overstated the magnitude of no trade*, even though they perfectly captured the trading mechanism. In a general-equilibrium perspective, the investor is less reluctant to trade, since the market maker has forecasted his customer's reaction when fixing the spreads.

Another feature is that the equilibrium magnitude of costs is not symmetric: ask spreads $1/q - 1 = 0.321$ are twice as big as bid ones $1 - s = 0.158$. In partial equilibrium models the exogenous spreads are usually taken to be equal, something which distorts the results even more than what Figure 1 says. In the current equilibrium, for instance, if the ask spreads were equal to the bid ones, the lower barrier would go up, as well as the upper barrier would go up if the bid were equated to the ask.

Trade is far from being continuous. The frequency of trade can be measured by the expected time that the process θ takes in order to reach either the upper barrier u or the lower one l, starting from the optimal mix θ^*. Between l and u, θ has drift $\alpha - r$ and diffusion σ. Standard results in the theory of the first passage time of a Brownian motion through either an upper or a lower boundary tell us that the expected time we are searching for is

$$t^* = \frac{1}{\frac{\sigma^2}{2} - (\alpha - r)} \left[\ln\left(\frac{\theta^*}{l}\right) - \frac{1 - \left(\frac{\theta^*}{l}\right)^{1 - \frac{2(\alpha - r)}{\sigma^2}}}{1 - \left(\frac{u}{l}\right)^{1 - \frac{2(\alpha - r)}{\sigma^2}}} \ln\left(\frac{u}{l}\right) \right] \tag{29}$$

In the general equilibrium just described, the expected time between transactions is very high, 13.4 years. It would be 26 years in the corresponding partial-equilibrium case. This happens because the barriers are distant from the Merton's line in both cases, but more in partial than in general equilibrium.

We still need to verify that, at least in the base-case, the *rate of growth of ex-*

pected utility moves in the right direction when going from a non-intermediated market to an intermediated one. To do so, let us comment on the last couple of equilibrium parameters, namely δ and β. Given that $\beta = r\gamma - \delta$, the higher is δ, the smaller the rate of growth of utility of the corresponding agent.[15] Since

$$\delta = 2\% > \delta_p = 1.4\%,$$

the investor's rate of growth of expected utility in the current equilibrium, β, is smaller than in the corresponding partial equilibrium ($\beta = -8\% < \beta_p = -7.42\%$).[16] The presence of a (monopolistic) market maker affects this rate in the expected direction.[17] Previous, partial equilibrium models with transaction costs then tended to underestimate the effects of intermediation on investors' utility.

3.3 Sensitivity analysis

This section explores the spread and trade implications of changing the participants' risk aversion, as done in Table 1.

[insert here Table 1]

By decreasing the risk aversion of the market maker, or making it more distant from the investor's one (from top to bottom of the table), we find equilibria characterized by lower s and q. When $1-\gamma' = 3$, so that $\gamma' - \gamma = 1$ - a case that we will explore also with outside option - the equilibrium bid prices decrease,

[15]Since utility stays bounded when discounted at the rate β, it grows at β when it is not artificially discounted.

[16]Please remember that the rate of growth can be negative even without transaction costs, since even in that case it depends on $\alpha, r, \sigma, \gamma$, according to (11). With the current paramaters, it would be -3.66%.

[17]Starting from this, we could determine the *liquidity discount* that investors would tolerate, costs being equal, in order to go from a general to a partial equilibrium or to a frictionless equilibrium. This means to determine under which $\alpha - r$ investors see their welfare growth unaffected by the strategic specialist's intervention. Practically, it means to solve for $\alpha - r$ the investor's problem with $\delta = \delta_p$ (provided that the solution still gives $\delta \leq (\geq)\delta^*$ if $\gamma < (>)0$).

the ask prices go up, the spread increases, the no-trade lower limit decreases, the upper goes up, so that the whole cone of no transaction opens up, following the behavior of the spread.

From top to bottom, even considering intermediate values of the risk aversion difference, the market maker decreases the bid and increases the ask price. So, the more distant agents are in risk aversion, the higher is the spread. In the microstructure, inventory-based models, such as Ho and Stoll (1981), risk aversion had an impact on prices together with trade size and inventories. In Ho and Stoll the bid-ask spread increases when the intermediary's risk aversion goes down, as from top to bottom in Table 1. This happens because the intermediary is more lenient towards the risky asset, and is counterpart arrives at the same (exogenous) rate, so that demand is inelastic with respect to his spread. Here the market maker takes also into account the elasticity of demand (the investor's behavior) when setting the prices, as commanded by his optimization conditions.

The bid and ask *prices* behavior is presented in Figure 2 below.

[insert here Figure 2]

Figure 3 shows the behavior of the barriers as a function of the difference between the market maker and investor's risk aversion.

[insert here Figure 3]

As a consequence of the spread and single prices behavior, the lower barrier decreases, the upper increases, and the cone, as clear from Table 1, opens up.

Assuming that barriers correspond to inventories in the microstructure literature, we observe that not only both the bid and the ask price separately depend on inventories, as was clear from the equilibrium definition, but also the bid-ask spread does, since an increase in the difference of risk aversions changes the inventories and opens up the cone. In traditional microstructure models, inventories disappear as determinants of the spread, while being determinants of its components (the bid and ask prices), because of symmetry and linearity assumptions in the demand by investors. O'Hara (1997) already anticipated that independence of the spread from the level of inventories was not very intuitive, and could probably be overcome by relaxing the traditional assumption of a constant fundamental - or pre-spread - value for the underlying good.[18] Our model has no symmetry and linearity assumptions on demand, which is endogenized. More than that, and consistently with O'Hara's intuition, our equilibrium builds on a non-constant fundamental value. The bid-ask spread $1/q - s$ increases with the width of the cone when going from top to bottom in the Table, as O'Hara's suggestion commands.

Consistently with the barriers behavior, the frequency of trade decreases when the risk aversion difference goes up. With $\gamma' - \gamma = 1$, the time to trade goes to 11 years.

We cannot compare across lines the growth rate of derived utility of the market maker, since his risk aversion (which enters it, since $\beta' \doteq r\gamma' - \delta'$) changes, but we can compare the growth rate of the investor, which is monotonical in the difference between the investor's and market maker's risk aversion, $\gamma' - \gamma$. The less apart are the risk aversions (top), the higher is δ and the smaller is β. At $\gamma' - \gamma = 0.15$, as a joint effect of a smaller spread and more frequent transactions, because the cone is less open, there is a smaller rate of growth ($\beta = -7.99\%$) than with more distant risk aversions, higher spread and less frequent interventions, since with $\gamma' - \gamma = 1$ we have $\beta = -7.80\%$. So, if the investor optimizes his interventions, he can offset the higher spread in equilibrium.

[18] All others equal, she claims that "the movement of a fixed spread around the true price may no longer be optimal if the price itself is moving".

4 Model set up with outside option

In this section we give investors the outside option to wait and trade in a competitive market, instead of trading with the intermediary. This should enable us to understand how competition affects equilibrium bid-ask spreads. In order to model the outside option, we assume that over the next instant the market can still be an intermediated one, or investors can find themselves in a state where they can trade competitively at no cost.

We investigate a continuous two-state Markov-regime model, meant to formalize the previous idea. In the first regime or state s_1 the investor can match his trade with other investors and transact without costs. In the second regime or state s_2 he can trade only with the market maker and undergoes transaction costs $1 - s$,$1/q - 1$. Switching among the two states $X_t = s_1, s_2$ is represented by a Markov transition matrix Q where the entries $Q_{i,j}$, $i, j \in 1, 2$ are defined as

$$Q_{i,j} = \begin{cases} \liminf_{h \to 0} \frac{1 - P(X(t+h) = s_i | X(t) = s_j)}{h} & i = j \\ \liminf_{h \to 0} \frac{P(X(t+h) = s_i | X(t) = s_j)}{h} & i \neq j \end{cases}$$

We specify Q as follows:

$$Q = \begin{pmatrix} -\lambda_1 & \lambda_1 \\ \lambda_2 & -\lambda_2 \end{pmatrix}$$

$$\lambda_1, \, \lambda_2 > 0.$$

So, λ_1, λ_2 are the instantaneous probabilities of switching respectively to the intermediated (the costly) and disintermediated (the costless) market, while $1 - \lambda_1, 1 - \lambda_2$ are the probabilities of staying in the costless and costly states.[19]

[19]The transition probability from one state to the other is

$$P(X(t+h) = j | X(t) = i) = \delta_{ij} + Q_{ij} h + o(h),$$

where δ_{ij} is the Kronecker delta.

