

Financial-market Equilibrium with Friction*

Adrian Buss Bernard Dumas

June 28, 2012

Abstract

When purchasing a security an investor needs not only have in mind the cash flows that the security will pay into the indefinite future, he/she must also anticipate his/her desire and ability to resell the security in the marketplace at a later point in time. In this paper, we show that the endogenous stochastic process of the liquidity of securities is as important to investment and valuation as is the exogenous stochastic process of their future cash flows.

For that purpose, we develop a general-equilibrium model with heterogeneous agents that have an every day motive to trade and pay transactions fees.

Our method delivers the optimal, market-clearing moves of each investor and the resulting ticker and transactions prices. We use it to show the effect of transactions fees on asset prices, on deviations from the classic consumption CAPM and on the time path of transactions prices and trades, including their total and quadratic variations.

*Previous versions circulated and presented under the title: “The Equilibrium Dynamics of Liquidity and Illiquid Asset Prices”. Buss is with Goethe University, Frankfurt (buss@finance.uni-frankfurt.de) and Dumas is with INSEAD, NBER and CEPR (bernard.dumas@insead.edu). Work on this topic was initiated while Dumas was at the University of Lausanne. Dumas’ research has been supported by the Swiss National Center for Competence in Research “FinRisk” and by grant #1112 of the INSEAD research fund. The authors are grateful for useful discussions to Yakov Amihud, Bruno Biais, Georgy Chabakauri, Massimiliano Croce, Magnus Dahlquist, Xi Dong, Stefano Giglio, Francisco Gomes, Harald Hau, John Heaton, Terrence Hendershott, Julien Hugonnier, Elyès Jouini, Andrew Karolyi, Francis Longstaff, Frédéric Malherbe, Alexander Michaelides, Pascal Maenhout, Stavros Panageas, Luboš Pástor, Tarun Ramadorai, Scott Richard, Barbara Rindi, Olivier Scaillet, Norman Schürhoff, Raman Uppal, Grigory Vilkov, Vish Viswanathan, Ingrid Werner, Fernando Zapatero and participants at workshops given at the Amsterdam Duisenberg School of Finance, INSEAD, the CEPR’s European Summer Symposium in Financial Markets at Gerzensee, the University of Cyprus, the University of Lausanne, Bocconi University, the European Finance Association meeting, the Duke-UNC Asset pricing workshop, the National Bank of Switzerland, the Adam Smith Asset Pricing workshop at Oxford University, the Yale University General Equilibrium workshop and the Center for Asset Pricing Research/Norwegian Finance Initiative Workshop at BI.

When purchasing a security an investor needs not only have in mind the cash flows that the security will pay into the indefinite future, he/she must also anticipate his/her desire and ability to resell the security in the marketplace at a later point in time. In this paper, we show that the endogenous stochastic process of the liquidity of securities is as important to investment and valuation as is the exogenous stochastic process of their future cash flows.

At any given time, an asset is more or less liquid as a function of three conceivable mechanisms and their fluctuating impact, taken in isolation or combined. The first mechanism is the fear of default of the counterparty to the trade. Trade is obviously hampered by the fear that contracts will not be abided by. The second mechanism is informed trading (asymmetric information) as in the market for “lemons” (Akerlof (1970)).¹ A vast Microstructure literature stemming from Copeland and Galai (1983), Glosten and Milgrom (1985) and Kyle (1985) has shown that informed trading indirectly generates transactions costs.² The third mechanism, which we examine here, is the presence of fees charged for transacting, stemming (in an unmodelled way) from order processing costs and inventory holding costs and holding risks, all items which Stoll (2000) refers to as “real frictions”, as did Demsetz (1968).³

Access to a financial market is a service that investors make available to each other. In the real world, investors do not trade with each other. They trade through intermediaries called broker and dealers, who incur physical costs, are faced with potentially informed customers and charge a fee that is close to being proportional to the value of the shares traded. This service charge aims to cover the actual physical cost of trading and the adverse-selection effect plus a profit. This paper is not about the pricing policy of broker-dealers. We bypass them and let the investors serve as dealers for, and pay the fees to each other.

As a way of providing a simple model, we assume that the trading fee is proportional to the value of the shares traded. Given the presence of that fee, an investor may decide not to trade, thereby preventing other investors from trading with him/her, which is an additional endogenous, stochastic and perhaps quantitatively more important consequence of the fee.

Our goal is to study, in terms both of price and volume, the dynamics of a financial-market equilibrium that we can expect to observe when there are frictions and when investors have an every-day motive for trading, such as shocks to their endowments, that is separate from the long-term need to trade for lifetime planning purposes. Actually, we assume long-lived investors who trade because they have differing risk aversions while they have access only to a menu of linear assets, which, absent transactions fees, would be sufficient to make the

¹Bhattacharya and Spiegel (1998) have shown the way in which the lemon problem can cause markets to close down.

²And, in Asset Pricing, an even larger literature stemming from Grossman and Stiglitz (1980) and Hellwig (1980) shows how the risk created by asymmetric information or heterogeneous expectations is priced. See, among many, Kyle (1985), Easley and O’Hara (1987), Admati and Pfleiderer (1988), Easley, Hvidkjaer and O’Hara (2002) and O’Hara (2003).

³On the various possible determinants of liquidity, see the synthesis paper of Vayanos and Wang (2009). The classic empirical decomposition of stock-market spreads by Stoll (1989) concluded that real frictions represented 53% of the bid-ask spread.

market dynamically complete. Dynamic completeness is, of course, killed by the presence of transactions fees. The imbalance of the portfolios, which investors have to hold because of transactions fees, acts as an inventory cost.

Our paper is related to the existing studies of portfolio choice under transactions costs such as Constantinides (1976a, 1976b, 1986), Davis and Norman (1990), Dumas and Luciano (1991), Edirisinghe, Naik and Uppal (1993), Genotte and Jung (1994), Shreve and Soner (1994), Leland (2000), Longstaff (2001), Nazareth (2002), Bouchard (2002), Obizhaeva and Wang (2005), Liu and Lowenstein (2002), Jang, Koo, Liu and Lowenstein (2007) and Gerhold, Guasoni, Muhle-Karbe and Schachermayer (2011) among others. As was noted by Dumas and Luciano, these papers suffer from a logical quasi-inconsistency. Not only do they assume an exogenous process for securities returns, as do all portfolio optimization papers, but they do so in a way that is incompatible with the portfolio policy that is produced by the optimization. The portfolio strategy is of a type that recognizes the existence of a “no-trade” region. Yet, it is assumed that prices continue to be quoted and trades remain available in the marketplace.⁴ Obviously, the assumption must be made that some traders, other than the one whose portfolio is being optimized, do not incur costs. In the present paper, we assume that all investors face the trading fee.

In one interpretation, the inventory of securities held by each investor can be viewed as a state variable in the dynamics of our equilibrium, a feature that is shared with the inventory-management model of a dealer that has been pioneered by Ho and Stoll (1980, 1983) and which is one of the main pillars of the Microstructure literature. In their work, however, Ho and Stoll focus exclusively on the dealer’s problem, taking the arrival of orders to the dealer as an exogenous random process. Here, we fully endogenize each investor’s decision to trade and we derive the full general equilibrium.⁵ Orders do not arrive at random; they implement optimal portfolio choices.

The papers of Heaton and Lucas (1996), Vayanos (1998), Vayanos and Vila (1999) and Lo, Mamaysky and Wang (2004) are direct ancestors of the present one in that they have exhibited the equilibrium behavior resulting from transactions costs, although they have postulated a physical, deadweight cost of transacting.⁶ In the neighborhood in which transactions take place, Heaton and Lu-

⁴Constantinides (1986) in his pioneering paper on portfolio choice under transactions costs attempted to draw some conclusions concerning equilibrium. Assuming that returns were independently, identically distributed (IID) over time, he claimed that the expected return required by an investor to hold a security was affected very little by transactions costs. Liu and Lowenstein (2002), Jang, Koo, Liu and Lowenstein (2007) and Delgado, Dumas and Puopolo (2010) have shown that this is generally not true under non IID returns. The possibility of falling in a “no-trade” region is obviously a massive violation of the IID assumption.

⁵Recently, a partial-equilibrium literature has developed aiming to model the optimal tactic of a trader who (for unmodelled reasons) needs to trade and determines how to optimally place his orders in a limit-order market. See: Parlour (1998), Foucault (1999), Foucault, Kadan and Kandel (2005), Goettler, Parlour and Rajan (2005), Rosu (2009).

⁶In a previous version of our paper, we had assumed that trading entailed physical deadweight costs proportional to the number of shares traded. Another predecessor is Milne and Neave (2003), which, however, contains few quantitative results. The equilibrium with other costs, such as holding costs and participation costs, has been investigated by Peress (2005),

cas (1996) derive a stationary equilibrium under transactions cost but, in the neighborhood of zero trade, the cost is assumed to be quadratic so that investors trade all the time in small quantities and equilibrium behavior is qualitatively different from the one we produce here. In Vayanos (1998) and Vayanos and Vila (1999), an investor’s only motive to trade is the fact that he has a finite lifetime. Transactions costs induce him to trade twice in his life: when young, he buys some securities that he can resell in order to be able to live during his old age. Here, we introduce a higher-frequency motive to trade. In the paper of Lo, Mamaysky and Wang (2004), costs of trading are fixed costs, all traders have the same negative exponential utility function, individual investors’ endowments provide the motive to trade (as in our paper) but aggregate endowment is not stochastic. In our current paper, fees are proportional, utility is a power utility that differs across traders and endowments are free to follow an arbitrary stochastic process. To our knowledge, ours is the first paper to reach that goal.

A form of restricted trading is considered by Longstaff (2009) where a physical asset traded by two logarithmic investors is considered illiquid if, after being bought at time 0, it must be held till some date T after which it becomes liquid again. The consequences for asset prices are drawn in relation to the length T of the freeze.

One can also capture liquidity considerations by means of a portfolio constraint. Holmström and Tirole (2001) study a financial-market equilibrium in which investors face an exogenous constraint on borrowing.⁷ When they hit their constraint, investors are said to be “liquidity constrained”. Gromb and Vayanos (2002) and Brunnermeier and Pedersen (2008) study situations in which the amount of arbitrage capital is constrained.⁸ It would be necessary to present some microfoundations for the constraint. A constraint on borrowing would best be justified by the risk of default on the loan. Equilibrium with default is an important but separate topic of research.

The paper that is closest to our work is Buss *et al.* (2011). Both papers derive an equilibrium in a financial market where investors incur a cost when they transact and both use the backward-induction procedure of Dumas and Lyasoff (2011) to solve the model. Technically, the main difference between the two papers is that Buss *et al.* (2011) use a “primal” formulation and we use a “dual” one. In the dual approach, the personal state prices of the investors are among the unknowns and the same system of equations applies in the entire space of values of state variables while the shadow costs of trading must be included among the state variables. The primal approach does not require

Tuckman and Vila (2010) and Huang and Wang (2010).

⁷As is apparent below, the cost and the constraint approaches are somewhat similar but are probably not equivalent to each other. As we show, transaction costs or fees give rise to shadow prices of potentially being unable to trade that are specific to each asset and each investor, whereas a constraint gives rise to a dual variable that is specific to each investor only.

⁸Distant antecedents of this idea in the macroeconomic literature can be found in the form of Clower and Bushaw (1954) constraints, which required a household to hold some money balance, as opposed to being able to borrow, when it wanted to consume, as well as the “cash-in-advance” model of Lucas (1982).

the addition of shadow costs of trading among the state variables and requires instead the solution of a different system of equations in different regions of the state space. The economic insights generated by the two papers are also quite different. In our paper, the focus is on the microstructure effects of transactions fees and the pricing of liquidity risk in a market where investors trade a riskless asset and a single risky asset. In Buss *et al.* (2011), the transactions cost is a deadweight cost and the focus is on its effect on the cross section of asset returns, and they consider a model with multiple risky assets. Both papers allow for idiosyncratic endowments (which Buss *et al.* (2011) call labor income) but Buss *et al.* (2011) assume a stochastic process for the labor income that is separate from the output process. Finally, the investors in Buss *et al.* (2011) have Epstein-Zin-Weil utility rather than power utility.

As far as the solution method is concerned, our analysis is closely related, in ways we explain below, to “the dual method” used by Jouini and Kallal (1995), Cvitanic and Karatzas (1996), Kallsen and Muhle-Karbe (2008) and Deelstra, Pham and Touzi (2002) among others.

In computing an equilibrium, one has a choice between a “recursive” method, which solves by backward induction over time, and a “global” method, which solves for all optimality conditions and market-clearing conditions of all states of nature and points in time simultaneously.⁹ The global method, often implemented in the form of a homotopy, is limited in terms of the number of periods it can handle. Here, we resort to a recursive technique, which requires the choice of state variables – both exogenous and endogenous – that track the state of the economy. Dumas and Lyasoff (2010) have proposed an efficient method to calculate incomplete-market equilibria recursively with a dual approach, which utilizes state prices as endogenous state variables. We use the same method here with the addition of dual state variables that capture the cost of trading. A crucial advantage of using dual variables as state variables to handle proportional-transactions costs problems is that the variables thus introduced evolve on a fixed domain, namely the interval set by the unit cost of buying and the cost of selling (with opposite signs), whereas primal variables, such as portfolio choices evolve over a domain that has free-floating barriers, to be determined.

Empirical work on equilibria with transactions costs has been couched in terms of a CAPM that recognizes a number of risk factors. Brennan and Subrahmanyam (1996), Pástor and Stambaugh (2003) and Acharya and Pedersen (2005) have recognized two or more risk factors, one of which is the market return (as in the classic CAPM) or aggregate consumption (as in the consumption-CAPM), and the others are meant to capture stochastic fluctuations in the degree of liquidity of the market, either taken as a whole or individually for each security. Liquidity fluctuations are proxied by fluctuations in volume or in the responsiveness of price to the order flow. The papers cited confirm that there exist in the marketplace significant risk premia related to these factors. Our model also identifies additional risk factors for the investors’ willingness

⁹For an implementation of the global solution, see Herings and Schmedders (2006).

to trade, in the form of shadow prices. However, there is one such per investor and they are not directly observable. We use our model to ascertain to what extent proxies used in the empirical literature are to any degree related to these shadow prices.^{10 11}

After writing down our model and specifying the solution method (Section 1), we focus our work on two main questions. First, we ask in Section 2 whether equilibrium securities prices conform to the famous dictum of Amihud and Mendelson (1986a), which says that they are reduced by the present value of transactions costs. In Section 3, we examine the behavior of the market over time, asking, for instance, to what degree price changes and transactions volume are related to each other and what effects transactions fees have on the point process of transaction prices. In Section 4, we quantify the additional premia that are created by transactions fees and which are deviations from the consumption CAPM. These are the drags on expected-return that empiricists would encounter as a result of the presence of transactions fees.

1 Problem statement: the objective of each investor and the definition of equilibrium

We start with a population of two investors $l = 1, 2$ and a set of exogenous time sequences of individual endowments $\{e_{l,t} \in \mathbb{R}_{++}; l = 1, 2; t = 0, \dots, T\}$ on a tree or lattice. For simplicity, we consider a binomial tree so that a given node at time t is followed by two nodes at time $t + 1$ at which the endowments are denoted $\{e_{l,t+1,u}, e_{l,t+1,d}\}$. The transition probabilities are denoted $\pi_{t,t+1,j}$ ($\sum_{j=u,d} \pi_{t,t+1,j} = 1$).¹² Notice that the tree accommodates the exogenous state variables only.^{13 14}

In the financial market, there are two securities, defined by their payoffs

¹⁰Transactions costs also constitute a “limit to arbitrage” and offer a potential explanation of the observed fact that sometimes securities that are closely related to each other do not trade in the proper price relationship. For these deviations to appear in the first place, however, and subsequently not be obliterated by arbitrage, some category of investors must introduce some form of “demand shock”, that can only result from some departure from von Neumann-Morgenstern utility. Here, we consider only rational behavior so that no opportunities for (costly) arbitrage arise in equilibrium.

¹¹Empirical work has also been done by Chordia *et al.* (2008) and others to track the dynamics of liquidity as it moves from one category of assets to another. In the present paper, the menu of assets is too limited to throw any light on the evidence presented by these papers.

¹²Transition probabilities generally depend on the current state but we suppress that subscript.

¹³As has been noted by Dumas and Lyasoff (2010), because the tree only involves the exogenous endowments, it can be chosen to be *recombining* when the endowments are Markovian, which is a great practical advantage compared to the global-solution approach, which would require a tree in which nodes must be distinguished on the basis of the values of not just the exogenous variables but also the endogenous ones.

¹⁴It would be straightforward to write the equations below for more agents and more complex trees. The implementation of the solution technique is much more computationally intensive with more than two agents while it is not more complicated with a richer tree.

$\{\delta_{t,i}; i = 1, 2; t = 0, \dots, T\}$.¹⁵ The “ticker” prices of the securities, which are not always transactions prices, are denoted: $\{S_{t,i}; i = 1, 2; t = 0, \dots, T\}$. The ticker price is an effective transaction price if and when a transaction takes place but it is posted all the time by the Walrasian auctioneering computer (which works at no cost).

Financial-market transactions entail transactions fees that are paid by investors to each other. When an investor sells one unit of security i , turning it into consumption good, he receives from the buyer in units of consumption goods the ticker price multiplied by $1 - \varepsilon_{i,t}$ and the buyer of the securities must pay to the seller the ticker price times $1 + \lambda_{i,t}$. However, when investor l decides to sell a security and to pay a fee for that service, he/she does not take into account the fact that his decision is linked to the other investor’s decision to buy the security from him, for which he/she will collect a fee. The fee collected is viewed as a lumpsum, which does not enter first-order conditions.¹⁶ With symbol $\theta_{l,t,i}$ standing for the number of units of Security i in the hands of Investor l after all transactions of time t , Investor l solves the following problem:¹⁷

$$\sup_{\{c_l, \theta_l\}} \mathbb{E}_0 \sum_{t=0}^T u_l(\tilde{c}_{l,t}, t)$$

subject to:

- terminal conditions:

$$\theta_{l,T,i} = 0,$$

- a sequence of flow budget constraints:

$$\begin{aligned} & c_{l,t} + \sum_{i=1,2} [\theta_{l,t,i} - \theta_{l,t-1,i}]^+ S_{t,i} \times (1 + \lambda_{i,t}) \\ & + \sum_{i=1,2} [\theta_{l,t,i} - \theta_{l,t-1,i}]^- S_{t,i} \times (1 - \varepsilon_{i,t}) \\ = & e_{l,t} + \sum_{i=1,2} \theta_{l,t-1,i} \delta_{t,i} + \sum_{i=1,2} [\theta_{l',t,i} - \theta_{l',t-1,i}]^+ S_{t,i} \lambda_{i,t} \quad (1) \\ & - \sum_{i=1,2} [\theta_{l',t,i} - \theta_{l',t-1,i}]^- S_{t,i} \varepsilon_{i,t}; \forall t; l' \neq l \end{aligned}$$

¹⁵It so happens that, without transactions fees, the market would be complete. But the derivations and the solution technique depend neither on the number of branches in the tree, nor on the number of securities. We could solve for the equilibrium with transactions fees in a market that would be incomplete to start with.

¹⁶When the buy and sell fees are equal, as is the case in our numerical illustrations below, this assumption is equivalent to investors being compensated (in the Hicksian sense) for the fees they incur on their transactions. As a result, transactions fees generate no income/wealth effect, only substitution effects.

¹⁷The tilda $\tilde{\cdot}$ is a notation we use to refer to a random variable.

- and given initial holdings:¹⁸

$$\theta_{l,-1,i} = \bar{\theta}_{l,i} \quad (2)$$

In the flow budget constraint, the term $\sum_{i=1,2} [\theta_{l,t,i} - \theta_{l,t-1,i}]^+ S_{t,i} (1 + \lambda_{i,t})$ reflects the net cost of purchases and the term $\sum_{i=1,2} [\theta_{l,t,i} - \theta_{l,t-1,i}]^- S_{t,i} (1 - \varepsilon)$ captures the net proceeds of sales of securities. And the terms on the right-hand side involving investor l' represent the net fees collected when the other investor trades.

The dynamic programming formulation of the investor's problem is:¹⁹

$$J_l(\{\theta_{l,t-1,i}\}, \cdot, e_{l,t}, t) = \sup_{c_{l,t}, \{\theta_{l,t,i}\}} u_l(c_{l,t}, t) + \mathbb{E}_t J_l(\{\theta_{l,t,i}\}, \cdot, \tilde{e}_{l,t+1}, t+1)$$

subject to the flow budget constraint written at time t only.

Writing $\theta_{l,t,i} = \hat{\theta}_{l,t,i} + \tilde{\theta}_{l,t,i} - \theta_{l,t-1,i}$, one can reformulate the same problem to make it more suitable for mathematical programming:

$$\begin{aligned} J_l(\{\theta_{l,t-1,i}\}, \cdot, e_{l,t}, t) &= \sup_{c_{l,t}, \{\hat{\theta}_{l,t,i}, \tilde{\theta}_{l,t,i}\}} u_l(c_{l,t}, t) \\ &+ \mathbb{E}_t J_l\left(\left\{\hat{\theta}_{l,t,i} + \tilde{\theta}_{l,t,i} - \theta_{l,t-1,i}\right\}, \cdot, \tilde{e}_{l,t+1}, t+1\right) \end{aligned} \quad (3)$$

subject to:

$$\begin{aligned} &c_{l,t} + \sum_{i=1,2} (\hat{\theta}_{l,t,i} - \theta_{l,t-1,i}) S_{t,i} (1 + \lambda_{i,t}) \\ &+ \sum_{i=1,2} (\tilde{\theta}_{l,t,i} - \theta_{l,t-1,i}) S_{t,i} (1 - \varepsilon_{i,t}) \\ &= e_{l,t} + \sum_{i=1,2} \theta_{l',t-1,i} \delta_{t,i} + \sum_{i=1,2} (\hat{\theta}_{l',t,i} - \theta_{l',t-1,i}) S_{t,i} \lambda_{i,t} \\ &- \sum_{i=1,2} (\tilde{\theta}_{l',t,i} - \theta_{l',t-1,i}) S_{t,i} \varepsilon_{i,t} \end{aligned} \quad (4)$$

$$\hat{\theta}_{l,t,i} \leq \theta_{l,t-1,i} \leq \hat{\theta}_{l,t,i} \quad (5)$$

Definition 1 An equilibrium is defined as a process for the allocation of consumption $c_{l,t}$, a process for securities prices $\{S_{t,i}\}$ such that the supremum of

¹⁸It is assumed that $\sum_{i=1,2} \bar{\theta}_{l,i} = 0$ or 1 depending on whether the security is assumed to be in zero or positive net supply.

¹⁹The form $J_l(\{\theta_{l,t-1,i}\}, \cdot, e_{l,t}, t)$ in which the value function is written refers explicitly only to investor l 's individual state variables. The complete set of state variables actually used in the backward induction is chosen below.

(3) is reached for all l, i and t and the market-clearing conditions:²⁰

$$\sum_{l=1,2} \theta_{l,t,i} = 0 \text{ or } 1; i = 1, \dots, 2 \quad (6)$$

are also satisfied with probability 1 at all times $t = 0, \dots, T$.

In Appendix A, we show, using a shift of equations proposed in the context of incomplete markets by Dumas and Lyasoff (2010), that the equilibrium can be calculated, for given initial values of some endogenous state variables, which are the dual variables $\{\phi_{l,t}, R_{l,t,i}\}$ – as opposed to given values of the original state variables, viz., initial positions $\{\theta_{l,t-1,i}\}$ –, by solving the following equation system written for $l = 1, 2; j = u, d; i = 1, 2$. The shift of equations amounts from the computational standpoint to letting investors at time t plan their time- $t+1$ consumption $c_{l,t+1,j}$ but choose their time- t portfolio $\theta_{l,t,i}$ (which will finance the time- $t+1$ consumption).²¹

1. First-order conditions for time $t+1$ consumption:

$$u'_l(c_{l,t+1,j}, t+1) = \phi_{l,t+1,j}$$

2. The set of time- $t+1$ flow budget constraints for all investors and all states of nature of that time:

$$\begin{aligned} & e_{l,t+1,j} + \sum_{i=1,2} \theta_{l,t,i} \delta_{t+1,i,j} - c_{l,t+1,j} \\ & - \sum_{i=1,2} (\theta_{l,t+1,i,j} - \theta_{l,t,i}) \times R_{l,t+1,i,j} \times S_{t+1,i,j} \\ & = \sum_{i=1,2} \left(\hat{\theta}_{l',t+1,i,j} - \theta_{l',t,i} \right) \times S_{t+1,i,j} \times \lambda_{i,t+1,j} \\ & - \sum_{i=1,2} \left(\hat{\theta}_{l',t+1,i,j} - \theta_{l',t,i} \right) \times S_{t+1,i,j} \varepsilon_{i,t+1,j} \end{aligned}$$

3. The third subset of equations says that, when they trade them, all investors must agree on the prices of traded securities and, more generally, they must agree on the posted “ticker prices” inclusive of the shadow prices R that make units of paper securities more or less valuable than units of consumption. Because these equations, which, for given values of $R_{l,t+1,i,j}$, are linear in the unknown state prices $\phi_{l,t+1,j}$, restrict these to

²⁰One equates $\sum_{l=1,2} \theta_{l,t}$ to 0 or 1 depending on whether the security is or is not in zero net supply.

²¹ u'_l denotes “marginal utility” or the derivative of utility with respect to consumption.

lie in a subspace, we call them the “kernel conditions”:

$$\begin{aligned} & \frac{1}{R_{1,t,i} \times \phi_{1,t}} \sum_{j=u,d} \pi_{t,t+1,j} \times \phi_{1,t+1,j} \times (\delta_{t+1,i,j} + R_{1,t+1,i,j} \times S_{t+1,i,j}) \\ & = \frac{1}{R_{2,t,i} \times \phi_{2,t}} \sum_{j=u,d} \pi_{t,t+1,j} \times \phi_{2,t+1,j} \times (\delta_{t+1,i,j} + R_{2,t+1,i,j} \times S_{t+1,i,j}) \end{aligned} \quad (7)$$

4. Definitions:

$$\theta_{l,t+1,i,j} = \widehat{\theta}_{l,t+1,i,j} + \widehat{\widehat{\theta}}_{l,t+1,i,j} - \theta_{l,t,i}$$

5. Complementary-slackness conditions:

$$\begin{aligned} (-R_{l,t+1,i,j} + 1 + \lambda_{i,t+1,j}) \times (\widehat{\theta}_{l,t+1,i,j} - \theta_{l,t,i}) &= 0 \\ (R_{l,t+1,i,j} - (1 - \varepsilon_{i,t+1,j})) \times (\theta_{l,t,i} - \widehat{\widehat{\theta}}_{l,t+1,i,j}) &= 0 \end{aligned}$$

6. Market-clearing restrictions:

$$\sum_{l=1,2} \theta_{l,t,i} = 0 \text{ or } 1$$

7. Inequalities:

$$\widehat{\widehat{\theta}}_{l,t+1,i,j} \leq \theta_{l,t,i} \leq \widehat{\theta}_{l,t+1,i,j}; 1 - \varepsilon_{i,t+1,j} \leq R_{l,t+1,i,j} \leq 1 + \lambda_{i,t+1,j};$$

This is a system of 24 equations (not counting the inequalities) where the unknowns are $\left\{ c_{l,t+1,j}, \phi_{l,t+1,j}, R_{l,t+1,i,j}, \theta_{l,t,i}, \widehat{\theta}_{l,t+1,i,j}, \widehat{\widehat{\theta}}_{l,t+1,i,j}; l = 1, 2; j = u, d \right\}$. This is a total of 24 unknowns. We solve the system by means of the Interior-Point algorithm, in a simplified version of the implementation of Armand *et al.* (2008).²²

Besides the exogenous endowments $e_{l,t+1,j}$, the “givens” are the time- t investor-specific shadow prices of consumption $\{\phi_{l,t}; l = 1, 2\}$ and of paper securities $\{R_{l,t,i}; l = 1, 2; i = 1, 2\}$, which must henceforth be treated as state variables and which we refer to as “endogenous state variables”. Actually, given the nature of the equations, the latter variables can be reduced to state variables: $\frac{R_{2,t,i}}{R_{1,t,i}}$

²²The Interior-Point method, which involves relaxed Karush-Kuhn-Tucker complementary-slackness conditions, turns inequality constraints into equations. It is more compatible with Newton solvers than the alternative method proposed earlier by Garcia and Zangwill (1981), which involves discontinuous functions such as $\max[\cdot, \cdot]$.

and $\frac{\phi_{1,t}}{\phi_{1,t}+\phi_{2,t}}$ all of which are naturally bounded a priori: $\frac{1-\varepsilon_{i,t}}{1+\lambda_{i,t}} \leq \frac{R_{2,t,i}}{R_{1,t,i}} \leq \frac{1+\lambda_{i,t}}{1-\varepsilon_{i,t}}$
and $0 \leq \frac{\phi_{1,t}}{\phi_{1,t}+\phi_{2,t}} \leq 1$.²³

In addition, the given securities' price functions $S_{t+1,i,j}$ are obtained by backward induction (see, in Appendix A, the third equation in System (15)):

$$\begin{aligned} S_{t,i} &= \frac{1}{R_{l,t,i}\phi_{l,t}} \sum_{j=u,d} \pi_{t,t+1,j} \phi_{l,t+1,j} \times (\delta_{t+1,i,j} + R_{l,t+1,i,j} \times S_{t+1,i,j}); \\ S_{T,i} &= 0 \end{aligned} \tag{8}$$

and the given future position functions $\theta_{l,t+1,i,j}$ (satisfying $\sum_{l=1,2} \theta_{l,t+1,i,j} = 0$ or $1; i = 1, ..2$) are also obtained by an obvious backward induction of $\theta_{l,t,i}$, the previous solution of the above system, with terminal conditions $\theta_{l,T,i} = 0$. All the functions carried backward are interpolated by means of piecewise, third-degree polynomials.

Moving back through time till $t = 0$, the last portfolio holdings we calculate are $\theta_{l,0,i}$. These are the post-trade portfolios held by the investors as they *exit* time 0. We need to translate these into *entering*, or pre-trade, portfolios holdings so that we can meet the initial conditions (2). The way to do that is explained in Appendix B.

2 Equilibrium asset holdings and prices at the initial point in time

In our benchmark setup, we consider two investors who have isoelastic utility and have different coefficients of relative risk aversion. One of them only (Investor $l = 1$) receives a flow endowment. In that sense, he has a “liquidity advantage.” The desire to trade arises from the differences in the endowments and in the risk aversions.

As for securities, the subscript $i = 1$ refers to a short-lived riskless security in zero net supply and the subscript $i = 2$ refers to equity *in positive supply*. We call “equity” a long-lived claim that pays the endowment of Investor 1 ($\delta = e_1$). Transactions fees are levied on trades of equity shares (the “less liquid” asset); none are levied on trades of the riskless asset, which is also, therefore, the “more liquid” asset. The economy is of a finite-horizon type with $T = 50$. The single exogenous process is the endowment process of the first investor, which is represented by a binomial tree, with constant geometric increments mimicking a geometric Brownian motion.

²³The two variables $\phi_{1,t}$ and $\phi_{2,t}$ are one-to-one related to the consumption shares of the two investors, so that consumption scales are actually used as state variables. Consumption shares of the two agents add up to 1 because of the transactions fees are paid in a reciprocal fashion.

The numerical illustration below cannot in any way be seen as being calibrated to a real-world economy.²⁴ Indeed, as has been noted in the introduction, investors in our model trade because they have different endowments and differing risk aversions while they have access only to a menu of linear assets. The imbalanced portfolios they have to hold because of transactions fees act as an inventory cost similar to the cost incurred in inventory-management model of the Ho-and-Stoll (1980, 1983) variety. But our model does not include two other motives for trading that are obviously present in the real world such as the liquidity-trading motive (arising from missing securities and endowment shocks that would be incompletely hedgeable even if the market were frictionless) and the speculative motive (arising from informed trading due to private signals or to differences of opinion). Above all, we have two traders, not millions. For these reasons, although our goal is to capture a higher-frequency motive to trade, the amount of trading we are able to generate is not sufficient to match high-frequency data quantitatively. We, therefore, keep a yearly trading interval because we need to cover a sufficient number of years to get some reasonable amount of trading. Even so, we are going to document interesting patterns that match real-world data qualitatively.

We demonstrate a property of scale invariance, which will save on the total amount of computation: all the nodes of a given point in time, which differ only by their value of the exogenous variable, are isomorphic to each other, where the isomorphy simply means that we can factor out the endowment. In this way, we do not need to perform a new calculation for each node of a given point in time; one suffices.²⁵

Table 1 shows all the parameter values.²⁶ The risk aversion of Investor 1 is lower than that of Investor 2, so that Investor 1 is a natural borrower, as far as

²⁴In this pure-exchange general-equilibrium economy, total consumption is equal to total endowments plus total dividends. And, in order to limit the number of exogenous processes, we have set dividends on the equity equal to the endowment. In order to capture some properties of real-world equity, we choose a process for all of these that reflects the behavior of dividends. The following set of papers document dividend dynamics. Lettau and Ludvigson (2005) write: “An inspection of the dividend data from the CRSP value-weighted index [] reveals that [] the average annual growth rate of dividends has not declined precipitously over the period since 1978, or over the full sample. The average annual growth rate of real, per capita dividends is in fact higher, 5.6%, from 1978 through 1999, than the growth rate for the period 1948 to 1978. The annual growth rate for the whole sample (1948-2001) is 4.2%.” Volatility is reported to be 12.24%. Earlier evidence includes Campbell and Shiller (1988) who report for periods up to 1986 dividend growth rates of around 4%. Recently, van Binsbergen and Koijen (2010) estimate a growth rate of 5.89%.

A good mean value given this evidence is then probably to use a drift of 4.5% with a volatility of 13%.

²⁵This property, which we prove in Appendix C, holds even though investors have different risk aversions. Remarkably, the property is valid when $\frac{R_{2,t,i}}{R_{1,t,i}}$ and $\frac{\phi_{1,t}}{\phi_{1,t}+\phi_{2,t}}$ are used as endogenous state variables of the backward recursion. With different risk aversions across investors, *it would not have held* if, as in the primal approach, the endogenous state variables had been $\{\theta_{l,t-1,i}\}$, the pre-trade portfolios held when entering each point in time t .