The expected time to switch from the costless to the costly state, or the expected lenght of time one spends in the costless state, when just arriving in it, is $1/\lambda_1$, while $1/\lambda_2$ is the expected time spent in the costly state. The stationary distribution, i.e. the long run proportion of time that the process spends in states s_1 and s_2 for $t \to \infty$ is

$$\pi = (\lambda_2/(\lambda_1 + \lambda_2), \lambda_1/(\lambda_1 + \lambda_2)) .$$

and is equal to 1/2 as long as $\lambda_1 = \lambda_2$. Obviously, the model without outside option explored so far obtains as a limit case when $\lambda_1 = 1, \lambda_2 \to 0$, while a competitive market, in which investors trade without intermediation, would obtain in the limit, for $\lambda_1 \to 0, \lambda_2 = 1$.

4.1 The investor problem

The maximization problem for the investor becomes a system in the two value functions J^{i_1}, J^{i_2} - which apply respectively when starting from the costless and costly states s_1 and s_2. In the no-cost state he keeps θ constant, in the costly state he will decide interventions according to a L, U policy. His problem can be written similarly to Dimitrakas (2008):

$$\begin{cases} \max\limits_{y,x} \left\{ rxJ^{i_1}_x + \alpha y J^{i_1}_y + \dfrac{\sigma^2}{2} y^2 J^{i_1}_{yy} - (\beta + \lambda_1) J^{i_1} + \lambda_1 J^{i_2} \right\} = 0 \\ \max\limits_{(L(t),U(t))} \left\{ rxJ^{i_2}_x + \alpha y J^{i_2}_y + \dfrac{\sigma^2}{2} y^2 J^{i_2}_{yy} - (\beta + \lambda_2) J^{i_2} + \lambda_2 J^{i_1} \right\} = 0 \end{cases} \tag{30}$$

which is a couple of PDEs for $J^{i_1}$and J^{i_2}. They are solved by verification.

For the costless state, namely the first PDE, we assume a solution of the type[20] $J^{i_1} = C(x+y)^\gamma$, since the value function should be the standard, Merton's one. Note however that, since the PDE for J^{i_1} is not the same as in the

[20]Notice that J^{i_1} is homogeneous of degree γ and at the optimum $J^{i_1}_x = J^{i_1}_y$

Merton's case (it includes J^{i_2}) , the optimal constant policy will not be θ^* as in (6). Call it θ^m.

For the costly state, and the second PDE, we guess a solution of the type $J^{i_2} = x^\gamma \boldsymbol{K}(y/x)$ by homotheticity; we transform the variables as follows: $\theta \doteq y/x$, $y = \theta/(1+\theta)W$, $x = 1/(1+\theta)W$ and substitute in the second equation. We get

$$(\delta - \lambda_2)\,\boldsymbol{K} + (\alpha - r)\,\theta \boldsymbol{K}' + \frac{\sigma^2}{2}\theta^2 \boldsymbol{K}'' + \lambda_2 (1+\theta)^\gamma = 0 \tag{31}$$

where $\delta \doteq r\gamma - \beta$, so that $\boldsymbol{K}$ is the solution of the corresponding homogeneous equation plus a particular solution of the complete equation. If the roots ρ_1, ρ_2 of the second degree equation

$$(\delta - \lambda_2) + (\alpha - r)\,\rho + \frac{\sigma^2}{2}\rho(\rho - 1) = 0. \tag{32}$$

are real, namely if

$$\delta \leq \frac{4r^2 - 8r\alpha + 4\alpha^2 + 8\lambda_2\sigma^2 + 4r\sigma^2 - 4\alpha\sigma^2 + \sigma^4}{8\sigma^2} \tag{33}$$

the solution for $\boldsymbol{K}$ is:

$$\boldsymbol{K}(\theta) = \mathbb{A}\theta^{\rho_1} + \mathbb{B}\theta^{\rho_2} + K_p(\theta). \tag{34}$$

Here

$$K_p(\theta) = \frac{2\lambda_2 \mathbb{C}}{\sigma^2(\rho_1 - \rho_2)} \left(\theta^{\rho_2} \int_{\theta^m}^{\theta} \frac{(1+t)^\gamma}{t^{\rho_2+1}} \mathrm{d}t - \theta^{\rho_1} \int_{\theta^m}^{\theta} \frac{(1+t)^\gamma}{t^{\rho_1+1}} \mathrm{d}t \right)$$

$\mathbb{A}, \mathbb{B}, \mathbb{C}$ are constant.

Substituting for the J^{i_1}, J^{i_2} functions in the first equation of the system and writing down the first order condition for the maximization with respect to θ^m, we get the following two equations:

$$\mathbb{C}\left[\gamma\left((\alpha-r)\frac{\theta^m}{1+\theta^m}-(1-\gamma)\frac{\sigma^2}{2}\left(\frac{\theta^m}{1+\theta^m}\right)^2\right)-\delta-\lambda_1\right] \tag{35}$$

$$= -\lambda_1\frac{\boldsymbol{K}(\theta^m)}{(1+\theta^m)^{\gamma}}$$

$$\left[\alpha-r-(1-\gamma)\sigma^2\frac{\theta^m}{1+\theta^m}\right]\mathbb{C} = \lambda_1\frac{\boldsymbol{K}(\theta^m)}{(1+\theta^m)^{-(1-\gamma)}} - \frac{\lambda^1}{\gamma}\frac{\boldsymbol{K}'(\theta^m)}{(1+\theta^m)^{\gamma-2}} \tag{36}$$

or, substituting for $\mathbb{C}$ from the first equation:

$$-\boldsymbol{K}(\theta^m)\left[\frac{\alpha-r-(1-\gamma)\sigma^2\frac{\theta^m}{1+\theta^m}}{\gamma\left((\alpha-r)\frac{\theta^m}{1+\theta^m}-(1-\gamma)\frac{\sigma^2}{2}\left(\frac{\theta^m}{1+\theta^m}\right)^2\right)-\delta-\lambda_1}-1+\theta^m\right]+\frac{\boldsymbol{K}'(\theta^m)}{\gamma(1+\theta^m)^{-2}}=0 \tag{37}$$

By value-matching and smooth-pasting, the BCs at the lower and upper levels $\theta = l, u$ of the no-transaction zone are:

$$\boldsymbol{K}'(l) = \frac{\gamma}{q+l}\boldsymbol{K}(l) \tag{38}$$

$$\boldsymbol{K}''(l) = \frac{\gamma-1}{q+l}\boldsymbol{K}'(l) \tag{39}$$

$$\boldsymbol{K}'(u) = \frac{\gamma}{1/s+u}\boldsymbol{K}(u) \tag{40}$$

$$\boldsymbol{K}''(u) = \frac{\gamma-1}{1/s+u}\boldsymbol{K}'(u) \tag{41}$$

Hence, the investor's problem is solved - and the solutions for (32) are real - when the equations from (37) to (41) are satisfied and (33) holds, since these represent the system of PDEs, their boundary conditions, the FOC for the optimal portfolio in the costless state and the requirement for non-negativity of the discriminant of (32). The system formed by the BCs themselves is linear homogeneous in the integration constants, so that its determinant cannot be zero (otherwise the value function nullifies). The problem is solved by equating to zero that determinant and solving (37) and the two equations which obtain by substituting for the BCs at l first, at u then into the ODE (31) and inserting the guessed solution for $\boldsymbol{K}$. This provides a total of four equations. If we had to solve the investor's problem on its own, we would solve these equations for l, u, θ^m, δ, using (33) as a constraint. Below we will solve these equations together with the specialist's ones.[21]

4.2 The market-maker problem

Also for the market maker there are two different value functions, depending on the state he starts from. Since the market maker does not transact in the costless state s_1, there is no optimization in that state; he optimizes only in state s_2. The two value functions J^{s_1},J^{s_2} are defined as:

$$J^{s_1}(x,y,t;T) \doteq \lim_{T\to\infty} \mathbb{E}\left[e^{-\beta'(T-t)}W(T)^{\gamma}/\gamma\right] \tag{42}$$

$$J^{s_2}(x,y,t;T) \doteq \lim_{T\to\infty} \sup \mathbb{E}\left[e^{-\beta'(T-t)}W(T)^{\gamma}/\gamma\right] \tag{43}$$

The system of equations which characterize these value functions can be written as:

$$\begin{cases} J^{s_1}_x rx + J^{s_1}_y \alpha y + J^{s_1}_{yy}\sigma^2 y^2/2 - (\beta' + \lambda_1)J^{s_1} + \lambda_1 J^{s_2} & =0 \\ \underset{s,q}{max}\left\{J^{s_2}_x rx + J^{s_2}_y \alpha y + J^{s_2}_{yy}\sigma^2 y^2/2 - (\beta' + \lambda_2)J^{s_2} + \lambda_2 J^{s_1}\right\} & =0 \end{cases} \tag{44}$$

To solve the system, it is still possible to guess a solution of the type $J^{s_{1,2}} = (-x)^{\gamma'} I^{s_{1,2}}(\theta)$ and transform the system itself into two differential equations in

[21] As in the no-outside option case, a similar solution for $\boldsymbol{K}$ (with hyperbolic sines and cosines) obtains when the roots of (32) are imaginary, since δ does not satisfy (33). We disregard this case here, but the reader can easily build it following the remapping of couples of imaginary roots described in Dumas-Luciano (1991) and used in the no-outside-option case.

$I^{s_{1,2}}(\theta)$.

$$\begin{cases} (\delta' - \lambda_1)I^{s_1} + (\alpha - r)\theta I^{s_1\prime} + \frac{\sigma^2}{2}\theta^2 I^{s_1\prime\prime} + \lambda_1 I^{s_2} = 0 \\ (\delta' - \lambda_2)I^{s_2} + (\alpha - r)\theta I^{s_2\prime} + \frac{\sigma^2}{2}\theta^2 I^{s_2\prime\prime} + \lambda_2 I^{s_1} = 0. \end{cases} \tag{45}$$

The system (45) in I^{s_1} and I^{s_2} can be solved obtaining I^{s_1} as a function of I^{s_2} from the second equation and substituting it in the first equation. The first equation becomes a differential equation of 4^{th} order in I^{s_2}. Solutions are in the form:

$$I^{s_2} = c_1\theta^{x_1} + c_2\theta^{x_2} + c_3\theta^{x_3} + c_4\theta^{x_4} \tag{46}$$

where $c_i, i = 1, ..4$ are constant, $x_{1,2} = \varsigma \pm \sqrt{\nu 1}$, $x_{3,4} = \varsigma \pm \sqrt{\nu 2}$,where

$$\begin{aligned} \varsigma &\doteq \frac{r - \alpha + \sigma^2/2}{\sigma^2} \\ \nu 1 &\doteq -8\delta'\sigma^2 + (-2r + 2\alpha - \sigma^2)^2 \\ \nu 2 &\doteq -4(2\delta' - 2\lambda_1 - 2\lambda_2)\sigma^2 + (-2r + 2\alpha - \sigma^2)^2 \end{aligned}$$

if $\nu 1$,$\nu 2$ are positive or null.[22] This occurs if δ' satisfies the following two inequalities:

$$\delta' \leq \frac{(-2r + 2\alpha - \sigma^2)^2}{8\sigma^2} \tag{47}$$

$$\delta' \leq \frac{8(\lambda_1 + \lambda_2)\sigma^2 + (-2r + 2\alpha - \sigma^2)^2}{8\sigma^2} \tag{48}$$

The first inequality implies the second.

Value matching and modified smooth pasting conditions must be derived in

[22] As in the outside-option investor's case, the solution could be written in terms of trigonometric sines and cosines if x_1, x_2, x_3, x_4 were all imaginary, while it could be written with real powers (or hyperbolic sines and cosines) as well as trigonometric ones if two roots were real, two imaginary. This would give four cases for the market maker, that have to be coupled with two cases for the investor, for a total of eight cases. We leave the specification to the reader.

the same vein as in the single-state case, and give:[23]

$$\begin{cases} lI^{s_2\prime}(l) = \gamma' I^{s_2}(l)\varepsilon_l \\ uI^{s_2\prime}(u) = \gamma' I^{s_2}(u)\varepsilon_u \\ \left[\frac{\partial l}{\partial q}(1-\gamma')+1\right] I^{s_2\prime}(l) + (q+l)\, I^{s_2\prime\prime}(l) = 0 \\ \left[\frac{\partial u}{\partial s}(1-\gamma')\, s + u\right] I^{s_2\prime}(u) + \frac{\partial u}{\partial s}(1+us)\, I^{s_2\prime\prime}(u) - \gamma' I^{s_2}(u) = 0 \end{cases} \tag{49}$$

Note that the BCs – considered as a system in the constants c_1, c_2, c_3, c_4 – are a linear homogeneous system. To avoid a null value function, we make sure that the corresponding determinant is equal to zero. Then we substitute first the BCs for l, then the ones for u in the ODE for I^{s_2}, obtaining two equations. The market maker problem is solved by the triple (δ', s, q) that nullifies the determinat and solves the equations so obtained, under the constraints on δ'.

5 Equilibrium

In order to find a solution, we solve for the seven unknowns (δ, δ', l, u, s, q, θ^m) the system of seven equations that corresponds to equating to zero both determinants, solving the BCs of both parties and satisfying (37), under the assumed constraints on δ, δ'. Given that the implicit function theorem allows us to compute the derivatives of the barriers in closed form, the solution is in quasi-closed form and unique up to the constants (δ, δ') here too.

5.1 Numerical illustration

We explore solutions for the asset parameters' case above, namely $\alpha - r = 5\%$, $\sigma^2 = 4\%, \gamma' - \gamma$ ranging from 0.15 to 1. As a starting point, we consider switching

[23] While in that case we obtained in explicit form the derivatives of l, u in the investor's problem and could substitute them in the market maker problem (distinguishing the cases of real versus imaginary solutions), here the optimization conditions for the investor can no longer be solved explicitly, and the derivatives of l, u with respect to the costs s, q must be computed numerically.

parameters $\lambda_1 = 10/100, \lambda_2 = 5/100$, which mean that the expected time spent in the costless state (its duration) is 10 years, the duration of the costly state is 20 years. The corresponding stationary distribution is $\pi = (1/3, 2/3)$: in the long run, the investor spends one third of the time in the costless state, two thirds in the costly state. We get the results in Table 2 below.

[insert here Table 2]

We can comment the results either by comparing Table 2 (row by row) with Table 1, namely comparing the case without and with limit orders' competition, or in terms of how the "with-outside option" market behaves when the risk aversion difference increases (from top to bottom of Table 2).

Note first, comparing row by row Tables 1 and 2, that in the outside-option case, equilibrium bid prices are higher, ask are lower, and spreads are smaller than in the no-outside-option case, as expected. When the difference in risk aversion is 0.15, for instance, bid prices are 99.99%, ask prices are 121%, higher than the ones we had with no outside option, namely 84% and 132 %. The same holds for the spread: with a switch to the costless state possible, even though the costless state is less likely than the costly one and its duration is lower, the bid-ask spread (21%) is lower than in the no-outside option case (48%).

As a consequence of the prices behavior, the lower barrier in the presence of switches to a frictionless market is higher, while the upper barrier is lower than in the case without switches. Intervention barriers are less far apart in the outside-option case than in the no-outside-option one.

The first time to trade requires a more sophisticated computation than in the single-state case, since trade can occur in both states. As a preliminary result useful just for the sake of comparison, we compute the first time-to-trade conditional on starting in the cost case. It is computed as the minimum time between the switching time to state s_1 (costless state), which is $1/\lambda_2$, and the crossing time of one the barriers, which is computed according to formula (29), with θ^* substituted by θ^m. Given that we are considering a low value of λ_2, in the numerical example the first time to trade coincides with the expected time to reach one of the trading barriers with the intermediary. The times to trade with outside option are 7.7 years with $\gamma' - \gamma = 0.15$ (when they were 13.4 without outside option) and 11.6 when $\gamma' - \gamma = 1$ (when they were 20.73). A reduction in the distance of the barriers from Table 1 to Table 2 brings about a reduction in the time to trade, as expected.