²⁶The initial holdings of equity $\bar{\theta}_{1,2}$ by Investor 1 are just that. Separately, Investor 1 receives his/her endowment, which is the same stream of consumption units as the equity stream.

Table 1: **Parameter Values and Benchmark Values of the State Variables.** This table lists the parameter values used for all the figures in the paper. The table also indicates the benchmark values of state variables, which are reference values taken by all state variables except for the particular one being varied in a given graph.

Name	Symbol	Value	Range
<i>Parameters for exogenous endowment</i>			
Horizon of the economy	T	50 years	
Expected growth rate of endowment		3.9%/year	
Time step of the tree		1 year	
Volatility of endowment		16.2%/year	
Initial endow. at $t = 0$ (cons. units)		1	
<i>Parameters for the investors</i>			
Investor 1's risk aversion	γ_1	2	
Investor's risk aversion	γ_2	4	
Investor 1's time preference	β_1	0.975	[0.9, 0.99]
Investor 2's time preference	β_2	0.975	
<i>Transactions fees per dollar of equity traded</i>			
When buying and when selling	$\lambda = \varepsilon$	1%	[0%, 3%]
<i>Benchmark values of the variables</i>			
Initial hold. of riskless asset by Inv. 1	$\theta_{1,1}$	0	[-30, 10]
Initial holding of equity by Investor 1	$\theta_{1,2}$	0	[-1, 1]
Initial hold. of riskless asset by Inv. 2	$\theta_{2,1}$	0	[-10, 30]
Initial holding of equity by Investor 2	$\theta_{2,2}$	1	[0, 2]

the riskless short-term security is concerned.

2.1 Equilibrium asset holdings

It is well-known from the literature on non-equilibrium portfolio choice that proportional transactions costs cause the investors to tolerate a deviation from their preferred holdings. The zone of tolerated deviation is called the “no-trade region”. In previous work, the no-trade region had been derived for a given stochastic process of securities prices. We now obtain the no-trade region in general equilibrium, when two investors make analogous portfolio decisions and prices are set to clear the market.

2.1.1 Equilibrium no-trade region

Figure 1, panel (a) plots the no-trade region for the different values of the initial holdings of securities. The lighter grey zone is specifically the no-trade region while the darker zone is the trade region. When the holdings with which Investor 1 enters the trading date are in the trade region, the investors trade to reach the edge of the no-trade region; to the contrary, when the holdings upon entering

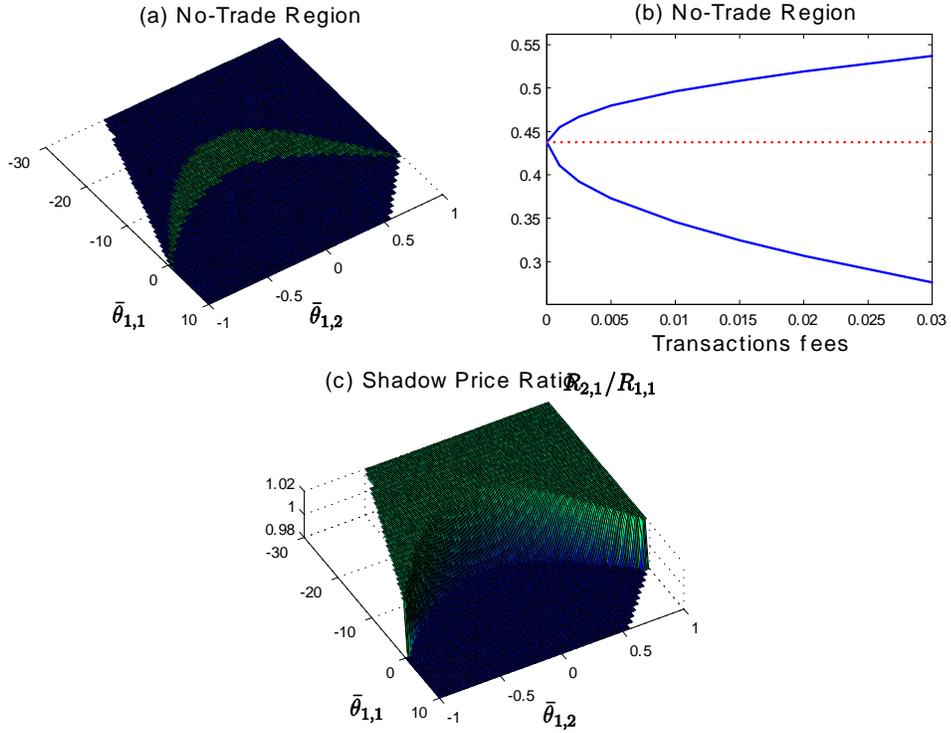


Figure 1: **Equilibrium no-trade region.** Panel (a) shows the no-trade region for different “entering” positions $\bar{\theta}$ of the agents. Transactions fees are equal to $1\% \times$ value of shares traded, while Panel (c) displays the ratio of shadow prices across the trade and no-trade regions. Panel (b) shows the no-trade region for different levels of transaction fees from 0% to 3%. Consumption shares are set at the value corresponding to the initial holdings of Table 1. In all panels, parameters are as in Table 1.

the trading date are within the no-trade region, the investors do nothing. The crescent shape of the no-trade zone is the result of the difference in risk aversions between the two investors: there exists a curve (not shown) inside the zone which would be the locus of holdings in a frictionless, complete market. The white zone of the figure, on both sides of the dark grey zone, is not admissible; when entering holdings are in that zone, there exists no equilibrium as one investor would, at equilibrium prices, be unable to repay his/her negative positions to the other investor. Panel (b) of the same figure displays, for the benchmark values of the variables, the width of the no-trade region against the rate of transactions fees. Panel (c) illustrates how the shadow prices vary across the trade and no-trade regions: in one trade region, the shadow price per unit of endowment of one investor is equal to $1 + \lambda$ (the buy transaction fee) while the other investor's shadow is equal to $1 - \varepsilon$ (the sell transaction fee) and in the other trade region, the opposite is true. The ratio between their two shadow prices is, therefore, $\frac{1+\lambda}{1-\varepsilon} = 1.02$ or 0.98 . Within the no-trade region the difference is between these two numbers, with a discrete-version of the smooth-pasting condition holding on the optimal boundary and causing the shadow-price difference to taper off smoothly. The result is analogous to the no-trade region and the relative price of the equilibrium shipping model of Dumas (1992), with the difference that the trades considered are not costly arbitrages between *geographic locations* in which physical resources have different prices but are, instead, costly arbitrages between *people* whose private valuations of paper securities differ.

Figure 2 shows, against the rate of transactions fees, the holdings of the stock and bond with which Investor 1 exits a trading period in which he enters with initial holdings $(0, 0)$. While this investor, who is less risk averse, is a natural borrower and thus chooses negative positions in the bond, increased transactions fees induce him to carry on with a smaller holding of equity. For that reason, he has to borrow less.

2.1.2 The clientele effect

Do more patient investors hold less liquid assets as in the “clientele effect” of Amihud and Mendelson (1986)? We now vary the patience parameter of the first investor between 0.9 and 0.99. Figure 3 provides a clear illustration of the clientele effect: as Investor 1 becomes more patient, he/she holds more of the stock, which is the illiquid security and less of the short-term bond, which is the more liquid one. The result, however, depends very much on the initial holdings, here assumed to be 0 of the short-term bond and 0 of the stock. The initial holdings are such that a trade occurs at time 0.

2.2 Asset prices

According to Amihud and Mendelson (1986a, Page 228), the price of a security in the presence of transactions costs is equal to the present value of the dividends to be paid on that security minus the present value of transactions costs subsequently to be paid by someone currently holding that security. A similar

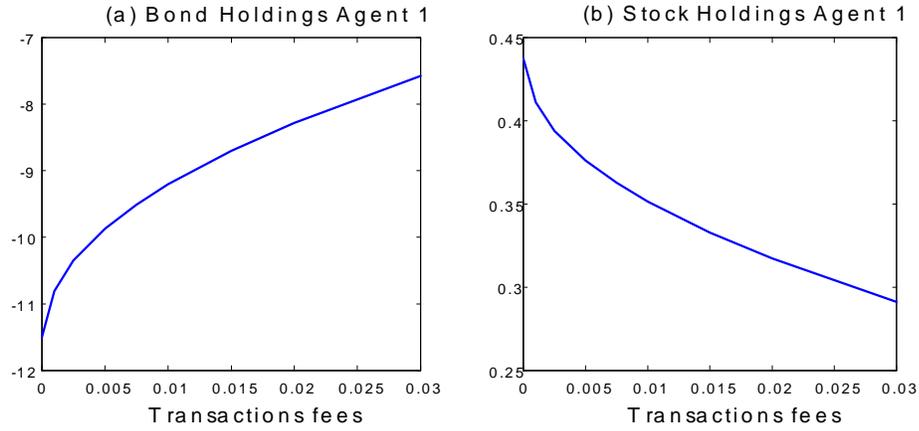


Figure 2: **Optimal “exiting” holdings θ of the securities.** Optimal bond and stock holdings of the first agent for different levels of transactions fees, in the range from 0% to 3%. All parameters and variables are set at their benchmark values indicated in Table 1 (entering holdings $(0, 0)$).

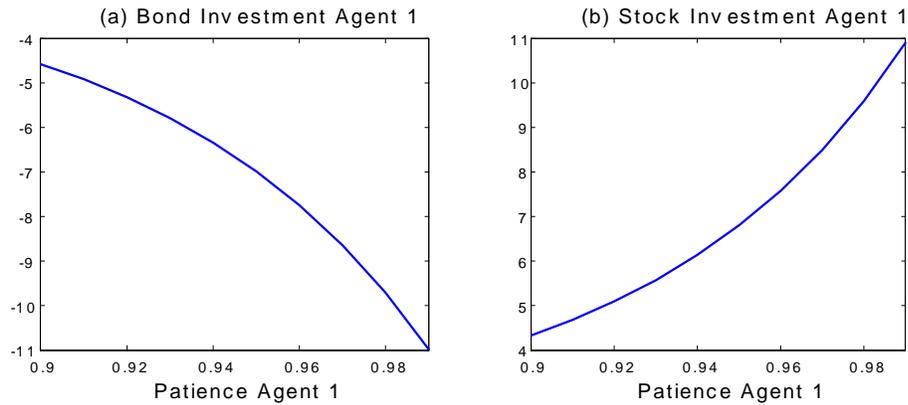


Figure 3: **Clientele effect.** Optimal bond and stock holdings of the first agent for different levels of patience, in the range from 0.95 to 0.99. All other parameters are as described in Table 1. Especially, transactions fees equal 1% of the value of shares traded.

conclusion was reached by Vayanos (1998, Page 18, Equation (31)) and Vayanos and Vila (1999, Page 519, Equation (5.12)).

There are many differences between our setting and the setting of Amihud and Mendelson. They consider a large collection of risk-neutral investors each of whom faces different transactions costs and are forced to trade. We consider two investors who are risk averse, face identical trading conditions and trade optimally. Nonetheless, their statement is an appealing conjecture to be investigated using our model.

Recall from Equation (8) that the securities' ticker prices $S_{t,i}$ are:

$$\begin{aligned} S_{t,i} &= \mathbb{E}_t \left[\frac{\phi_{l,t+1}}{R_{l,t,i}\phi_{l,t}} \times (\delta_{t+1,i} + R_{l,t+1,i} \times S_{t+1,i}) \right]; \\ S_{T,i} &= 0 \end{aligned}$$

where the terms $R_{l,t,i}$ ($1 - \varepsilon_{i,t} \leq R_{l,t,i} \leq 1 + \lambda_{i,t}$) capture the effect of current and anticipated trading fees.

We now present two comparisons. First, we compare equilibrium prices to the present value of dividends on security i calculated at the Investor l 's equilibrium state prices *under transactions fees*. We denote this private valuation $\hat{S}_{t,i,l}$:

Definition 2

$$\hat{S}_{t,i,l} \triangleq \frac{1}{\phi_{l,t}} \sum_{j=u,d} \pi_{t,t+1,j} \phi_{l,t+1,j} \times (\delta_{t+1,i,j} + \hat{S}_{t+1,i,j}); \hat{S}_{T,i} = 0$$

We show that:

Proposition 3

$$R_{l,t,i} \times S_{t,i} = \hat{S}_{l,t,i} \tag{9}$$

Proof. In Appendix D ■

which means that the ticker prices of securities can at most differ from the private valuation of their dividends as seen by Investor l by the amount of the transactions fees incurred or imputed by Investor l at the current date only. Figure 4, panel (b) plots the ticker price and the private valuation of dividends for different values of transactions fees, thus illustrating the decomposition of Equation (9). For instance, for transactions fees of 3%, the price difference is in the range $[-3\%, +3\%]$ of endowment, where we achieve the boundaries of this range when the system hits the boundaries of the trade region. Within the no-trade region, it is somewhere within the range. Second, we compare equilibrium asset prices that prevail in the presence of transactions fees to those that would prevail in a frictionless economy, based, that is, on state prices that would obtain *under zero transactions fees*. Denoting all quantities in the zero-transactions fees economy with an asterisk *, and defining:

$$\Delta\phi_{l,t} \triangleq \frac{\phi_{l,t}}{\phi_{l,t-1}} - \frac{\phi_{l,t}^*}{\phi_{l,t-1}^*}$$

we show that:

Proposition 4

$$R_{l,t,i} \times S_{t,i} = S_{t,i}^* + \mathbb{E}_t \left[\sum_{\tau=t+1}^T \frac{\phi_{l,\tau-1}}{\phi_{l,t}} \cdot \Delta\phi_{l,\tau} \cdot (\delta_\tau + S_\tau^*) \right] \quad (10)$$

Proof. In Appendix E ■

That is, the two asset prices differ by two components: (i) the current shadow price $R_{l,t,i}$, acting as a factor, of which we know that it is at most as big as the one-way transactions fees, (ii) the present value of all future price differences arising from the change in state prices and consumption induced by the presence of transactions fees.

While the ticker price S and the present value of dividends \hat{S} differ from each other at most by one round of transactions fees, both of them are reduced by the presence of transactions fees because, over some range, the state prices ϕ are lower with transactions fees than without them. Panel (c) of the same figure illustrates the decomposition of Equation (10).

Since transactions fees are paid in a reciprocal fashion, the reason for the drop is not that the investors incur large amounts of fees in the future but that they do not hold the optimal frictionless holdings and, therefore, also have consumption schemes that differ from those that would be optimal in the absence of transactions fees. The differences in consumption schemes then influence the future state prices and accordingly the present values of dividends.

Because the affected state prices are applied by investors to all securities, the change in the state prices is also reflected in the one-period bond price which varies (non monotonically) as we vary the transactions fees applied to equity, as is illustrated in panel (a) of figure 4.²⁷

3 Time paths of prices and holdings

We now study the behavior of the equilibrium over time and the transactions that take place. Figure 5 displays a simulated sample path illustrating how our financial market with transactions fees operates over time. In an attempt to remove the effects of the finite horizon on trade decisions, we only display the first 25 periods, although the economy runs for 50 periods ($T = 50$).²⁸

Panel (a) shows a sample path of: (i) stock holdings as they would be in a zero-transaction fee economy, (ii) the actual stock-holdings with a 1% transaction fee and (iii) the boundaries of the no-trade zone, which fluctuate over time. The boundaries fluctuate very much in parallel with the optimal frictionless holdings, allowing a tunnel of deviations on each side. Within that tunnel, the actual holdings move up or down whenever they are pushed up or down by

²⁷Vayanos (1998) had even noted that prices can be increased by the presence of transactions costs.

²⁸If the equilibrium of this economy had been a stationary one, it would have been useful to introduce also a number of “run-in” periods, in an attempt to render the statistical results of this section independent of initial conditions. But, with investors of different risk aversions, equilibrium is rarely stationary (see Dumas (1989)).

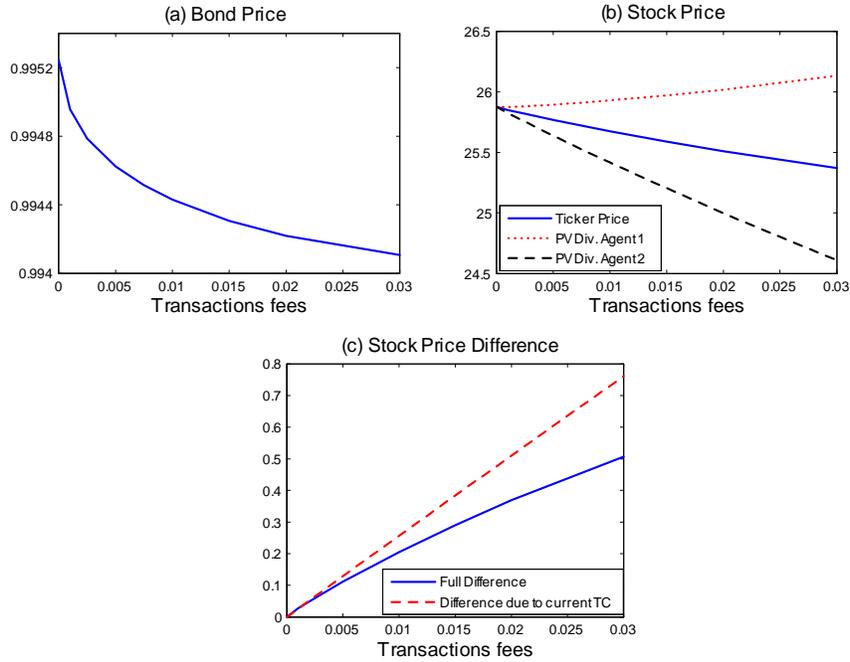


Figure 4: **Initial asset prices.** Panel (a) shows the initial period's bond price for different levels of transactions fees in the range from 0% to 3%. All parameters and variables are set at their benchmark values indicated in Table 1 (entering holdings $(0,0)$ for agent 1). Panel (b) shows the initial period's stock price and the two agents' present values of dividends $\hat{S}_{t,i,l}$ for different levels of transactions fees. Panel (c) shows the difference between the initial stock price in an economy with transactions fees and the stock price in economies without transactions fees. In addition, we show the component of the stock price difference that is due to the current amount of transactions fees.

the movement of the boundaries, with a view to reduce transactions fees and making sure that no wasteful round trip ever occurs. These three paths clarify the logic behind the actual holdings.