As for the behavior of the case with outside option in itself when risk aversion of the market maker goes down, and the difference in risk aversion goes up, as from top to bottom in Table 2, the behavior is similar to the one without ouside option. The bid-ask spread and the level of the barriers are shown respectively in Figure 4,5 below

[insert here Figure 4]

[insert here Figure 5]

The bid price decreases, the ask increases and the spread goes up. So, the possibility of encountering a counterpart with whom to trade competitively (at no cost), even if this happens with low probability, does not affect the traditional result of Ho and Stoll, which we encountered without outside option. Now, however, the spread behavior is not the result of inelasticity, but of both the strategic behavior of the market maker and competition from limit orders.When $\gamma' - \gamma$ goes from 0.15 to 1, transaction costs at the upper barrier go from $1-s = 0.1\%$ to 5.7 5%, while those at the lower barrier go from $1/q-1 = 20.85\%$ to 24.31%. The asymmetry between bid and spreads, with the former much smaller than the latter, already encountered without outside option, but not included in partial equilibrium models, is very strong here too.

As a consequence of the costs and spreads behavior, and as intuition would suggest, the cone opens up along the table. The boundaries still contain the Merton's ratio, where the investor optimally sets his portfolio when no costs exist.

As for the rate of growth of expected utility within the outside-option case, remind that we cannot compare across lines β'. β increases, so that the reaction of the investor, faced with an higher spread, succeeds in increasing his rate, as in the no-outside-option case. In Table 3 we observe that β increases from -0.1006, with $\gamma' - \gamma = 0.15$, to -0.0999, with $\gamma' - \gamma = 1$.

[insert here Table 3]

5.2 Sensitivity analysis

This section explores the effect of having the same switching parameters to the costly and costless state, and then raising one at a time, so as to understand the dynamics of spreads and barriers when the duration and the long-run probability of the two states moves. We then compare our results with the general equilibrium results for OTC consol bonds' markets in Duffie *et al.* (2005).

We introduce in Table 4 the equilibrium results when the switching parameters are equal, so that at each point in time the probability of being able to trade with another agent instead of being forced to pay the spread to the market maker (the probability of going from the costly to the costless state), or viceversa, are the same. We do this for a level of risk aversion difference equal to 1, and for two levels of this (common) probability, 5% and 10%:

[insert here Table 4]

The Table shows that, as the switching probabilities increase at the same time (from the first to the second line, namely from $\lambda_1 = \lambda_2 = 5/100$ to

$\lambda_1 = \lambda_2 = 1/10$), the bid and ask price both increase, and consequently the lower barrier increases and the upper decreases, with the total spread and cone shrinking. In Figure 6 we plot the behavior of the bid and ask prices as a function of the level of $\lambda_1 = \lambda_2$: in order to interpret the results, we can decompose the effect into the effect of raising λ_1 for given λ_2 and viceversa.

[insert here Figure 6]

To understand the effect of raising λ_1 for given λ_2, we must compare the first line in Table 4 ($\lambda_1 = \lambda_2 = 5/100$) with the line corresponding to a risky aversion difference of 1 in Table 2, where $\lambda_1 = 1/10, \lambda_2 = 5/100$. By doing so, we keep the probability of being able to trade with another investor constant (at $\lambda_2 = 5/100$), and increase from 5% to 10% λ_1, the instantaneous probability of being forced to trade through an intermediary. Equivalently, we are increasing the long-run probability of staying in the intermediated state, and lowering the duration of the costless state. The result is that the bid price goes down, the ask goes up, the spread increases, the lower barrier decreases, the upper increases, and the cone of no transaction widens (with its pivotal value θ^m changing too). So, if all others equal, the long run probability of being in the costly state goes up, and the long run probability of being free from spreads goes down, the price that the market maker is willing to pay decreases, the one he asks is higher. He can increase transaction costs. The investor trades less with the intermediary, in order to incur less costs as a whole, as intuition would suggest.

To understand the effect of raising λ_2 for given λ_1, we must compare the line corresponding to a risky aversion difference of 1 in Table 2, where $\lambda_1 = 1/10, \lambda_2 = 5/100$, with the second line in Table 4 ($\lambda_1 = \lambda_2 = 5/100$). By so doing, we are increasing the instantaneous probability of finding a counterpart who is not a market maker, increasing the long-run probability of staying in the disintermediated state, and decreasing the duration of the intermediated state.

The effect is the following: the bid price goes up, the ask goes down, the spread decreases, the lower barrier increases, the upper decreases, and the cone of no transaction shrinks (with its pivotal value θ^m changing too). Actually, if we increase λ_2 to 30%, the spread reaches a realistic value of 2.7%, with the cone even smaller.

So, all others equal, if the long run probability of being in the costless state goes up (λ_2 up), or the long run probability of being subject to spreads goes down (λ_1 down), the price that the market maker is willing to pay increases, the one he asks is lower. He decreases transaction costs, and, with our risk aversion and return parameters, a probability of $3/4$ to trade in the disintermediated market, with an expected time spent with the market maker of three years, leads to a spread of 2.7%.

In the two panels of Figure 7 below we plot the effects of changing either λ_1 (upper panel) or λ_2 (lower panel) on the bid and ask prices.

[insert here Figure 7]

What the two panels show is that increasing λ_1, the instantaneous probability of finding an intermediary, or being forced to trade in an intermediated market, and decreasing λ_2, the instantaneous probability of finding a counterpart with whom to trade at no cost, have similar effects, as expected. We also understand from the panels that the overall behavior when both λ_1 and λ_2 increase at the same pace, as represented in Figure 6, is caused by λ_1 dominating on the ask price, and being dominated on the bid. However, in Figure 6 the long run probabilities and expected time in the two states are equal, with the former staying at $1/2$, ewhile this is not the case in Figures 7. The only thing that changes in Figure 6 are the instantaneous, or conditional, probabilities of switching, while in Figure 7 both it and long-run values change.

In Figure 8 below we plot the effects of changing λ_1 and λ_2 together on the transaction boundaries, while in the two panels of Figure 9 we let λ_1 (upper panel) and λ_2 (lower panel) change, while keeping the other parameter fixed.

[insert here Figures 8 and 9]

The boundaries in Figure 8 behave as the ones in the lower panel of Figure 9, as if the effect of λ_2 prevailed over λ_1. As it happens with Figure 6 versus 7, one must be careful though in this interpretation, since in Figure 8 the long-run probability stays at $1/2$ both for the intermediated and disintermediated state, while this does not happen in Figure 9. So, as with prices, the joint effect of conditional probabilities keeps the long run constant, while their separate effect does not keep it constant. Note also that Figures 8 and 9 are in sharp contrast with the prediction of the corresponding partial equilibrium model (the solution of the investor's only problem, for given costs), in which both cones opened up, and the effect of λ_1 was in the same direction, but numerically much less important, than the one of λ_2. That was interpreted by Dimitrakas as saying that the existence, not the duration, of the costless state, opened up the cone marginally. In our case the adjustment of costs is endogenous and optimally chosen by the intermediary, when the probability exogenously change. This makes the difference with respect to the partial equilibrium result.

Our predictions are in line with empirical regularities and consistent with the general equilibrium through a search model in Duffie *et al.* (2005). Two facts are relevant. First, "smaller" investors, those with fewer search options in Duffie, for whom the duration of the costly state is higher (smaller λ_2 here), receive less favourable prices . This is consistent also with the lab experiments in Lamoreux and Schnitzlein (1997). Duffie *et al.* (2005) recall that the existence in equilibrium of higher spreads for smaller investors makes their contribution profoundly different from the traditional information-based literature, which assigns greater spreads to more informed - intuitively, "bigger" - investors. From our model we also know that smaller investors choose wider no intervention regions, while there is no optimal policy of that type in Duffie. In addition, since trade frequency is endogenous in our setting, the traders which deserve higher spreads are the ones which intervene less frequently. This, on top of the smaller likelihood of finding a counterpart with whom they can trade competitively, is consistent with them being "small" traders. Last but not least, we can study how spread are affected by changes in the difference in risk aversion of market participants, something which is not possible in Duffie *et al.* (2005), because all market participants are risk neutral, and in Duffie *et al.* (2007), where risk aversion is equal across agents and market makers do not exist.