Panel (b) shows the stock ticker price (expressed in units of the consumption good), with transaction dates highlighted by a circle. While the ticker price forms a stochastic process with realizations at each point in time, transactions prices materialize as a “point process” with realizations at random times only. Panel (c) displays the difference between individual private valuations (i.e., present values of dividends) and ticker price divided by ticker price, as in decomposition (9). The ticker price is thus seen as an average of the two private valuations. When the two valuations differ by more than the sum of the one-way transaction fees for the two investors, a transaction takes place. As figure 7, panel (a) below further illustrates, agents trade more often after an up-move than after a down-move. The direction of the trade depends, of course, on the sign of the difference between private valuations. The increments in the private valuations of Investor 1 are more highly correlated with the increments in the ticker price than those of Investor 2. In fact, Investor 2 does not buy on an up move in the ticker price. In our benchmark example, Investor 1 has a lower risk aversion. Although ours remains a Walrasian market and not a dealer market, Investor 1 is closer to the proverbial “market maker” of the Microstructure literature, who is traditionally assumed to be risk neutral, and Investor 2 may be viewed as a “customer”. If we wanted to push the analogy further, we could define the “bid” and the “ask” prices as being equal to Investor 1’s private valuation plus and minus transactions fees and we would call a purchase by Investor 2 a “buy”.²⁹

Panel (d) of the figure shows the fluctuations of Amihud’s LIQ measure, which is defined below. It will be useful to us later on.

Finally, panels (e) and (f) illustrate decomposition (10) over time. Deviations are here expressed relative to the price that would prevail if transactions fees were zero. For example, for the bond, the quantity is: $\frac{S_{1,t}^* - S_{1,t}}{S_{1,t}^*}$ where $S_{1,t}^*$ denotes the price in a zero-transactions fee economy. Panel (f) shows along the same path, again in relative terms, the components of the difference, as seen by Investor 1, between the stock price in a frictionless economy and in an economy with transactions fees, the two components reflecting the current amount of (shadow or actual) transactions fees and the future difference in pricing (state prices) respectively.

We now demonstrate some properties of the sample paths. We first investigate univariate properties of trades on the one hand and of asset price increments on the other. Then we investigate bivariate properties of trades and price changes.

²⁹The pattern is reminiscent of Lee and Ready (1991) but would be opposite to their rule. When, in empirical work, the direction of trade is not observed, they recommend to classify the transaction as a buy (by the customer) if it occurs on an “uptick”.

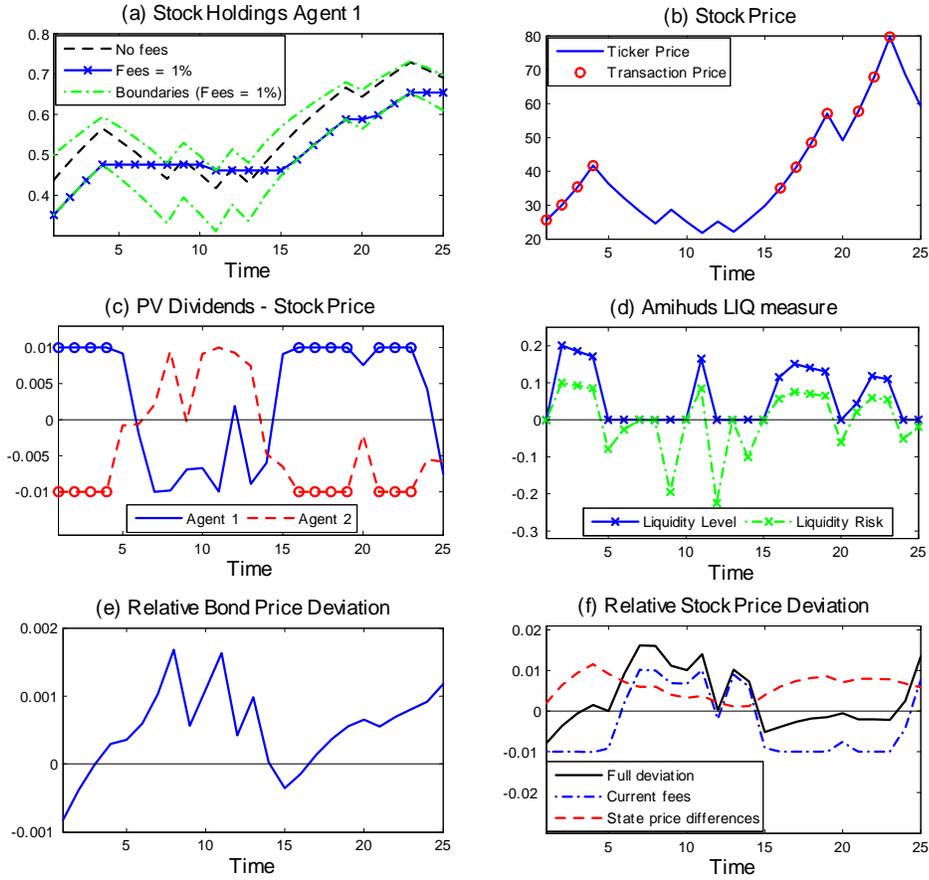


Figure 5: **Sample time paths of stock holdings, the stock price and the difference between the stock price and each investor’s value of the present value of dividends.** Panel (a) shows stock holdings of the first agent along the paths for zero and 1% transactions fees. All parameters and variables are set at their benchmark values indicated in Table 1 (time-0 holdings (0,0) for agent 1). Panel (b) shows the stock ticker price along the sample path for 1% transactions fees. Panel (c) shows the present values of future dividends from the points of view of the two agents along the same path. Transactions are highlighted by a circle. Panel (d) shows Amihud’s LIQ measure along the same path as well as its unanticipated or permanent component (marked Liquidity Risk). Panel (e) shows the relative deviation between the asset price in a zero-transactions fees economy and an economy with transactions fees along the same path. Panel (f) shows the components of the difference between the stock price in a zero-transactions fees economy and an economy with transactions fees, along the same path.

3.1 Trades over time

We examine the trading volume and the waiting times between trades. The trading volume is defined as the sum of the absolute values of changes in θ_2 (shares of the stock) over the first 25 periods of the tree. The average trading volume is shown in figure 6, panel (a); as one would expect, it decreases with transactions fees. Correspondingly, the average waiting (panel (b)) between trades rise. We also show in panel (c) the volatility of the waiting time, which is a first measure of the (endogenous) liquidity risk that the investor has to bear because he/she operates in a market with friction. We examine in section 4 below how this risk is priced.

The Microstructure literature has established that trades are autocorrelated and the order flow is predictable (Hasbrouck (1991a, 1991b) and Foster *et al.* (1993)). Looking at the time path in panel (a) of figure 5, we have already pointed out that the investors smooth their trades over time in order to keep transactions fees low. We investigate the matter more systematically in figure 7, panel (b), which displays the average of 20000 simulations. The microstructure literature usually ascribes the autocorrelation of trades to a trader’s desire to avoid price impact by, for instance, breaking up large trades into smaller ones, a form of behavior known as “order fragmentation”. But, here we see that the desire simply to avoid wasteful round trips, in a Walrasian market, also leads to a strong autocorrelation.

3.2 Prices over time

We are interested in determining in which way, as one decreases transactions fees, the point process of transactions prices approaches the process that would prevail in the absence of transactions fees, which in the limit of continuous time would be a continuous-path process. As is well-known, the Brownian motion is characterized by the fact that its total variation, calculated over a finite period of time, is infinite while its quadratic variation is finite.³⁰ The transactions prices are like the result of infrequent sampling of ticker prices, with frequency of sampling rising as transactions fees go down. If ticker prices behave roughly like random walks, the total absolute variation should rise with the frequency of sampling and the quadratic variation should stay about the same. This should be approximately true for any fixed, extended time period. It should not be exactly true because here the sampling (the occurrence of transactions) is not independent of the price movements.

We generate many simulated paths of the stock price for zero transactions fees and calculate average (across paths) total variation and quadratic variation over the first 25 periods. Then we generate the same paths of transactions prices and holdings with transactions fees increasing to 3% and we calculate again average (across paths and dates) total variation and quadratic variation. These are plotted against transactions fees in figure 8.

³⁰Total variation is the sum of the absolute values of the segments making up a path or connecting the dots, whereas quadratic variation is the sum of their squares.

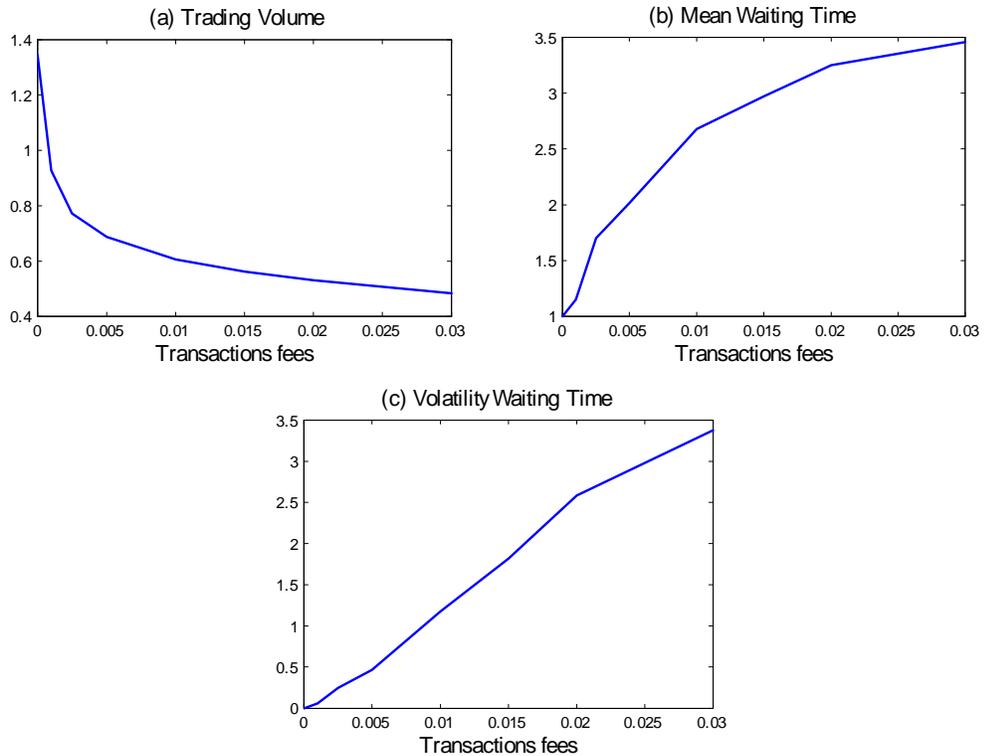


Figure 6: **Trading volume and waiting time against transactions fees.** Panels (a) and (b) show the average (across paths and dates) stock trading volume/year and the average waiting time between trades (measured in years) respectively, up to period 25 for different levels of transactions fees, in the range from 0% to 3%. Panel (c) shows the standard deviation of the waiting time calculated the same way. All parameters and variables are set at their benchmark values indicated in Table 1. We use 20,000 simulations along the tree. Panel (c) shows the standard deviation of waiting time computed the same way.

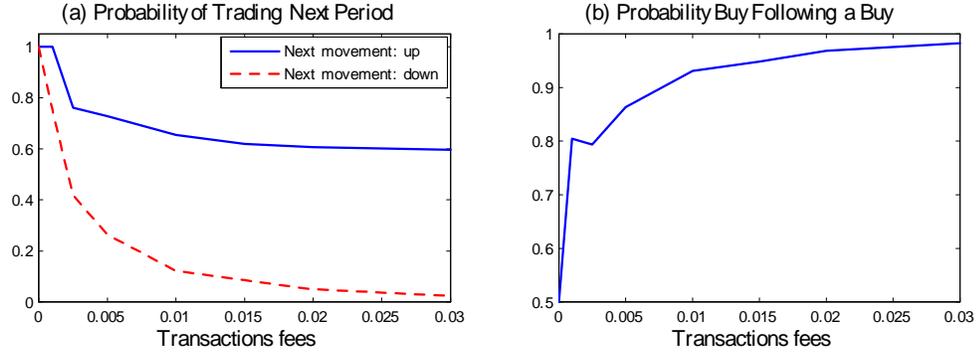


Figure 7: **Trading patterns.** Panel (a), the frequency (across paths and dates) of a buy transaction coinciding with an up or down move in price or endowment. Panel (b), *Serial dependence of trades*: the frequency (across paths and dates) of a buy transaction following a previous buy transaction. All parameters and variables are set at their benchmark values indicated in table 1. We use 20,000 simulations along the tree. In each, the first 25 periods only are used.

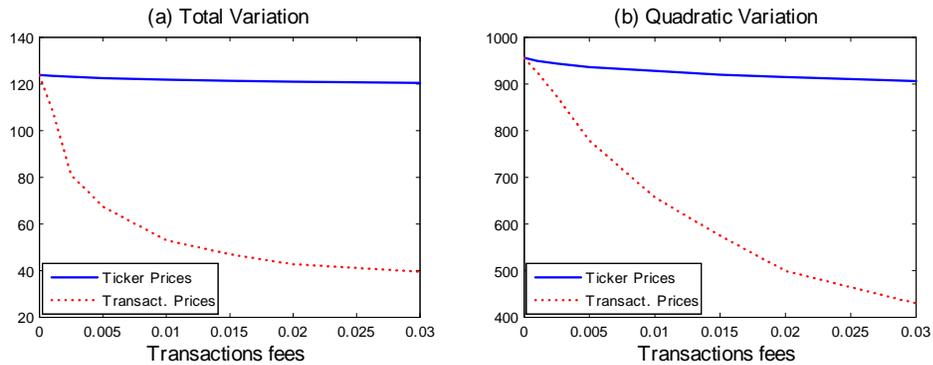


Figure 8: **Total and quadratic variations of stock price depending on transactions fees.** Panel (a) shows the total variation (defined in footnote 30) up to period 25 for different levels of transactions fees, in the range from 0% to 3%. All parameters and variables are set at their benchmark values indicated in table 1. We use 20,000 simulations along the tree. Panel (b) shows the quadratic variation computed the same way.

The total variation of the ticker price is practically invariant to transactions fees. It is finite because this is a finite tree but, if one took the limit of continuous time, it would be infinite, as is the case for Brownian motions. When reducing transactions fees, transactions become more and more frequent and the total variation of the transactions prices rises rapidly to approach the total variation of the ticker price but then is capped by it. If one took the limit to continuous time, it would also approach infinity.

As can be expected from the reasoning above, the quadratic variation of the ticker price is approximately constant (note the vertical scale). The quadratic variation of the transactions prices rises modestly.

The serial dependence of prices plays a crucial role in the empirical Microstructure literature. It serves to decompose real frictions from information frictions. Roll (1984) originally proposed to use the “bid-ask bounce” to measure the effective spread, an approach which was later generalized by Stoll (1989). As Stoll (2000) explains, “Price changes associated with order processing, market power, and inventory are transitory. Prices ‘bounce back’ from the bid to the ask (or from the ask to the bid) to yield a profit to the supplier of immediacy. Price changes associated with adverse information are permanent adjustments in the equilibrium price.” The presumption is that, in response to random customer arrivals (Stoll (2000)), “bid and ask prices are lowered after a dealer purchase in order to induce dealer sales and inhibit additional dealer purchases, and bid and ask prices are raised after a dealer sale in order to induce dealer purchases and inhibit dealer sales.”

In our model, there is information coming in (but no information asymmetry).³¹ Figure 9 displays the frequency of an up move in price being followed by an up move in price. Because a move up in the ticker price can only be associated with a move up in the endowment/dividend, and because we have assumed endowment/dividend up and down moves that are IID, that frequency is tautologically equal to 1/2 when transactions fees are zero and when considering the ticker price. When considering transactions prices, however, the frequency quickly rises with transactions fees, to above 0.8. Far from displaying a bid-ask bounce, prices display momentum. Evidently, the absence of a bounce, if observed by an econometrician, should not be regarded as evidence of absence of real frictions.

How can we account for the differing conclusions? Even though our market is Walrasian, we could define a concept of bid and ask as being the prices inclusive of transactions fees at which a person would be willing to buy or sell. More precisely, the bid price of a person could be defined as being equal to the person’s private valuation of dividends minus the transactions fees to be paid in case the person buys. Had we done that, Figure 5, panel (c) above implies that bid and ask prices would have moved as much as transactions prices and would also have exhibited momentum. In our model of optimal customer arrival, there can be no buy or sale order coming to the market place unless some information about

³¹The empirical Micro literature seems to use the same “martingale” specification for both without distinction.

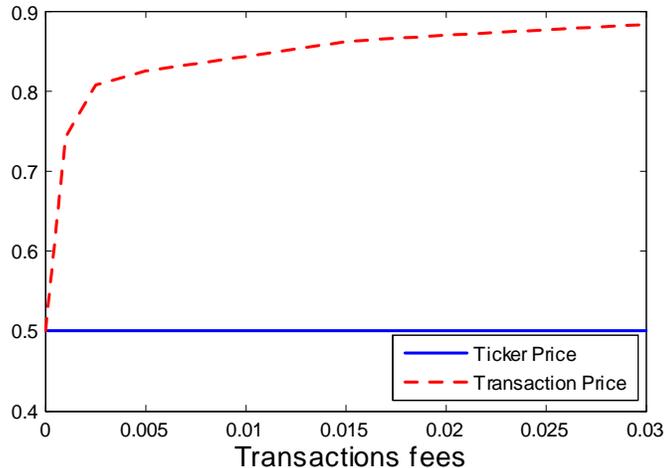


Figure 9: **Serial dependence of price changes:** the frequency (across paths and dates) of a price increase following a previous price increase. All parameters and variables are set at their benchmark values indicated in table 1. We use 20,000 simulations along the tree. In each, the first 25 periods only are used.

the fundamental has also arrived. It is not the case that customers act randomly and dealers accomodate them temporarily in an optimal fashion. Everyone here acts optimally.

3.3 Joint behavior of transactions prices and trades

We now explore the joint behavior of prices and transactions, which is a favorite topic of the empirical Microstructure literature, aiming to measure the “price impact” of trades, when customer trades arrive randomly.³² Much of the literature relates price impacts to traders’ hedging and speculative motives (the latter arising from the presence of informed traders) and possibly also to their strategic behavior. We want to determine whether the empirical phenomena that have been unearthed could also be explained, in a more mundane fashion, by transactions fees and the heterogeneity of tastes of the investor population, when customers’ orders do not arrive randomly but are, instead, those of intertemporally optimizing agents seeking to economize on the cost of transacting. Our findings are not meant to oppose the informed-trading interpretation offered by the Microstructure literature. Indeed, following Glosten and Milgrom (1985)’s dealership theory, we know that bid-ask spreads, which in the real world are a

³²See the surveys by Biais *et al.* (2005), Amihud *et al.* (2005), the monographs by Hasbrouck (2007) and by de Jong and Rindi (2009), and the works of Roll (1984), Campbell *et al.* (1993), Llorente *et al.* (2002) and Sadka (2006).

large component of transactions costs, arise from informed trading.

We investigate the relationship between transactions fees and three popular measures of price impact. A first popular measure is the ILLIQ measure of Amihud (2002). We interpret it as being equal to the average over time of the absolute values of the change in the ticker price divided by the contemporaneous absolute volume of trade. However, in most sample paths, there are nodes with zero trades. We prefer, therefore, to compute a LIQ measure equal to the average of volume of trade over the absolute price change. In figure 5, panel (d), we have exhibited the fluctuations of LIQ along a sample path, illustrating the way it varies with the volume of trade and the shadow prices of investors. But, as has been emphasized by Acharya and Pedersen (2005), Pástor and Stambaugh (2003) and Sadka (2006), the risk borne by an investor is not given by the totality of the fluctuations of LIQ but by the innovations in the process. For that reason, we have also shown in the same graph the sample path of the innovation in the LIQ process (defined as the realized value of LIQ minus the conditional expectation of LIQ computed from the model). These are used below (in Section 4) in our discussion of liquidity pricing.

Figure 10, panels (a) and (b) show how LIQ, which is commonly used to estimate effective trading costs, is, on average, related to the given one-way transactions fees of our model and to the average effective fee for the investors captured by the average ratio of their shadow costs R . We compute, at each node where there is a trade, the price change since the last trade as well as the purchase or sale at that node and collect the ratios of those. We then compute the average. Panel (a) shows that this average LIQ is monotonically related to one-way trading fees. For panel (b) we have, in addition, calculated an average across paths and dates of the shadow costs ratio; the panel again shows a monotonic relationship.

More formal methods to measure price impact are based on reduced forms of theoretical Microstructure models. Some are motivated by the desire to capture informed trading (Roll (1984), extended by Glosten and Harris (1988)). Others (Ho and Macris (1984)) are motivated by inventory considerations. Madhavan and Smidt (1991) run a regression which is meant to capture both effects. We implement their idea in the following way. At each node where there is a trade, we collect the price change since the last trade, the signed amount of purchase or sale by Investor 2 and the current equity holding of Investor 1. We then regress, across nodes of various times, the price change on these two variables. The responsiveness of price to order quantity, often referred to as Kyle's λ , is displayed in figure 10, panel (c), against transactions fees. It is also mostly rising with transactions fees.

We also calculate average PIN (Probability of Informed Trading). We implemented the procedure described in Easley *et al.* (2002). The parameters of the underlying sequential trade model are estimated by Maximum Likelihood. In empirical studies what is typically done is the following: given data for a specific horizon, say, a year, one counts the buy and sells on a pre-defined unit of time, e.g., day or week, so that one finally has a vector containing the number of buy and sell trades for each unit in the year. Here, we counted for each simulation

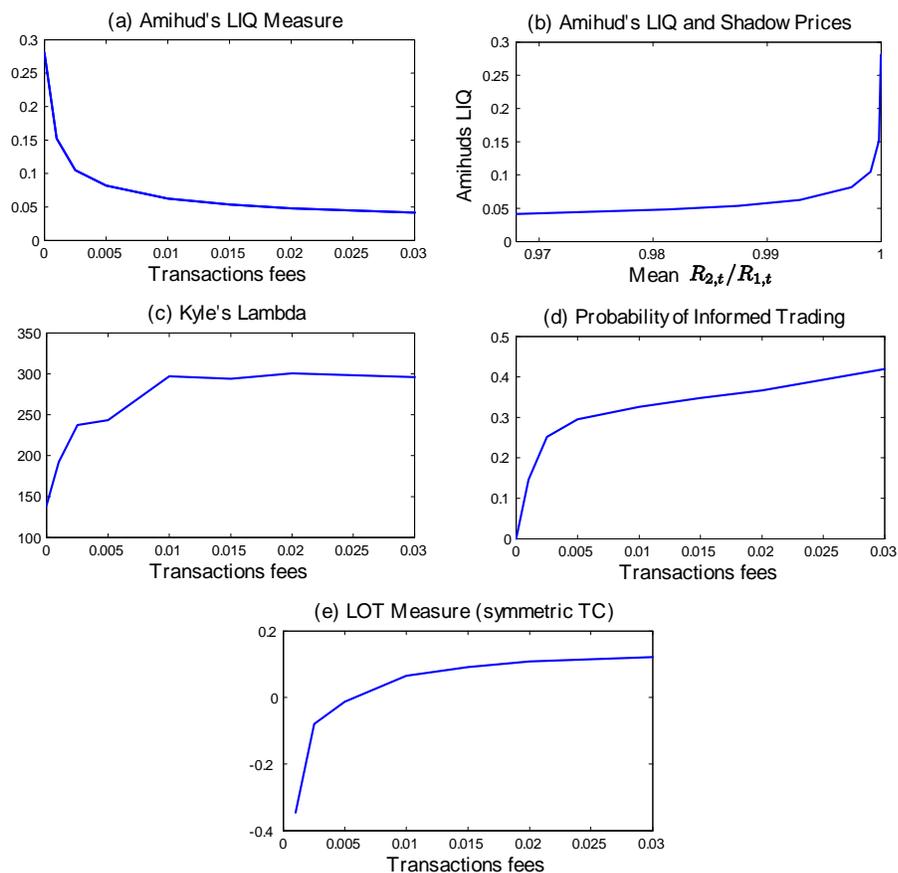


Figure 10: **Liquidity variables.** The first two panels show the average across paths and dates of Amihud's LIQ measure, computed using simulated results up to period 25 for different levels of transactions fees, in the range from 0% to 3% (Panel (a)) and against the average shadow price ratio (Panel (b)). All parameters and variables are set at their benchmark values indicated in table 1. We use 20,000 simulations along the tree. Panel (c) shows Kyle's lambda and computed in the same way using Madhavan-Smidt regression. Panel (d) shows the average PIN measure. Panel (e) shows the LOT/FHT measure based on frequency of no trade.

the number of buys and sells and thus arrived at a vector of numbers of buy and sell trades for each simulation path.³³ Thus, we can simply estimate the parameters of the model for each level of transaction fee and compute the PIN measure. Panel (d) displays the results. For an economy without transactions fees the PIN measure is zero. If we increase transaction fees, the PIN measure quickly increases. The level of the PIN measure is also quite high.³⁴

Finally, we calculate LOT (after Lesmond *et al.* (1999)) who suggest a measure of transaction costs that does not depend on information about quotes or the order book. Instead, LOT is calculated from daily returns. It uses the frequency of zero returns to estimate an implicit trading cost.³⁵ Here, we use, instead, the frequency of no trade and implement the measure as described in Fong *et al.* (2010) under the acronym “FHT”. That measure assumes symmetric transactions costs and applies to a single stock independent of a market return.³⁶

Overall, figure 10 demonstrates that commonly used measures of price impact are not necessarily measures of the degree of informed trading present in the marketplace but could also be the mechanical result of intertemporal optimization in the presence of market frictions.

An additional aspect of the joint behavior of prices and volume has been pointed out to us by our colleague Xi Dong of INSEAD. Although we could not find any academic work having specifically documented that phenomenon, it is plain on any stock-price path diagram obtainable, for instance, from Yahoo.com, that very large price drops are immediately followed by an increase in volume, as though the market was waiting for that drop to occur before they would trade again. In figure 11, we have plotted the average volume occurring after 1, 2, 3 or 4 consecutive down movements in price relative to the unconditional average volume. That is, we collected all dates from our simulations with 1, 2, 3 or 4 consecutive down-movements in the ticker price - to pick up negative market movements. We computed the average trading volume on these dates as well as the overall average trading volume (on all dates). We then computed the ratio. That is, values greater than 1 mean that liquidity is higher after down moves compared to overall liquidity. The effect is strikingly present.³⁷

³³This means that we assimilate a simulation run with one record in the empirical data and instead of having several successive records we have several simulation runs.

³⁴For example, in Easley *et al.* (2002), the mean of the PIN measure is 0.2 with a typical 95% percentile of 0.3.

³⁵The LOT cost is an estimate of the implicit cost required for a stock’s price not to move when the market as a whole moves. To see the intuition behind this measure, consider the simple market-factor model $R_{i,t} = a_i + b_i R_{m,t} + \varepsilon_{i,t}$, where $R_{i,t}$ is the return on security i at time t , $R_{m,t}$ is the market return at time t , a is a constant term, b is a regression coefficient, and ε is an error term. In this model, for any change in the market return, the return of security i should move according to the market-factor model. If it does not, it could be that the price movement that should have happened is not large enough to cover the costs of trading. Lesmond *et al.* estimate how wide the transaction cost band around the current stock price has to be to explain the occurrence of no price movements (zero returns). The wider this band, the less liquid the security. Lesmond *et al.* show that their transaction cost measure is closely related to the bid-ask spread.

³⁶For low levels of transactions fees, where the agents always trade, the LOT measure is $-\infty$, accordingly, not shown in the picture.

³⁷It can be argued that the dollar amount of transactions fees is lower when the price of the

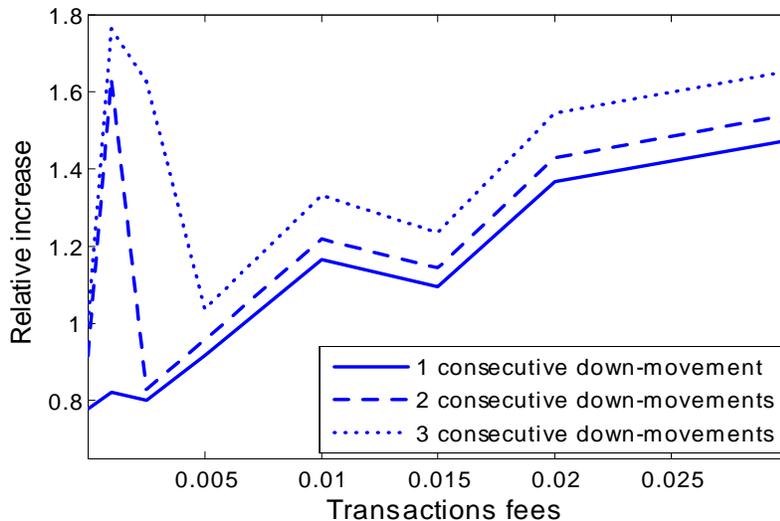


Figure 11: **Volume post price drop:** average over all 25000 simulated paths and 25 dates of volume after 1, 2, 3 or 4 consecutive ticker price drops, compared to regular volume, for various levels of transactions fees.

4 The pricing of liquidity and of liquidity risk

Based on a pure portfolio-choice reasoning, Constantinides (1986) argued that transactions costs make little difference to risk premia in the financial market. Liu and Lowenstein (2002) and Delgado, Dumas and Puopolo (2012), still on the basis of portfolio choice alone, challenge that view by pointing out that the conclusion of Constantinides holds only when rates of return are identically, independently distributed (IID) over time. We go one step further than these authors, in that we now get the deviations in a full general-equilibrium model, when endowments are IID but returns themselves are not, and investors must also face the uncertainty about the dates at which they can trade.

4.1 Deviations from the classic consumption CAPM under transactions fees

In our equilibrium, the capital-asset pricing model is Equation (8) above. The dual variables R (in addition to the intertemporal marginal rates of substitution ϕ) drive the prices of assets that are subject to transactions fees, as do, in the

share is lower, which could be an incentive to wait for a low price before one trades. In order to address that issue, we have redone the calculation with fees that were proportional to the number of shares traded, as opposed to their value. The effect was still present.

“LAPM” of Holmström and Tirole (2001), the shadow prices of the liquidity constraints.³⁸

It can be rewritten as:³⁹

$$\begin{aligned} \mathbb{E}_t [r_{t+1,i}] &= r_{t+1,1} - cov_t \left(r_{t+1,i}, \frac{\phi_{l,t+1}}{\mathbb{E}_t [\phi_{l,t+1}]} \right) \\ &+ \mathbb{E}_t [\tau_{l,t+1,i}] + cov_t \left(\tau_{l,t+1,i}, \frac{\phi_{l,t+1}}{\mathbb{E}_t [\phi_{l,t+1}]} \right); i \neq 1 \end{aligned} \quad (11)$$

where:

$$r_{t+1,i,j} \triangleq \frac{\delta_{t+1,i,j} + S_{t+1,i,j}}{S_{t,i}}$$

is the gross rate of return on asset i ($r_{t+1,1}$ being the gross rate of interest from time t to time $t + 1$) and:

$$\tau_{l,t+1,i,j} \triangleq (1 - R_{l,t+1,i,j}) \times \frac{S_{t+1,i,j}}{S_{t,i}} - (1 - R_{l,t,i}) \times r_{t+1,1}$$

both referring to state of nature j of time $t + 1$.

Equation (11) constitutes a decomposition exercise similar to that performed by Acharya and Pedersen (2005). Here, however, the terms have received a formulation that is explicitly related to the optimal decision of investors to trade or not to trade and they have explicit dynamics. The first part of the expression is exactly the CCAPM expression of a frictionless market. The remainder is a deviation from the CCAPM, which we can split into the following parts: $\mathbb{E}_t [\tau_{l,t+1,i}]$ is the expected change in shadow transactions costs applying to the future stock price relative to the current shadow cost adjusted for the time value of money, and $cov \left(\tau_{l,t+1,i}, \frac{\phi_{l,t+1}}{\mathbb{E}_t [\phi_{l,t+1}]} \right)$ which is a liquidity risk premium.

We now define deviations from the classic consumption CAPM that occur in our equilibrium with fees as:

Definition 5

$$CCAPM \text{ deviation} \triangleq \mathbb{E}_t [\tau_{l,t+1,i}] + cov_t \left(\tau_{l,t+1,i}, \frac{\phi_{l,t+1}}{\mathbb{E}_t [\phi_{l,t+1}]} \right) \quad (12)$$

The deviation from the classic CCAPM being the sum of expected change in transactions fees (or expected change in liquidity) and a premium for the liquidity risk created by transactions fees, figure 12, panels (a) and (b) shows these two components from the classic consumption CAPM from the standpoint of individual investors. These are computed using simulated returns up to period 25 for different levels of transactions fees, in the range from 0% to 3%.

³⁸Holmstrom and Tirole (2001) assume that their liquidity constraint is always binding. Here, the inequality constraints (5) bind whenever it is optimal for them to do so.

³⁹Recall that the security numbered $i = 2$ is equity and the security numbered $i = 1$ is the short-term bond.