The other empirical regularity both our model and Duffie explain is that even a monopolistic market maker, as the one we have, may have smaller spreads if traders can actively meet each other (λ_2 goes up): it suffices to go from the last line in Table 2 to the second line of Table 3 (i.e., to keep $\lambda_1 = .1$, while increasing λ_2 from 5 to 10%), or to look at the lower panel of Figure 7, to note that the bid price goes up, the ask goes down, and the spread decreases. Realistic levels of spreads can then be obtained also in our non-calibrated model, when λ_2 goes closer to 1.

Duffie *et al.* consider the presence of low spreads for monopolistic market makers a good explanation of the fact that reforms allowing "the public", namely other investors, to compete with dealers in the Nasdaq were very effective in reducing spreads. Our model, which captures risk aversion and risk aversion heterogeneity between market makers and investors. adds a further dimension: for given (and high) λ_2, the spread is the smaller the smaller is the risk aversion difference. It suffices to compare the spread for $\lambda_2 = 10\%$ and unit difference (16%), with the one for difference equal to .15 (13%).

5.3 Welfare analysis

An interesting effect on welfare is provided by increasing the volatility of capital, to which intermediaries give access. How are β and β' affected by a change in σ? Bid prices increase, ask go down, the spread does too, the cone shrinks, so that trade becomes more frequent. This is interesting per se, since we usually connect an increase in volatility to an increase in trade. Here the effect is endogenous. It is accompanied by an increase in β and β'. So, a decrease in spreads due to an increase in in exogenous volatility, even if accompanied by more frequent trade, makes investors better off; at the same time, market makers, even if faced by smaller trade, intervene more frequently, for given λs, and can be better off. It is one situation in which increased volatility is not harmful, and goe through trade.

Another interesting perspective on welfare consists in assessing the effects of intermediary's capital constraints. He and Krishnamurthy (2013) indeed, with the aim of assessing the effect of intermediation, developed an equilibrium model in which the risk premium - both its level, the time it takes to recover from shocks and the effect of government policies - depend on intermediaries being capital-constrained or not. In their model, developed, as our own, in order not to treat intermediaries as a "veil", but to fully uncover their role in equilibrium, capital constraints arise because households are not willing to "invest" more than n of the intermediary's wealth in him. Given this investment, the intermediary can take positions in the risky and riskless asset. Since the riskless asset is in zero-net supply, though, the constraint in He and Krishnamurthy corresponds to the following constraint in our model:

$$(n-1)y \leq nx$$

With $x > 0$, as in our solutions, and $n > 1$, as calibrated by He and Krishnamurthy on the base of ownership data of banks and compensations from hedge funds, the constraint on the intermediary becomes: $\theta \leq n/(n-1)$, and $\theta \leq 4/3$ according to their specific calibration to match how much of stock markets return goes on average to intermediaries (20%). So, the effects of capital constraints in our model can be judged from the effect on the rates of growth of derived utility of upper boundaries on θ, with an upper boundary of 4/3 if we directly apply their calibration. However, to mimick the situation of He and Krishnamurthy, it would be better to calibrate n so that the bid-ask spread is 20%, which would deliver a different upper bound depending on the other parameters $(\gamma, \gamma', \alpha, r, \sigma)$. For $\gamma' - \gamma = 1$, $\lambda_1 = \lambda_2 = 5\%$, any upper bound smaller or equal to 77.3% is binding in our numerical examples.m

As a general rule, an upper bound on θ may be irrelevant, if the upper bound is greater or equal than u in equilibrium, or binding, in the opposite case. Obviously, a binding constraint will prevent both the intermediary and the investor to reach their first best, and will decrease their welfare consequently. In the numerical examples provided so far, 4/3 is never binding, but for instance an upper bound smaller or equal to 77.3% would be binding, when $\gamma' - \gamma = 1$, $\lambda_1 = \lambda_2 = 5\%$. This shows that, as in He and Krishnamurthy, capital constraints may affect market outcome. Not in risk premia, as in their case, but in trading frequency and - above all - welfare.

5.4 Endogenous switching probabilities

One could be interested in determining the optimal level of λ_1, for given λ_2, if the market maker could intervene on it. Symmetrically, we could be interested in determining the optimal level of λ_2, for given λ_1, if investors could intervene on it. We consider the first problem, which parallels the discussion in Duffie *et al.* (2005): even in our model, in principle, the market maker could intervene on λ_1, by increasing its search for customers, without being able to affect the intensity by which limit order are executed and investors trade among themselves. We do not take for granted that pushing λ_1 to 1, for given λ_2, is his optimal choice, since we know from the previous Tables that it increases the spread and makes the barriers more distant, at least for the risk aversion levels studied so far (Including non-reported values at $\gamma' - \gamma = .15$). In general, one should find, if it exists

$$\max_{\lambda_1} \lim_{T \to \infty} K^s(x_s, y_s, t, T)$$

which is equivalent, if we assume that we start at 0 from the intermediated case, and from $\theta = \theta^m$, to searching for

$$\max_{\lambda_1} \exp^{-(r\gamma' - \delta'(\lambda_1))} I^{s_2}(\theta^m)$$

where the expression for the value function I^{s_2} is (46), and we have recalled that, in equilibrium, δ' is a function of λ_1. Substituting and recalling that the maximization has to be performed in equilibrium, the market maker should maximize

$$g(\lambda_1) \doteq$$
$$\exp^{-(r\gamma' - \delta'(\lambda_1))} \left[c_1(\lambda_1) (\theta^m)^{\sqrt{\nu 1(\lambda_1)}} + c_2(\lambda_1) (\theta^m)^{-\sqrt{\nu 1(\lambda_1)}} + c_3(\lambda_1) (\theta^m)^{\sqrt{\nu 2(\lambda_1)}} + (\theta^m)^{-\sqrt{\nu 1(\lambda_1)}} \right]$$

where the constants c_1, c_2, c_3 solve the system (49) and the fourth constant c_4 has been set to one, since the whole system has rank three.

5.5 Policy interventions

The asymmetric effect of increasing λ_1 versus λ_2, the predictions on spreads applied to small investors in contrast with information-based theories and the possibility of having relatively small spreads even in the presence of a monopolistic market makers, if investor can, as an alternative, wait and trade competitively - submit limit orders, in the microstructure language - does not only confirm empirical phenomena, but is also the basis of policy implications of our paper.

In the first respect (asymmetric effect of increasing λ_1 versus λ_2), policy manouvres that favour disintermediation can work either by decreasing λ_1 or increasing λ_2. These have indeed been recognized above to have similar effects.

In the second respect, increasing λ_2 can be interpreted as decreasing the difficulties and costs of small investors, which would not be the case with an information-theory based approach.

In the third respect, the competition from a disintermediated market is a valid alternative to competition between market makers.

Last, we can interpret the measures against proprietary trading of the so-called Volcker rule in the Dodd-Frank or Bank-Holding-Company Act in the US as an increase in λ_2. It is common believed that the implementation of the Volcker rule is going to disincetivize market making by regulated entities, thus reducing liquidity and the easiness of access and immediacy of execution of market orders by investors (see Duffie, 2012). In our model this raises bid, lower ask prices, reduces the bid-ask spread, leads to closer barriers, higher frequency of trade, and higher liquidity. So, our model argues for a positive effect of the Volcker rule.