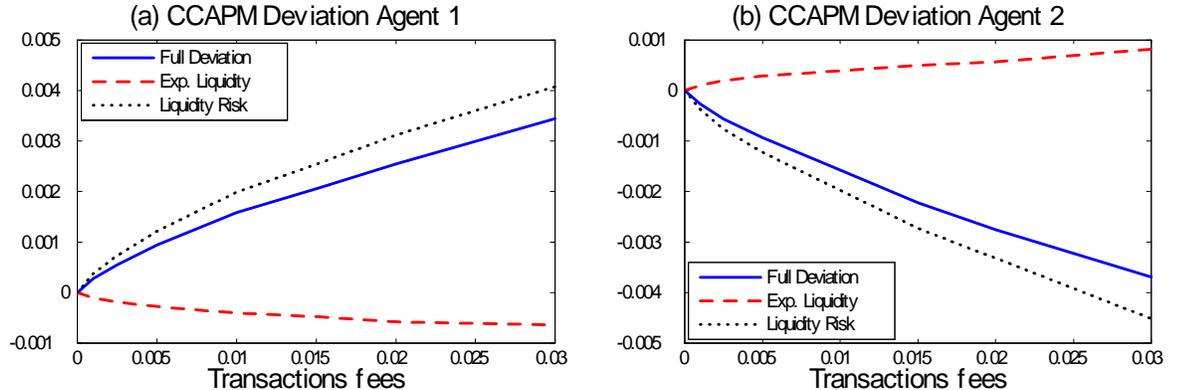


Figure 12: **CCAPM deviations.** The panels show the average (across paths and dates) of deviations from the classic consumption CAPM, computed using results up to period 25 for different levels of transactions fees, in the range from 0% to 3%. The “expected liquidity” component is $\mathbb{E}_t[\tau_{l,t+1,i}]$; the “liquidity risk” premium component is $cov_t\left[\tau_{l,t+1,i}, \frac{\phi_{l,t+1}}{\mathbb{E}_t[\phi_{l,t+1}]}\right]$. All parameters and variables are set at their benchmark values indicated in table 1.

As expected, the absolute CCAPM deviation is increasing in transactions fees. The CCAPM deviation is positive for the first investor, i.e., the less risk-averse investor demands a higher expected return in an economy with transactions fees whereas the more risk-averse Investor 2 demands a lower expected return. This is due to the fact that the covariance between the first investor’s pricing kernel and the τ return, i.e., liquidity risk, is positive.

In Amihud and Mendelson (1986a), it was explained that the total premium should be concave in the size of transactions fees. For that reason, Amihud and Mendelson (1986b) fitted the cross section of equity portfolio returns to the log of the bid-ask spread of the previous period and found a highly significant relationship. Our figure does exhibit that concavity property.⁴⁰

For high transactions fees of 3%, the total deviation reaches 40bp, which is less than the transaction fees themselves, measured as a percentage of the value of each trade. That deviation is much too small to be able to account for the several percentage points of returns that empirical researchers commonly attribute to liquidity premia.⁴¹ But it does show that trading frictions can play a role when we try to explain empirical deviations from classic asset pricing

⁴⁰See also Figure 3.1 in Amihud et al. (2005). The analogy between what we do and what they do is not perfect as they display a cross-section of firms affected differently by transactions costs and we display a single premium for different levels of transactions fees. But the underlying rationale is identical.

⁴¹Furthermore, the terms being of opposite signs for the two investors, their values would be even smaller in any CAPM that would be somehow aggregated across investors.

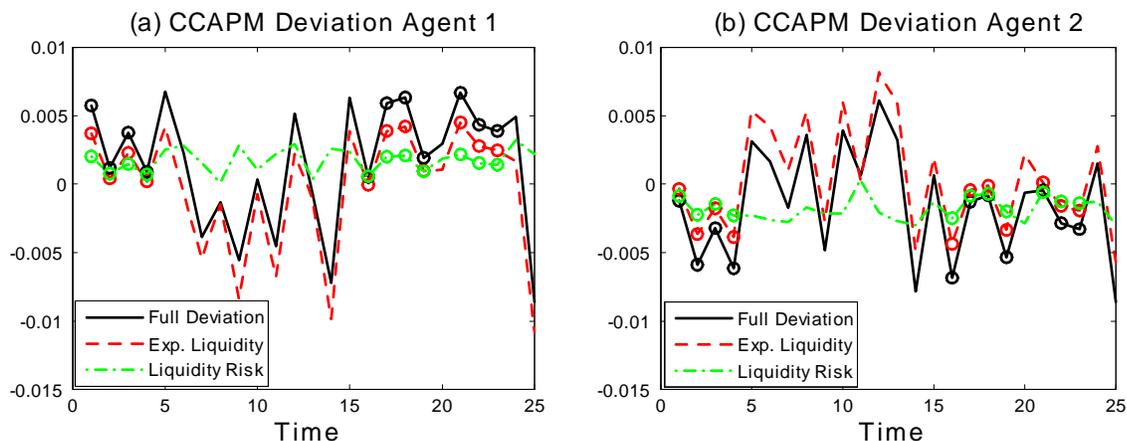


Figure 13: **Fluctuations of the components of the deviations from the classic CCAPM**, along the same sample path as in figure 5.

models.

An important benefit of our model, in which the liquidity variable is endogenized, is that we can study the variation of each of the terms over time. Whereas Figure 12 shows that the unconditional average value of the CAPM deviation is mostly due to the liquidity risk premium, figure 13 reveals that the expected liquidity term is mostly responsible for the fluctuations over time of the CCAPM deviation, the liquidity risk premium being approximately constant at 20bp when the transaction fee is 1%. This theoretical contrast between the conditional and the unconditional pictures should provide guidance for empirical researchers working on currently trying on a number of illiquid markets and trying to decide which of the two terms is more important. In a recent contribution, Bongaerts, De Jong and Driessen (2012), for instance, study very thoroughly the effect of liquidity on corporate-bond expected returns and “find a strong effect of expected liquidity and equity market liquidity risk on expected corporate bond returns, while there is little evidence that corporate bond liquidity risk exposures explain expected corporate bond returns, even during the recent financial crisis.” The model shows that, here especially, empirical conclusions could vary a lot depending on conditioning.

4.2 Liquidity and asset pricing

In our CAPM (11), the shadow prices are generally not observable and most empiricists would choose to replace them with proxies.

In figure 5, we have displayed a sample path of the time variation of an empirical measure of liquidity and of its unanticipated component. They were seen to fluctuate widely over time. For that reason, liquidity fluctuations, in

addition to current and expected liquidity, have been regarded as a source of risk, and as a risk that receives a price in the market place. In a number of empirical papers,⁴² tests were conducted on a cross-section of monthly portfolio returns, looking at changes in market liquidity as a new risk factor. Brennan and Subrahmanyam (1996) and Brennan, Chordia, Subrahmanyam, and Tong (2012) base their tests on Kyle's lambda as a measure of liquidity. Brennan, Chordia and Subrahmanyam (1998) use volume of trading. Acharya and Pedersen (2005) and Bongaerts, De Jong and Driessen (2012) use LIQ (or ILLIQ) as a liquidity measure.⁴³

Our model, however, says that liquidity risk should be captured by the fluctuations in a combination of shadow prices and stock prices defined above as τ , not by the empirical variable LIQ. We now ask whether the unanticipated component of LIQ is a good proxy for the correct measure.⁴⁴ As far as the liquidity risk premium is concerned, that question is answered in figure 14, which shows the average, across sample paths and dates, of the conditional correlation between the two variables. LIQ seems to capture liquidity risk in a time-series dimension for all but very small values of the transactions fees values. For reasonable values of transactions fees, LIQ, being highly correlated with τ , has the potential to be an adequate proxy in tests of the CAPM.⁴⁵ Other measures of liquidity could be subjected to the same theoretical validation test before being used.

In empirical work, the gross rate of return on a security is commonly computed as $\frac{\delta_{t+1,i,j} + S_{t+1,i,j}}{S_{t,i}}$ between fixed, equally spaced calendar points in time, between which the security is held. However, the concept of holding period is quite arbitrary. Absent transactions fees, since investors are ready to trade at any time, the only holding period that would make sense is one approaching zero. Armed with the current model, we have determined the holding period endogenously in the presence of transactions fees. In a model with more than two agents, holding periods would generally differ across people. With two agents, who can only trade with each other, the holding periods are identical across agents but, between trades, their desires to trade differ. That desire is reflected in the investor-specific shadow prices, which must be taken into account if rates of return continue to be based on fixed, equally spaced points in time. If one wanted to test our CAPM, a better way would be not to use the standard concept of rate of return measured between fixed points in time. Instead, one would use transactions prices only, which do not occur at fixed time intervals, and one would substitute out in the model the values of the prices that are unobserved for lack of transaction.⁴⁶ That, however, is not the way empirical tests have

⁴²Brennan and Subrahmanyam (1996), Brennan, Chordia, and Subrahmanyam (1998), and Easley, Hvidkjaer and O'Hara (1999), Acharya and Pedersen (2005) and Pástor and Stambaugh (2003).

⁴³Pástor and Stambaugh (2003) use a different measure, which we cannot replicate here.

⁴⁴We remind the reader that the separation between anticipated and unanticipated component is made by means of the conditional expected value of LIQ provided by the model.

⁴⁵The behavior is similar for both agents, only with a reversed sign due to the restriction on the R variables.

⁴⁶See the discussion on page 89 of Hasbrouck (2007).

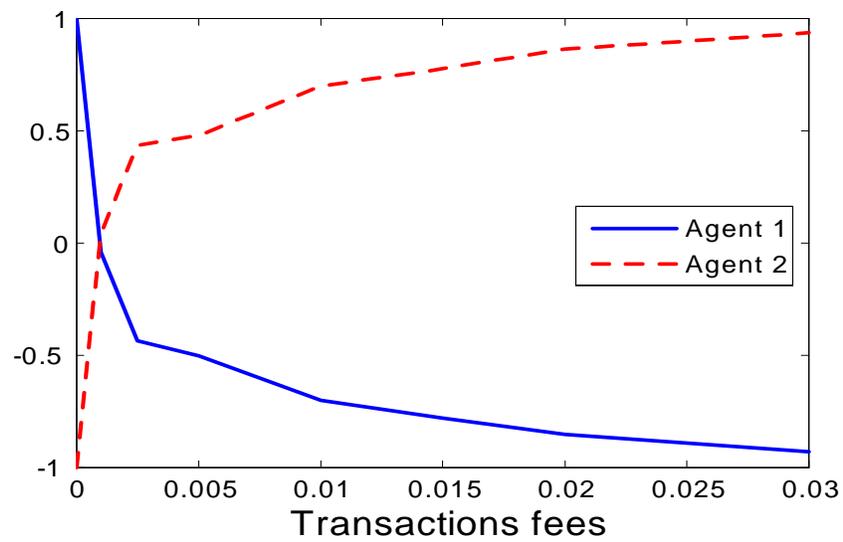


Figure 14: **Average conditional correlation between the LIQ variable and the shadow liquidity variable τ .** The computation uses results up to period 25 for different levels of transactions fees, in the range from 0% to 3%. All parameters and variables are set at their benchmark values indicated in table 1. We use 20,000 simulations along the tree.

been conducted by previous authors. Further work is needed to develop the econometric method.

5 Conclusion

We have developed a new method to compute financial-market equilibria in the presence of proportional transactions fees. For a given rate of transactions fees, our method delivers the optimal, market-clearing moves of each investor and the resulting ticker and transactions prices. In our model, it is not the case that customers act randomly and dealers accommodate them temporarily in an optimal fashion. Here, everyone behaves optimally in reaction to the information they receive.

We have concluded that transactions fees have a strong effect on investors' asset holdings, that deviations in asset prices from a frictionless economy are equal at most to current transactions fees only plus all future state-price differences. We have studied the behavior over time of trades, ticker prices and asset holdings, showing that they can match many of the empirical Microstructure results. We found, however, that transactions prices exhibit momentum and no reversal.

We have presented a transactions-fees adjusted CAPM model, identified the risk factors and displayed their relative sizes and movements over time. We confirmed, however, the view expressed in prior work saying that explicitly observable transactions fees cannot account for the size of what is commonly measured as a liquidity premium. We have commented, in the light of our theoretical model, on the adequacy of extant empirical tests of CAPMs that include a premium for liquidity risk. Shadow prices that properly capture liquidity are generally not observable but our model validates the variables often used in those tests to proxy for time-varying liquidity. Further work is needed to develop the econometric method that would be most powerful given the theoretical equilibrium model.

Future theoretical work should aim to model an equilibrium in which trading would not be Walrasian. In it, the rate of transactions fees would not be a given and investors would submit limit and market orders. The behavior of the limit-order book would be obtained. This would be similar to the work of Parlour (1998), Foucault (1999), Foucault, Kadan and Kandel (2005), Goettler, Parlour and Rajan (2005) and Rosu (2009), except that trades would arrive at the time and in quantities of the investor's choice, and would not be driven by an exogenous process.⁴⁷

⁴⁷Recently, Kühn and Stroh (2010) have used the dual approach to optimize portfolio choice in a limit-order market and may have shown the way to do that.

Appendixes

A Proof of the equation system of Section 1.

The Lagrangian for problem (3) is:

$$\begin{aligned}
\mathcal{L}_l(\{\theta_{l,t-1,i}\}, \cdot, e_{l,t}, t) &= \sup_{c_{l,t}, \{\widehat{\theta}_{l,t,i}, \widehat{\theta}_{l,t,i}\}} \inf_{\phi_{l,t}} u_l(c_{l,t}, t) \\
&+ \sum_{j=u,d} \pi_{t,t+1,j} J_l \left(\left\{ \widehat{\theta}_{l,t,i} + \widehat{\theta}_{l,t,i} - \theta_{l,t-1,i} \right\}, \cdot, e_{l,t+1,j}, t+1 \right) \\
&\quad + \phi_{l,t} \left[e_{l,t} + \sum_{i=1,2} \theta_{l,t-1,i} \delta_{t,i} - c_{l,t} \right. \\
&\quad + \sum_{i=1,2} \left(\widehat{\theta}_{l',t,i} - \theta_{l',t-1,i} \right) S_{t,i} \lambda_{i,t} - \sum_{i=1,2} \left(\widehat{\theta}_{l',t,i} - \theta_{l',t-1,i} \right) S_{t,i} \varepsilon_{i,t} \\
&\quad \left. - \sum_{i=1,2} \left(\widehat{\theta}_{l,t,i} - \theta_{l,t-1,i} \right) S_{t,i} (1 + \lambda_{i,t}) + \sum_{i=1,2} \left(\widehat{\theta}_{l,t,i} - \theta_{l,t-1,i} \right) S_{t,i} (1 - \varepsilon_{i,t}) \right] \\
&\quad + \sum_{i=1,2} \left[\mu_{1,l,t,i} \left(\widehat{\theta}_{l,t,i} - \theta_{l,t-1,i} \right) + \mu_{2,l,t,i} \left(\theta_{l,t-1,i} - \widehat{\theta}_{l,t,i} \right) \right]
\end{aligned}$$

where $\phi_{l,t}$ is obviously the Lagrange multiplier attached to the flow budget constraint (4) and μ_1 and μ_2 are the Lagrange multipliers attached to the inequality constraints (5). The Karush-Kuhn-Tucker first-order conditions are:

$$\begin{aligned}
u'_l(c_{l,t}, t) &= \phi_{l,t} \\
e_{l,t} + \sum_{i=1,2} \theta_{l,t-1,i} \delta_{t,i} - c_{l,t} + \sum_{i=1,2} \left(\widehat{\theta}_{l',t,i} - \theta_{l',t-1,i} \right) S_{t,i} \lambda_{i,t} - \sum_{i=1,2} \left(\widehat{\theta}_{l',t,i} - \theta_{l',t-1,i} \right) S_{t,i} \varepsilon_{i,t} \\
- \sum_{i=1,2} \left(\widehat{\theta}_{l,t,i} - \theta_{l,t-1,i} \right) S_{t,i} (1 + \lambda_{i,t}) + \sum_{i=1,2} \left(\widehat{\theta}_{l,t,i} - \theta_{l,t-1,i} \right) S_{t,i} (1 - \varepsilon_{i,t}) &= 0 \\
\sum_{j=u,d} \pi_{t,t+1,j} \frac{\partial J_{l,t+1,j}}{\partial \theta_{l,t,i}} \left(\left\{ \widehat{\theta}_{l,t,i} + \widehat{\theta}_{l,t,i} - \theta_{l,t-1,i} \right\}, \cdot, e_{l,t+1,j}, t+1 \right) & \quad (13) \\
&= \phi_{l,t} \times S_{t,i} \times (1 + \lambda_{i,t}) - \mu_{1,l,t,i} \\
\sum_{j=u,d} \pi_{t,t+1,j} \frac{\partial J_{l,t+1,j}}{\partial \theta_{l,t,i}} \left(\left\{ \widehat{\theta}_{l,t,i} + \widehat{\theta}_{l,t,i} - \theta_{l,t-1,i} \right\}, \cdot, e_{l,t+1,j}, t+1 \right) \\
&= \phi_{l,t} \times S_{t,i} \times (1 - \varepsilon_{i,t}) + \mu_{2,l,t,i} \\
\widehat{\theta}_{l,t,i} \leq \theta_{l,t-1,i} \leq \widehat{\theta}_{l,t,i}; \mu_{1,l,t,i} \geq 0; \mu_{2,l,t,i} \geq 0 \\
\mu_{1,l,t,i} \times \left(\widehat{\theta}_{l,t,i} - \theta_{l,t-1,i} \right) = 0; \mu_{2,l,t,i} \times \left(\theta_{l,t-1,i} - \widehat{\theta}_{l,t,i} \right) = 0
\end{aligned}$$

where the last two equations are referred to as the “complementary-slackness” conditions. Two of the first-order conditions imply that

$$\phi_{l,t} \times S_{t,i} \times (1 + \lambda_{i,t}) - \mu_{1,l,t,i} = \phi_{l,t} \times S_{t,i} \times (1 - \varepsilon_{i,t}) + \mu_{2,l,t,i}$$

Therefore, we can merge two Lagrange multipliers into one, $R_{l,t,i}$, defined as:

$$\phi_{l,t} \times R_{l,t,i} \times S_{t,i} \triangleq \phi_{l,t} \times S_{t,i} \times (1 + \lambda_{i,t}) - \mu_{1,l,t,i} = \phi_{l,t} \times S_{t,i} \times (1 - \varepsilon_{i,t}) + \mu_{2,l,t,i}$$

and recognize one first-order condition that replaces two of them:

$$\begin{aligned} \sum_{j=u,d} \pi_{t,t+1,j} \frac{\partial J_{l,t+1,j}}{\partial \theta_{l,t,i}} \left(\left\{ \widehat{\theta}_{l,t,i} + \widehat{\theta}_{l,t,i} - \theta_{l,t-1,i} \right\}, \cdot, e_{l,t+1,j}, t+1 \right) & \quad (14) \\ & = \phi_{l,t} \times R_{l,t,i} \times S_{t,i} \end{aligned}$$

In order to eliminate the value function from the first-order conditions, we differentiate the Lagrangian with respect to $\theta_{l,t-1,i}$ and then make use of (14):

$$\begin{aligned} \frac{\partial J_l}{\partial \theta_{l,t-1,i}} & = \frac{\partial L_l}{\partial \theta_{l,t-1,i}} \\ & = - \sum_{j=u,d} \pi_{t,t+1,j} \frac{\partial J_{l,t+1,j}}{\partial \theta_{l,t,i}} \left(\left\{ \widehat{\theta}_{l,t,i} + \widehat{\theta}_{l,t,i} - \theta_{l,t-1,i} \right\}, \cdot, e_{l,t+1,j}, t+1 \right) \\ & \quad + \phi_{l,t} [\delta_{t,i} + S_{t,i} \times (1 + \lambda_{i,t}) + S_{t,i} \times (1 - \varepsilon_{i,t})] - \mu_{1,l,t,i} + \mu_{2,l,t,i} \\ & = - \sum_{j=u,d} \pi_{t,t+1,j} \frac{\partial J_{l,t+1,j}}{\partial \theta_{l,t,i}} \left(\left\{ \widehat{\theta}_{l,t,i} + \widehat{\theta}_{l,t,i} - \theta_{l,t-1,i} \right\}, \cdot, e_{l,t+1,j}, t+1 \right) \\ & \quad + \phi_{l,t} \delta_{t,i} + 2\phi_{l,t} \times R_{l,t,i} \times S_{t,i} \\ & = \phi_{l,t} \times (\delta_{t,i} + R_{l,t,i} \times S_{t,i}) \end{aligned}$$

so that the first-order conditions can also be written:

$$\begin{aligned} u'_l(c_{l,t}, t) & = \phi_{l,t} \\ e_{l,t} + \sum_{i=1,2} \theta_{l,t-1,i} \delta_{t,i} - c_{l,t} - \sum_{i=1,2} \left(\widehat{\theta}_{l,t,i} + \widehat{\theta}_{l,t,i} - 2 \times \theta_{l,t-1,i} \right) \times R_{l,t,i} \times S_{t,i} \\ & \quad - \sum_{i=1,2} \left(\widehat{\theta}_{l',t,i} - \theta_{l',t-1,i} \right) S_{t,i} \lambda_{i,t} + \sum_{i=1,2} \left(\widehat{\theta}_{l',t,i} - \theta_{l',t-1,i} \right) S_{t,i} \varepsilon_{i,t} = 0 \\ \sum_{j=u,d} \pi_{t,t+1,j} \times \phi_{l,t+1,j} \times (\delta_{t+1,i,j} + R_{l,t+1,i,j} \times S_{t+1,i,j}) & = \phi_{l,t} \times R_{l,t,i} \times S_{t,i} \end{aligned} \quad (15)$$

$$\begin{aligned} \widehat{\theta}_{l,t,i} & \leq \theta_{l,t-1,i} \leq \widehat{\theta}_{l,t,i} \\ 1 - \varepsilon_{i,t} & \leq R_{l,t,i} \leq 1 + \lambda_{i,t}; \\ (-R_{l,t,i} + 1 + \lambda_{i,t}) \times \left(\widehat{\theta}_{l,t,i} - \theta_{l,t-1,i} \right) & = 0 \\ (R_{l,t,i} - (1 - \varepsilon_{i,t})) \times \left(\theta_{l,t-1,i} - \widehat{\theta}_{l,t,i} \right) & = 0 \end{aligned}$$

As has been noted by Dumas and Lyasoff (2010) in a different context, the system made of (15) and (6) above has a drawback. It must be solved simultaneously (or globally) for all nodes of all times. As written, it cannot be solved recursively in the backward way because the unknowns at time t include consumptions at time t , $c_{l,t}$, whereas the third subset of equations in (15) if rewritten as:

$$\begin{aligned} \sum_{j=u,d} \pi_{t,t+1,j} \times u'_l(c_{l,t+1,j}, t) \times [\delta_{t+1,i,j} + R_{l,t+1,i,j} \times S_{t+1,i,j}] \\ = \phi_{l,t} \times R_{l,t,i} \times S_{t,i}; l = 1, 2 \end{aligned}$$

can be seen to be a restriction on consumptions at time $t + 1$, which at time t would already be solved for.

In order to “synchronize” the solution algorithm of the equations and allow recursivity, we first shift all first-order conditions, except the third one, forward in time and, second, we no longer make explicit use of the investor’s positions $\theta_{l,t-1,i}$ held when entering time t , focusing instead on the positions $\theta_{l,t+1,i,j}$ ($\sum_{l=1,2} \theta_{l,t+1,i,j} = 0$ or 1) held when exiting time $t + 1$, which are carried backward. Regrouping equations in that way leads to the equation system of Section 1.

B Time 0

After solving the equation system of Section 1, it remains to solve at time 0 the following equation system ($t = -1, t + 1 = 0$) *from which the kernel conditions only have been removed*:⁴⁸

1. First-order conditions for time 0 consumption:

$$u'_l(c_{l,0}, 0) = \phi_{l,0}$$

2. The set of time-0 flow budget constraints for all investors and all states of nature of that time:

$$\begin{aligned} e_{l,0} + \sum_{i=1,2} \theta_{l,-1,i} \delta_{0,i} - c_{l,0} - \sum_{i=1,2} (\theta_{l,0,i} - \theta_{l,-1,i}) \times R_{l,0,i} \times S_{0,i} \\ + \sum_{i=1,2} (\widehat{\theta}_{l',0,i} - \theta_{l',-1,i}) S_{0,i} \lambda_{i,0} - \sum_{i=1,2} (\widehat{\theta}_{l',0,i} - \theta_{l',-1,i}) S_{0,i} \varepsilon_{i,0} = 0 \end{aligned}$$

3. Definitions:

$$\theta_{l,0,i} = \widehat{\theta}_{l,0,i} + \widehat{\theta}_{l,0,i} - \theta_{l,-1,i}$$

⁴⁸There could be several possible states j at time 0 but we have removed the subscript j .

4. Complementary-slackness conditions:

$$\begin{aligned} (-R_{l,0,i} + 1 + \lambda_{i,0}) \times (\widehat{\theta}_{l,0,i} - \theta_{l,-1,i}) &= 0 \\ (R_{l,0,i} - (1 - \varepsilon_{i,0})) \times (\theta_{l,-1,i} - \widehat{\theta}_{l,0,i}) &= 0 \end{aligned}$$

5. Market-clearing restrictions:

$$\sum_{l=1,2} \theta_{l,-1,i} = 0 \text{ or } 1$$

This system can be handled in one of two ways:

1. We can either solve for the unknowns $\left\{ c_{l,0}, \theta_{l,-1,i}, \widehat{\theta}_{l,0,i}, \widehat{\theta}_{l,0,i}; l = 1, 2; j = u, d \right\}$ as functions of $\{\phi_{l,0}\}$ and $\{R_{l,0,i}\}$. If we plot $\theta_{l,-1,i}$ as functions of $\{\phi_{l,0}\}$ and $\{R_{l,0,i}\}$, we have the “Negishi map”.⁴⁹ If it is invertible, we can then invert that Negishi map to obtain the values of $\{\phi_{l,0}\}$ and $\{R_{l,0,i}\}$ such that $\theta_{l,-1,i} = \widehat{\theta}_{l,0,i}$. If the values $\widehat{\theta}_{l,0,i}$ fall outside the image set of the Negishi map, there simply does not exist an equilibrium as one investor would, at equilibrium prices, be unable to repay his/her debt to the other investor.
2. Or we drop the market-clearing equation also and solve directly this system for the unknowns: $\left\{ c_{l,0}, \phi_{l,0}, R_{l,0,i}, \widehat{\theta}_{l,0,i}, \widehat{\theta}_{l,0,i}; l = 1, 2; j = u, d \right\}$ with $\theta_{l,-1,i}$, replaced in the system by the given $\widehat{\theta}_{l,0,i}$.

In this paper, the second method has been used.

C Scale-invariance property

Assuming that the risky asset pays as dividends the endowment of Investor 1, i.e. $\delta_{T,2,j} = e_{1,T,j}$, we now show that all the nodes of a given point in time, which differ only by their value of the exogenous variable, are isomorphic to each other, where the isomorphy simply means that we can factor out the endowment.

Time T-1

Given the fact that we have zero transactions fees in the last period T , using the first-order conditions for consumption, and rewriting the investors’ consumptions in terms of consumption shares $\omega_{l,T,j}$ the system of equations at time $T - 1$ can be re-written as:

$$e_{l,T,j} + \sum_{i=1,2} \theta_{l,T-1,i} \delta_{T,i,j} - \omega_{l,T,j} \times (e_{1,T,j} + \delta_{T,2,j}) = 0$$

⁴⁹For a definition of the “Negishi map” in a market with frictions, see Dumas and Lyasoff (2010).

$$\begin{aligned}
& \beta_1 \sum_{j=u,d} \pi_{T-1,T,j} \left(\frac{\omega_{1,T,j}}{\omega_{1,T-1}} \times \frac{(e_{1,T,j} + \delta_{T,2,j})}{(e_{1,T-1} + \delta_{T-1,2})} \right)^{-\gamma_1} \\
&= \beta_2 \sum_{j=u,d} \pi_{T-1,T,j} \left(\frac{\omega_{2,T,j}}{\omega_{2,T-1}} \times \frac{(e_{1,T,j} + \delta_{T,2,j})}{(e_{1,T-1} + \delta_{T-1,2})} \right)^{-\gamma_2} \\
&= \frac{\beta_1}{R_{1,T-1,i}} \sum_{j=1,2} \pi_{T-1,T,j} \times \left(\frac{\omega_{1,T,j}}{\omega_{1,T-1}} \times \frac{(e_{1,T,j} + \delta_{T,2,j})}{(e_{1,T-1} + \delta_{T-1,2})} \right)^{-\gamma_1} \times \delta_{T,i,j} \\
&= \frac{\beta_2}{R_{2,T-1,i}} \sum_{j=1,2} \pi_{T-1,T,j} \times \left(\frac{\omega_{2,T,j}}{\omega_{2,T-1}} \times \frac{(e_{1,T,j} + \delta_{T,2,j})}{(e_{1,T-1} + \delta_{T-1,2})} \right)^{-\gamma_2} \times \delta_{T,i,j} \\
& \quad \sum_{l=1,2} \theta_{l,T-1,1} = 0; \quad \sum_{l=1,2} \theta_{l,T-1,2} = 1
\end{aligned}$$

with unknowns $\{\omega_{l,T,j}; l = 1, 2; j = 1, 2\}$, $\{\theta_{l,T-1,i}; l = 1, 2; i = 1, 2\}$.

We can solve the flow budget equation for $j = 1$ for the holdings in the first asset:

$$\theta_{l,T-1,1} = 2 \times e_{1,T-1} \times u \times \left(\omega_{l,T,1} - \frac{1}{2} 1_{l,E} - \frac{1}{2} \theta_{l,T-1,2} \right), \quad (16)$$

where $1_{l,E}$ denotes an indicator for receiving endowment, i.e., $1_{1,E} = 1$ and $1_{2,E} = 0$, and we define $\frac{e_{1,t+1,1}}{e_{1,t}} = u$ as well as $\frac{e_{1,t+1,2}}{e_{1,t}} = d$. Plugging this expression into the flow budget equation for $j = 2$, we can solve for $\theta_{l,T-1,2}$:

$$\theta_{l,T-1,2} = \frac{1_{l,E} \times (d - u) + 2u \times \omega_{l,T,1} - 2d \times \omega_{l,T,2}}{u - d}. \quad (17)$$

Rewriting the kernel conditions and reducing the system using (16) and (17), we get a system with unknowns $\{\omega_{l,T,j}; l = 1, 2; j = 1, 2\}$ only:

$$\begin{aligned}
& \beta_1 \sum_{j=u,d} \pi_{T-1,T,j} \left(\frac{\omega_{1,T,j}}{\omega_{1,T-1}} \right)^{-\gamma_1} r_j^{-\gamma_1} = \beta_2 \sum_{j=u,d} \pi_{T-1,T,j} \left(\frac{\omega_{2,T,j}}{\omega_{2,T-1}} \right)^{-\gamma_2} r_j^{-\gamma_2} \\
&= \frac{\beta_1}{R_{1,T-1,i}} \sum_{j=u,d} \pi_{T-1,T,j} \left(\frac{\omega_{1,T,j}}{\omega_{1,T-1}} \right)^{-\gamma_1} r_j^{-\gamma_1+1} \\
&= \frac{\beta_2}{R_{2,T-1,i}} \sum_{j=u,d} \pi_{T-1,T,j} \left(\frac{\omega_{2,T,j}}{\omega_{2,T-1}} \right)^{-\gamma_2} r_j^{-\gamma_2+1} \\
& \quad (\omega_{1,T,1} + \omega_{2,T,1}) \frac{d}{d-u} + (\omega_{1,T,2} + \omega_{2,T,2}) \frac{d}{u-d} = 0 \\
& \quad \frac{(d-u) + 2u \times \omega_{1,T,1} - 2d \times \omega_{1,T,2}}{u-d} + \frac{2u \times \omega_{2,T,1} - 2d \times \omega_{2,T,2}}{u-d} = 1
\end{aligned}$$

where $r_j = u$ for $j = 1$ and $r_j = d$ for $j = 2$.

Importantly this system of equations *does not depend on the current or future levels of endowment*, i.e. it is enough to solve the system for one node at time $T - 1$ as long as u and d are not state (node) dependent.

After solving this system, one can compute the implied holdings and asset prices. From (17) we get that the stock holdings are independent of $T - 1$ endowment, while from (16) we know that the bond holdings are scaled by the $T - 1$ endowment:

$$\begin{aligned}\theta_{l,T-1,1} &= 2 \times e_{1,T-1} \times u \times \left(\omega_{1,T,1} \frac{d}{d-u} + \omega_{1,T,2} \frac{d}{u-d} \right) \\ &= 2 \times e_{1,T-1} \times \bar{\theta}_{l,T-1,1},\end{aligned}\quad (18)$$

where $\bar{\theta}_{l,T-1,1}$ denotes the normalized bond holdings for $e_{1,T-1} = 0.5$, i.e. $2 \times e_{1,T-1} = 1$. Moreover, we get that the bond price does not depend on $T - 1$ endowment:

$$S_{T-1,1} = \beta_1 \sum_{j=u,d} \pi_{T-1,T,j} \left(\frac{\omega_{1,T,j}}{\omega_{1,T-1}} \right)^{-\gamma_1} r_j^{-\gamma_1},$$

and that the stock price is scaled by the $T - 1$ endowment:

$$\begin{aligned}S_{T-1,2} &= 2 \times e_{1,T-1} \times \left[\frac{\beta_1}{R_{1,T-1,i}} \sum_{j=u,d} \pi_{T-1,T,j} \left(\frac{\omega_{1,T,j}}{\omega_{1,T-1}} \right)^{-\gamma_1} \frac{r_j^{-\gamma_1+1}}{2} \right] \\ &\triangleq 2 \times e_{1,T-1} \times \bar{S}_{T-1,2},\end{aligned}\quad (19)$$

where $\bar{S}_{T-1,2}$ denotes the normalized price for $e_{1,T-1} = 0.5$.