A common interpretation of the Volcker rule though is that it was not meant to lower intermediation per se, but only speculative proprietary trading. In order to analyze the impact of the Volcker rule in the "speculative" position of the market maker in our model, we have to assess how the position he would take when he can trade on his own, in state 1. To do that, we have to solve for his value function in that state, something that we did not do so far, to write down the FOC with respect to his optimal holdings in that case, that we will call θ^{**}, and to compute its derivative with respect to λ_2. If the derivative is positive, this means that, all others (including risk aversion of the intermediary itself) equal, he will lever more (invest more in the risky in comparison to the riskless asset) when λ_2 - the chance to enter the possibly-speculative state, but also the long-run probability of being in it - increases. To solve for θ^{**}, we recall that the discounted value function in the corresponding state, I^{s_1}, is the solution to the second equation in (45), once we substitute in it I^{s_2} as given by (46), with any triple of integration constants c_1, c_2, c_3 that solve the system (49). The equation that defines θ^{**}, namely the FOC with respect to it, is

$$\left(\delta' - \lambda_2 + \alpha - r\right) I'\left(\theta^{**}\right) + \left(\alpha - r + \sigma^2\right) \theta^{**} I^{''}\left(\theta^{**}\right) + \sigma^2 \left(\theta^{**}\right)^2 I^{'''}\left(\theta^{**}\right)/2 = 0 \tag{50}$$

where we wrote I instead of I^{s_2} for simplicity. Note that both δ', the integration constants and powers in I^{s_2}, as well as θ^{**} depend on λ_2. Calling $F(\lambda_2, \theta^{**}; \alpha, r, \sigma, \gamma, \gamma')$ the left hand side of (50), the derivative we are interested in - by the implicit function theorem - is

$$\frac{\partial \theta^{**}}{\partial \lambda_2} = -\frac{\partial F/\partial \lambda_2}{\partial F/\partial \theta^{**}}$$

which can be computed explicitly.

6 Summary and conclusions

We characterized equilibrium bid-ask spreads and infrequent trade in symmetric-information, intermediated markets. We actually specified two cases: either investors are obliged to trade with the monopolistic market maker and incur into transaction costs, or they can wait until another trader - with whom they can trade at no cost - arrives. In each economy, we provided the optimality conditions for market participants. These conditions determine the equilibrium bid and ask spreads, as well as the value functions of the agents and intervention barriers - or trade - of the investor. Quasi-closed form solutions - unique up to two constants - are provided.

Our major contribution consists in endogenizing spreads and infrequent trade. When investors have the outside-option to wait and trade competitively, as expected, the magnitude of the bid-ask spreads goes down. It goes from 30% when the long-run probability of encountering another investor is 33% to 2.7% when it becomes 75%, without wiping infrequent trade out. It could easily go down further, when that probability goes towards 1. This shows that the competition from a disintermediated market is a valid alternative to competition between market makers.

Capital constraints on intermediaries are likely to prevent the first best outcome: their effect is not measured on the risk premium (as it is in He and Krishnamurthy), but on trade frequency and welfare.

We show that increasing the instantaneous probability of having a limit order executed, or the long-run probability of staying in the disintermediated state, or decreasing the duration of the intermediated state has the same effect than decreasing the instantaneous probability of being forced to trade with the market maker, or the long-run probability of facing him, or increasing the duration of the non-intermediated state. So, policy interventions can act on both. In particular, our model argues for a positive effect of the Volcker rule on spreads (reduced) and frequency of trade or liquidity (augmented), and enables us to compute the effects on speculation.

As a consequence of the previous effects, smaller investors are worse off (face bigger spreads) than big investors, opposite to what information theory would predict. This is the testable implication of the model, as it was for the spreads on consol bonds in Duffie (2005). In our case, the magnitude of the effect is related to risk aversion and is connected to trade frequency too.

7 Appendix A

The three steps for solving the optimization problem of the investor are as follows. First, we recognize that a candidate solution for the value function is either the first or the second expression in (12), according to whether $\delta > (\leq)\delta_c$. Indeed, the algebraic equation corresponding to (7), which provides the roots $x_{1,2}$, i.e. (15), has imaginary solutions in the first case, real in the second.

Second, we substitute both the first and second order BCs (8) into the ODE (7), so as to obtain a second degree equation for the optimal barriers l and u, through their transforms ε_l and ε_u. These are respectively the smaller and the bigger root of the following equation:

$$\delta + \gamma\left(\alpha - r\right)\varepsilon + \gamma\left(\gamma - 1\right)\sigma^2\varepsilon^2/2 = 0 \tag{51}$$

whose discriminant we denote as Δ :

$$\Delta \doteq \gamma^2(\alpha - r)^2 - 2\delta\gamma(\gamma - 1)\sigma^2 \tag{52}$$

The sign of the discriminant depends on γ and on whether $\delta \geq (\leq)\delta^*$, as stated in the text.

Third, we make the determinant of the value-matching BCs, considered as equations in (A, B) or $(\mathcal{A}, \mathcal{B})$, equal to zero. This guarantees that the value function is non-null. The determinant is equated to zero by a proper choice of the artificial discount rate β, via δ. This means solving for δ the algebraic equation $a(l,q)b(u,s) - c(u,s)d(l,q) = 0$. The solution requires substitution of the expressions for $l, u, \epsilon_l, \epsilon_u$ in terms of the parameters $\alpha - r, \sigma$ and δ itself. These expression obtain from (51).

Note that, in order to solve for l, u it is not necessary to solve for the constants A and B (or $\mathcal{A}, \mathcal{B}$).

8 Appendix B

In order to compute the derivatives in (25), we first use the definition of ε_l and ε_u, namely (9) and (10), determine explicitly the investor's barriers:

$$\begin{cases} l = \frac{N}{D-N} q, \\ u = \frac{N'}{D-N'} \frac{1}{s} \end{cases} \tag{53}$$

where

$$D = \gamma(\gamma - 1)\sigma^2, \tag{54}$$

$$N = -\gamma(\alpha - r) - \sqrt{\Delta}, \tag{55}$$

$$N' = N + 2\sqrt{\Delta}, \tag{56}$$

Based on them, dependence of l on q and u on s acts both directly and via the discount rate δ

$$\frac{\partial l}{\partial q} = \frac{1}{D - N} \left[N + q \frac{D^2}{(D - N)\sqrt{\Delta}} \frac{\partial \delta}{\partial q} \right] \tag{57}$$

$$\frac{\partial u}{\partial s} = \frac{1}{s\,(D - N')} \left[\frac{-N'}{s} - \frac{D^2}{(D - N')\sqrt{\Delta}} \frac{\partial \delta}{\partial s} \right] \tag{58}$$

Using the implicit function theorem to derive the discount rate sensitivities, $\frac{\partial \delta}{\partial q}, \frac{\partial \delta}{\partial s}$, one has:

$$\frac{\partial \delta}{\partial q} = -\frac{b(u,s)\frac{\partial a(l,q)}{\partial q} - c(u,s)\frac{\partial d(l,q)}{\partial q}}{\frac{\partial (ab-cd)}{\partial \delta}} \quad (59)$$

$$\frac{\partial \delta}{\partial s} = -\frac{a(l,q)\frac{\partial b(u,s)}{\partial s} - d(l,q)\frac{\partial c(u,s)}{\partial s}}{\frac{\partial (ab-cd)}{\partial \delta}}. \quad (60)$$

where the derivatives of a, b, c, d are easily obtained in closed form (separately for the imaginary and real case). Plugging (59) into (57) and (60) into (58), we have

$$\frac{\partial l}{\partial q} = \frac{1}{D-N}\left[N - q\frac{D^2}{(D-N)\sqrt{\Delta}}\frac{b(u,s)\frac{\partial a(l,q)}{\partial q} - c(u,s)\frac{\partial d(l,q)}{\partial q}}{\frac{\partial (ab-cd)}{\partial \delta}}\right] \quad (61)$$

$$\frac{\partial u}{\partial s} = \frac{1}{s(D-N')}\left[\frac{-N'}{s} + \frac{D^2}{(D-N')\sqrt{\Delta}}\frac{a(l,q)\frac{\partial b(u,s)}{\partial s} - d(l,q)\frac{\partial c(u,s)}{\partial s}}{\frac{\partial (ab-cd)}{\partial \delta}}\right] \quad (62)$$

which need to be substituted in the "modified" smooth pasting conditions (25) as well as in conditions (26) and (27). The latter enter into the equilibrium computation.

9 References

Amihud, Y. and H. Mendelson, 1986, Asset Pricing and the Bid-Ask Spread, *Journal of Financial Economics,* 17, pp. 223-49.

Buss, A., and B. Dumas, 2013, Financial-Market Equilibrium with Frictions, CEPR Discussion Paper No. DP9524 .

Constantinides, G., 1986, Capital Market Equilibrium with Transaction Costs, *Journal of Political Economy,* 94, pp. 842-862.

Dimitrakas, V., Portfolio selection with stochastic transaction costs, PhD dissertation, INSEAD, 2008.

Duffie, D., Market Making under the Proposed Volcker Rule, wp, Stanford University, 2012

Duffie, D., N. Gârleanu and L.H. Pedersen, 2005, Over-the-Counter Markets, *Econometrica*, 73, pp. 1815-1847.

Duffie, D., N. Gârleanu, and L.H. Pedersen, 2007, Valuation in Over-the-Counter Markets, *The Review of Financial Studies*, 20, pp. 1865-1900.

Dumas, B. and E. Luciano, 1991, An exact solution to a dynamic portfolio choice problem with transaction costs, *Journal of Finance*, 46, pp. 577-595.

Gerhold, S., Guasoni, P., Muhle-Karbe, J., and W. Schachermayer, 2011, Transaction Costs, Trading Volume, and the Liquidity Premium, *Finance and Stochastics,* 18 (1), 1-37.

He, Z., and A. Krishnamurthy. Intermediary Asset Pricing, *American Economic Review*, 103, 732-770

Ho, T., and H. Stoll, 1981, Optimal Dealer Pricing under Transactions and Return Uncertainty, *Journal of Financial Economics*, 9, pp. 47-73.

Lo, A., H. Mamaysky and J. Wang, 2004, Asset Pricing and Trading Volume under Fixed Transaction Costs, *Journal of Political Economy,* 112, pp. 1054-1090.

O'Hara, M., 1997, *Market microstructure theory*, Basil Blackwell, Cambridge, MA, 2nd ed.

Pagano, M., and A. Roell, 1993, Auction markets, dealership markets and execution risk, ch 8 in *Financial market liberalization and the role of banks*, ed. by Hamaui, R., Conti, V., Cambridge University Press.

Peskir, G. and A. Shyriaev, 2006, *Optimal Stopping and Free-Boundary Problems,* (Springer Verlag, Berlin).

Stoll, H., 1978, The supply of broker services in Securities Markets, *Journal of Finance*, 33, 1133-1151

Vayanos, D., 1998, Transaction Costs and Asset Prices: A dynamic equilibrium model, *The Review of Financial Studies*, 11, pp. 1-58.

Table 1: Bid-ask prices $(s, 1/q)$, spread, barriers (l, u) and optimal asset ratio (θ^*), without outside option, as a function of the difference in risk aversion $(\gamma' - \gamma)$.

Table 1: Equilibrium without outside option

$\gamma' - \gamma$	**s**	**q**	**1/q**	**spread**	**l**	θ^*	**u**	t^*
0.15	0.842	0.757	1.321	0.479	0.1809	0.4545	0.9041	13.40
0.5	0.841	0.750	1.333	0.492	0.1608	0.4545	0.9670	16.02
1	0.788	0.691	1.446	1.446	0.1336	0.4545	1.0937	20.73

Table 2: Bid-ask prices $(s, 1/q)$, spread, barriers (l, u) and optimal asset ratio (θ^*), $\lambda_1 = 10\%$, $\lambda_2 = 5\%$, with outside option, as a function of the difference in risk aversion $(\gamma' - \gamma)$.

Table 2: Equilibrium with outside option

$\gamma' - \gamma$	**s**	**q**	**1/q**	**spread**	**l**	θ^*	**u**	t^*
0.15	0.9999	0.8275	1.2085	0.2086	0.1823	0.4623	0.6918	7.661
0.5	0.9999	0.8206	1.2186	0.2187	0.1817	0.4610	0.7139	8.3489
1	0.9425	0.8044	1.2431	0.3006	0.1623	0.4584	0.8004	11.596

Table 3: δ and rate of growth of derived utility for the specialist and the investor (β' and β), $\lambda_1 = 10\%$, $\lambda_2 = 5\%$, with outside option, as a function of $\lambda_1 = \lambda_2$.

Table 3: Equilibrium with outside option

$\gamma' - \gamma$	δ'	δ	β'	β
0.15	0.01125	0.0406	-0.0683	-0.1006
0.5	0.01125	0.04056	-0.0613	-0.10056
1	0.01125	0.0399	-0.0513	-0.0999

Table 4: Bid-ask prices $(s, 1/q)$, spread, barriers (l, u) and optimal asset ratio (θ^*), with outside option, as a function of the difference in risk aversion $(\gamma' - \gamma = 1)$.

Table 4: Equilibrium with outside option

$\lambda_1 = \lambda_2$	**s**	**q**	**1/q**	**spread**	**l**	θ^*	**u**	t^*
5%	0.9500	0.8751	1.1428	0.1928	0.1793	0.4544	0.7735	10.26
10%	0.9999	0.8635	1.1581	0.1582	0.1902	0.4580	0.6616	6.68

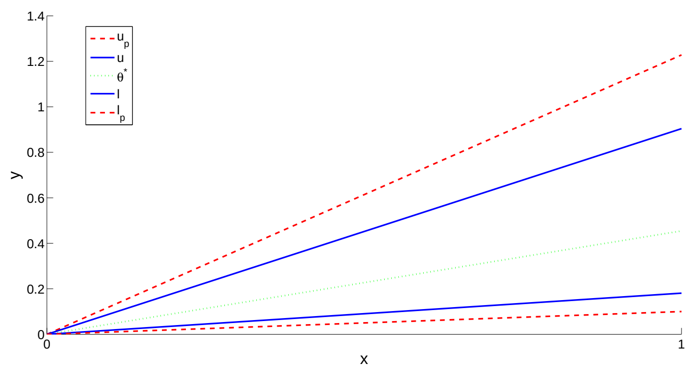

Figure 1: Cone of no-transactions without outside option, partial versus general equilibrium, at $\gamma' - \gamma = 0.15$.

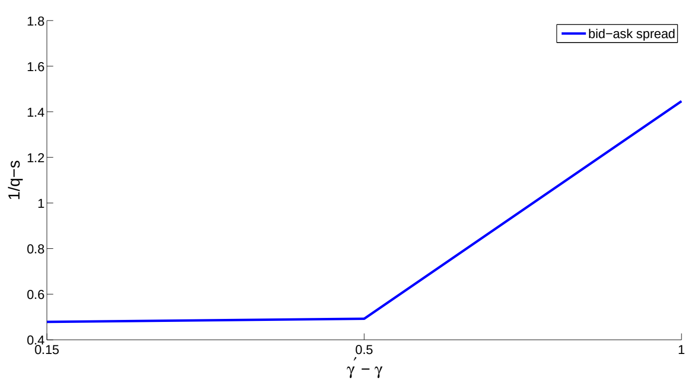

Figure 2: Bid-ask spread as a function of the difference between the investor's and market maker's risk aversion (without outside option).

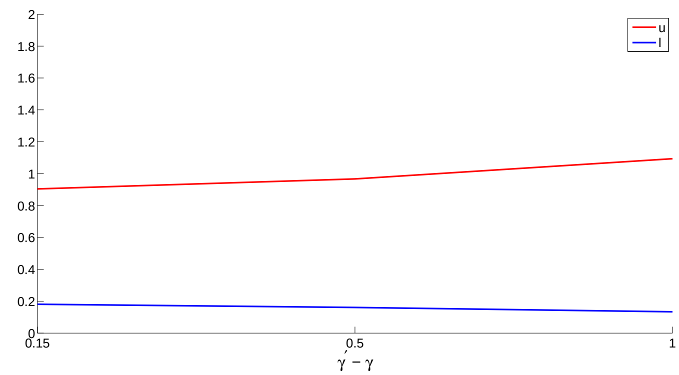

Figure 3: Barriers as a function of the difference between the investor's and market maker's risk aversion(without outside option).

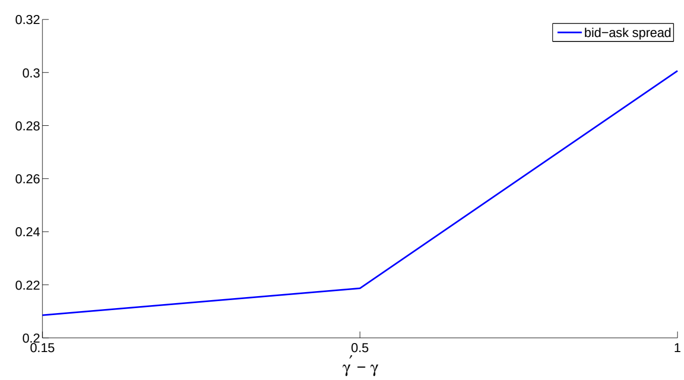

Figure 4: Bid-ask spread as a function of the difference between the investor's and market maker's risk aversion, at $\lambda_1 = 10\%$, $\lambda_2 = 5\%$ (with outside option).

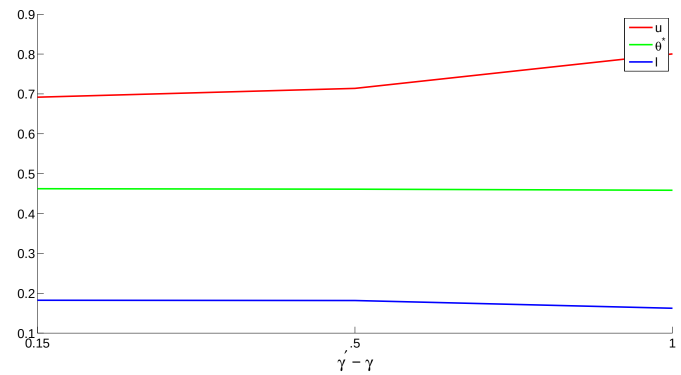

Figure 5: Barriers as a function of the difference between the investor's and market maker's risk aversion, at $\lambda_1 = 10\%$, $\lambda_2 = 5\%$ (with outside option).

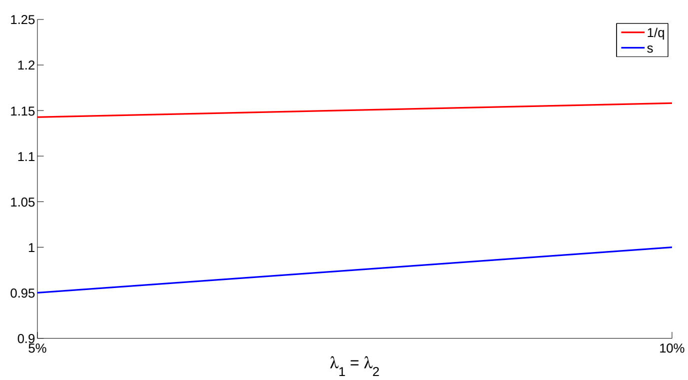

Figure 6: Bid-ask prices as a function of the level of $\lambda_1 = \lambda_2$, at risk aversion $\gamma' - \gamma = 1$ (with outside option).

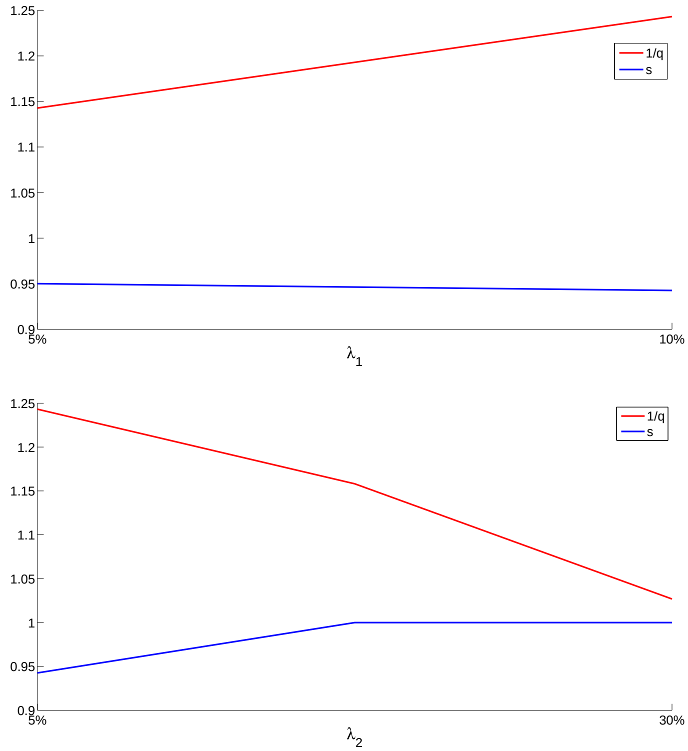

Figure 7: Bid-ask prices as a function of the level of λ_1 (upper panel) and λ_2 (lower panel), at risk aversion $\gamma' - \gamma = 1$ (with outside option).

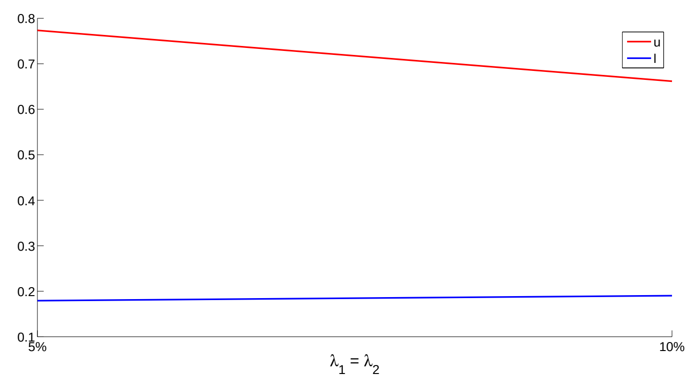

Figure 8: Barriers as a function of the level of $\lambda_1 = \lambda_2$, at risk aversion $\gamma' - \gamma = 1$ (with outside option).

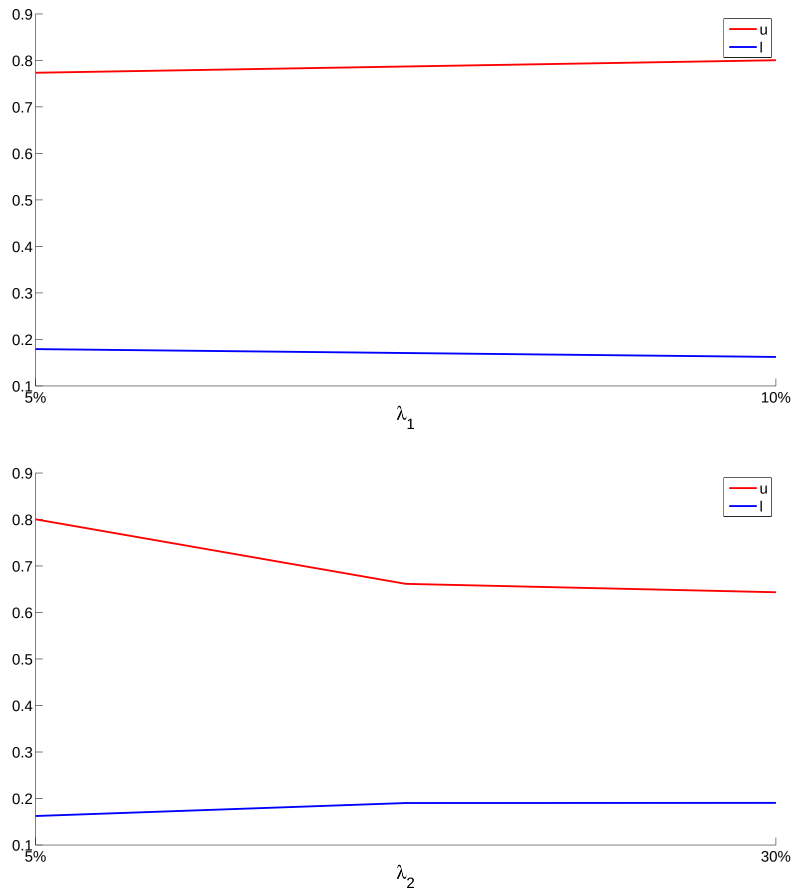

Figure 9: Barriers as a function of the level of λ_1 (upper panel) and λ_2 (lower panel), at risk aversion $\gamma' - \gamma = 1$ (with outside option).