Time $t < T-1$

For time $t < T - 1$ the system of equations is the system of Section 1. Rewriting $c_{l,t+1,j} = \omega_{l,t+1,j} \times 2 \times e_{1,t+1,j}$, replacing $S_{t+1,2}$ and $\theta_{l,t+1,1}$ with expressions (19) and (18), and solving the flow budget equation for $j = 1$ for $\theta_{l,t,1}$, we get:

$$\theta_{l,t,1} = (2 \times e_{1,t}) \times u \times \left(\kappa_{l,t+1,1} - \theta_{l,t,2} \times \left(R_{l,t+1,2,1} \times \bar{S}_{t+1,2,1} + \frac{1}{2} \right) \right)$$

where

$$\begin{aligned}\kappa_{l,t+1,j} &= \omega_{l,t+1,j} - \frac{1}{2} 1_{l,E} + \bar{\theta}_{l,t+1,1,j} \bar{S}_{t+1,1,j} + \bar{\theta}_{l,t+1,2,j} R_{l,t+1,2,j} \bar{S}_{t+1,2,j} \\ &\quad - \left(\hat{\theta}_{l',t+1,2,j} - \theta_{l',t,2} \right) \bar{S}_{t+1,2,j} \lambda_{2,t+1,j} + \left(\hat{\theta}_{l',t+1,2,j} - \theta_{l',t,2} \right) \bar{S}_{t+1,2,j} \varepsilon_{2,t+1,j}\end{aligned}$$

Plugging this into the budget equation for $j = 2$, and solving for $\theta_{l,t,2}$ we get:

$$\theta_{l,t,2} = \frac{\kappa_{l,t+1,1} \times u - \kappa_{l,t+1,2} \times d}{u \left(R_{l,t+1,2,1} \bar{S}_{t+1,2,1} + \frac{1}{2} \right) - d \left(R_{l,t+1,2,2} \bar{S}_{t+1,2,2} + \frac{1}{2} \right)} \quad (20)$$

such that

$$\theta_{l,t,1} = (2 \times e_{1,t}) \times u \times \left(\kappa_{l,t+1,1} - \frac{(\kappa_{l,t+1,1} \times u - \kappa_{l,t+1,2} \times d) \times (R_{l,t+1,2,1} \times \bar{S}_{t+1,2,1} + \frac{1}{2})}{u (R_{l,t+1,2,1} \bar{S}_{t+1,2,1} + \frac{1}{2}) - d (R_{l,t+1,2,2} \bar{S}_{t+1,2,2} + \frac{1}{2})} \right) \quad (21)$$

Rewriting the kernel conditions, we can write the system as:

$$\begin{aligned} \beta_1 \sum_{j=u,d} \pi_{T-1,T,j} \left(\frac{\omega_{1,T,j}}{\omega_{1,T-1}} \right)^{-\gamma_1} r_j^{-\gamma_1} &= \beta_2 \sum_{j=u,d} \pi_{T-1,T,j} \left(\frac{\omega_{2,T,j}}{\omega_{2,T-1}} \right)^{-\gamma_2} r_j^{-\gamma_2} \\ &= \frac{\beta_1}{R_{1,t,2}} \sum_{j=u,d} \pi_{t,t+1,j} \left(\frac{\omega_{1,t+1,j}}{\omega_{1,t}} \right)^{-\gamma_1} r_j^{-\gamma_1} \left(R_{1,t+1,2,j} \times \bar{S}_{t+1,2,j} \times r_j + \frac{r_j}{2} \right) \\ &= \frac{\beta_1}{R_{2,t,2}} \sum_{j=u,d} \pi_{t,t+1,j} \left(\frac{\omega_{2,t+1,j}}{\omega_{2,t}} \right)^{-\gamma_2} r_j^{-\gamma_2} \left(R_{2,t+1,2,j} \times \bar{S}_{t+1,2,j} \times r_j + \frac{r_j}{2} \right) \\ \theta_{l,t+1,2,j} &= \hat{\theta}_{l,t+1,2,j} + \widehat{\hat{\theta}}_{l,t+1,2,j} - \theta_{l,t,2} \\ (-\bar{R}_{l,t+1,2,j} + 1 + \bar{\lambda}_{2,t+1,j}) \times (\hat{\theta}_{l,t+1,2,j} - \theta_{l,t,2}) &= 0 \\ (\bar{R}_{l,t+1,2,j} - (1 - \bar{\varepsilon}_{2,t+1,j})) \times (\theta_{l,t,2} - \widehat{\hat{\theta}}_{l,t+1,2,j}) &= 0 \\ \sum_{l=1,2} \theta_{l,T-1,1} &= 0; \quad \sum_{l=1,2} \theta_{l,T-1,2} = 1 \end{aligned}$$

with unknowns $\left\{ \omega_{l,t+1,j}; R_{l,t+1,2,j}; \hat{\theta}_{l,t+1,2,j}; \widehat{\hat{\theta}}_{l,t+1,2,j}; l = 1, 2; j = 1, 2 \right\}$. The holdings implied are given by (21) and (20). Note, one can show that the endowment $e_{1,t}$ cancels out in the market clearing conditions for the bond. Thus, the full system *does not depend on the level of endowment* $e_{1,t}$, only on u as well as d , and therefore we only need to solve the system at one node at time t .

As backward interpolated values we use the bond price $S_{t+1,2,j}$ and stock holdings $\theta_{l,t+1,2,j}$ as well as the normalized stock price $\bar{S}_{t+1,2,j}$ and normalized bond holdings $\theta_{l,t+1,1,j}$. After solving the system we can compute the implied time t holdings and prices. Again, holdings in the bond and the stock price are scaled by $2 \times e_{1,t}$, while the holdings in the stock and the bond price are not scaled. *Using backward induction the scaling invariance holds for any time t .*

D Proof of Proposition 3

The proof is by induction.

At date $t = T - 1$, agent's l present value of dividends δ is given by:

$$\hat{S}_{T-1,i,l} = \mathbb{E}_{T-1} \left[\frac{\phi_{l,T}}{\phi_{l,T-1}} \times \delta_{T,i} \right].$$

whereas Equation (8) applied to time $T - 1$ is:

$$\begin{aligned} R_{l,T-1,i} \times S_{T-1,i} &= \mathbb{E}_{T-1} \left[\frac{\phi_{l,T}}{\phi_{l,T-1}} \times \delta_{T,i} \right] \\ &= \hat{S}_{T-1,i,l} \end{aligned} \tag{22}$$

At $t = T - 2$, agent's l present value of dividends is:

$$\hat{S}_{T-2,i,l} = \mathbb{E}_{T-2} \left[\frac{\phi_{l,T-1}}{\phi_{l,T-2}} \times \left(\delta_{T-1,i} + \hat{S}_{T-1,i,l} \right) \right]$$

whereas Equation (8) applied to time $T - 2$ is:

$$\begin{aligned} R_{l,T-2,i} \times S_{T-2,i} &= \mathbb{E}_{T-2} \left[\frac{\phi_{l,T-1}}{\phi_{l,T-2}} \times \left(\delta_{T-1,i} + R_{l,T-1,i} \times S_{T-1,i} \right) \right] \\ &= \mathbb{E}_{T-2} \left[\frac{\phi_{l,T-1}}{\phi_{l,T-2}} \times \left(\delta_{T-1,i} + \hat{S}_{T-1,i,l} \right) \right] \\ &= \hat{S}_{T-2,i,l} \end{aligned}$$

where we used equation (22) to replace $R_{l,T-1,i} \times S_{T-1,i}$.

By an induction argument one can show the final result (3).

E Proof of Proposition 4

The proof is by induction.

At date $t = T - 1$, the stock price in an economy without transactions fees is given by:

$$S_{T-1}^* = \mathbb{E}_{T-1} \left[\frac{\phi_{l,T}^*}{\phi_{l,T-1}^*} \delta_T \right]$$

whereas Equation (8) applied to time $T - 1$ is:

$$R_{l,T-1} \times S_{T-1} = \mathbb{E}_{T-1} \left[\frac{\phi_{l,T}}{\phi_{l,T-1}} \delta_T \right]$$

which can be rewritten as:

$$\begin{aligned} R_{l,T-1} \times S_{T-1} &= \mathbb{E}_{T-1} \left[\frac{\phi_{l,T}^*}{\phi_{l,T-1}^*} \delta_T \right] + \mathbb{E}_{T-1} \left[\left(\frac{\phi_{l,T}}{\phi_{l,T-1}} - \frac{\phi_{l,T}^*}{\phi_{l,T-1}^*} \right) \delta_T \right] \\ &= \mathbb{E}_{T-1} \left[\frac{\phi_{l,T}^*}{\phi_{l,T-1}^*} \delta_T \right] + \mathbb{E}_{T-1} [\Delta \phi_{l,T} \times \delta_T] \end{aligned}$$

where we defined:

$$\Delta\phi_{l,T} \triangleq \frac{\phi_{l,T}}{\phi_{l,T-1}} - \frac{\phi_{l,T}^*}{\phi_{l,T-1}^*}.$$

We can thus derive the following relation between the stock price in a zero-transactions fees economy S_{T-1}^* and the stock price in an economy with transactions fees S_{T-1} :

$$R_{l,T-1} \times S_{T-1} - S_{T-1}^* = \mathbb{E}_{T-1} [\Delta\phi_{l,T}\delta_T] \quad (23)$$

At $t = T - 2$, the stock price in an economy without transactions costs is given by:

$$S_{T-2}^* = \mathbb{E}_{T-2} \left[\frac{\phi_{l,T-1}^*}{\phi_{l,T-2}^*} (\delta_{T-1} + S_{T-1}^*) \right]$$

whereas Equation (8) applied to time $T - 2$ is:

$$R_{l,T-2} \times S_{T-2} = \mathbb{E}_{T-2} \left[\frac{\phi_{l,T-1}}{\phi_{l,T-2}} (\delta_{T-1} + R_{l,T-1} \times S_{T-1}) \right]$$

Replacing $R_{l,T-1} \times S_{T-1}$ with expression (23), this can be rewritten as:

$$\begin{aligned} R_{l,T-2} \times S_{T-2} &= \mathbb{E}_{T-2} \left[\frac{\phi_{l,T-1}}{\phi_{l,T-2}} (\delta_{T-1} + S_{T-1}^* + \mathbb{E}_{T-1} [\Delta\phi_{l,T}\delta_T]) \right] \\ &= \mathbb{E}_{T-2} \left[\frac{\phi_{l,T-1}^*}{\phi_{l,T-2}^*} (\delta_{T-1} + S_{T-1}^*) \right] \\ &\quad + \mathbb{E}_{T-2} [\Delta\phi_{l,T-1} (\delta_{T-1} + S_{T-1}^*)] + \mathbb{E}_{T-2} \left[\frac{\phi_{l,T-1}}{\phi_{l,T-2}} \Delta\phi_{l,T}\delta_T \right] \\ &= S_{T-2}^* \\ &\quad + \mathbb{E}_{T-2} \left[\Delta\phi_{l,T-1} (\delta_{T-1} + S_{T-1}^*) + \frac{\phi_{l,T-1}}{\phi_{l,T-2}} \Delta\phi_{l,T}\delta_T \right] \end{aligned}$$

We can thus derive the following relation between the stock price in a zero-transactions fees economy S_{T-2}^* and the stock price in an economy with transactions fees S_{T-2} :

$$R_{l,T-2} \times S_{T-2} - S_{T-2}^* = \mathbb{E}_{T-2} \left[\Delta\phi_{l,T-1} (\delta_{T-1} + S_{T-1}^*) + \frac{\phi_{l,T-1}}{\phi_{l,T-2}} \Delta\phi_{l,T}\delta_T \right]$$

By an induction argument one can show the final result (4).

References

- Acharya, V. and L. H. Pedersen, 2005, "Asset Pricing with Liquidity Risk," *Journal of Financial Economics*, 77, 375-4010.
- Admati, A., and P. Pfleiderer, 1988, "A theory of intraday patterns: Volume and price variability," *Review of Financial Studies*, 1, 3-40.
- Akerlof, G., 1970, "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *Quarterly Journal of Economics*, 84, 488-500.
- Amihud, Y. and H. Mendelson, 1986a, "Asset Pricing and the Bid-ask Spread," *Journal of Financial Economics*, 17, 223-249.
- Amihud, Y. and H. Mendelson, 1986b, "Liquidity and Stock Returns," *Financial Analysts Journal*, 42, 43-48.
- Amihud, Y., H. Mendelson and L. Pedersen, 2005, "Liquidity and Asset Prices," *Foundations and Trends in Finance*, 1, 269-364.
- Akerlof, G. A., 1970, "The Market for 'Lemons': Quality Uncertainty and the Market Mechanism," *The Quarterly Journal of Economics*, 84, 488-500.
- Armand, P., J. Benoist and D. Orban, 2008, "Dynamic updates of the barrier parameter in primal-dual methods for nonlinear programming," *Computational Optimization and Applications*, 41, 1-25.
- Bhattacharya, U. and M. Spiegel, 1998, "Anatomy of a Market Failure: NYSE Trading Suspensions (1974-1988)," *Journal of Business & Economic Statistics*, 16, 216-226.
- Bernard, V. and J. Thomas, 1989, "Post-earnings-announcement Drift: Delayed Price Response or Risk Premium?" *Journal of Accounting Research*, 27 (Suppl.), 1-48.
- Biais, B., L. Glosten and C. Spatt, 2005, "Market Microstructure: a Survey of Microfoundations, Empirical Results, and Policy Implications," *Journal of Financial Markets*, 8, 217-264.
- Bouchard, B., 2002, "Utility Maximization on the Real Line under Proportional Transactions Costs," *Finance and Stochastics*, 6, 495-516.
- Brennan, M. J., T. Chordia, and A. Subrahmanyam, 1998, "Alternative Factor Specifications, Security Characteristics, and the Cross-section of Expected Returns," *Journal of Financial Economics* 49, 345-373.
- Brennan, M. J., T. Chordia, A. Subrahmanyam, and Q. Tong, 2012, "Sell-order liquidity and the cross-section of expected stock returns," *Journal of Financial Economics*, 105, 523-541.
- Brennan, M. J., and A. Subrahmanyam, 1996, "Market Microstructure and Asset Pricing: On the Compensation for Illiquidity in Stock Returns," *Journal of Financial Economics*, 41, 441-464.
- Brunnermeier, M. and L. H. Pedersen, 2008, "Market Liquidity and Funding Liquidity," *Review of Financial Studies*, 22, 2201-2238.
- Buss, A., R. Uppal and G. Vilkov, 2011, "Asset Prices in General Equilibrium with Transactions Costs and Recursive Utility," working paper, University of Frankfurt.

- Campbell, J. Y., S. J. Grossman and J. Wang, 1993, "Trading Volume and Serial Correlation in Stock Returns," *Quarterly Journal of Economics*, 108, 905-939.
- Campbell, J. Y., and R. J. Shiller, 1988, "The Dividend-Price Ratio and Expectations of Future Dividends and Discount Factors," *Review of Financial Studies*, 1, 195-228.
- Chordia, T., R. Roll and A. Subrahmanyam, 2008, "Liquidity and Market Efficiency", *Journal of Financial Economics*, 87.
- Clower, R. W. and D. W. Bushaw, 1954. "Price Determination in a Stock-Flow Economy," *Econometrica*, 22, 328-343.
- Constantinides, G. M., 1976a, "Optimal Portfolio Revision with Proportional Transactions Costs: Extension to HARA Utility Functions and Exogenous Deterministic Income," *Management Science*, 22, 921-923.
- Constantinides, G. M., 1976b, "Stochastic Cash Management with Fixed and Proportional Transactions Costs," *Management Science*, 22, 1320-1331.
- Constantinides, G. M., 1986, "Capital Market Equilibrium with Transactions Costs," *Journal of Political Economy*, 94, 842-862.
- Copeland, T. C., and D. Galai, 1983, "Information effects of the bid-ask spread," *Journal of Finance*, 38, 1457-1469.
- Cvitanic J. and I. Karatzas, 1996, "Hedging and Portfolio Optimization under Transactions Costs: a Martingale Approach," *Mathematical Finance*, 6, 113-165.
- Davis, M. H. and A. R. Norman, 1990, "Portfolio Selection with Transaction Costs," *Mathematics of Operations Research*, 15, 676-713.
- Deelstra, G., Pham, H. and N. Touzi, 2001, "Dual Formulation of the Utility Maximization Problem under Transaction Costs," *The Annals of Applied Probability*, 11, 1353-1383.
- de Jong, F. and B. Rindi, 2009, *The Microstructure of Financial Markets*, Cambridge University Press.
- Delgado, F., Dumas, B. and G. Puopolo, 2012, "Hysteresis Bands, Holding Period and Transactions Costs," working paper, INSEAD and Bocconi University.
- Demsetz, H., 1968, "The Cost of Transacting," *Quarterly Journal of Economics*, 82, 33-53.
- Bongaerts, D., F. De Jong and J. Driessen, 2012, "An Asset Pricing Approach to Liquidity Effects in Corporate Bond Markets," Available at SSRN: <http://ssrn.com/abstract=1762564> or <http://dx.doi.org/10.2139/ssrn.1762564>
- Dumas, B., 1989, "Two-person Dynamic Equilibrium in the Capital Market," *Review of Financial Studies*, 2 (1989), 2, 157-188.
- Dumas, B., 1992, "Capital Market Equilibrium and the Real Exchange Rate in a Spatially Separated World," *Review of Financial Studies*, 5, 153-180.
- Dumas, B. and E. Luciano, 1991, "An Exact Solution to a Dynamic Portfolio Choice Problem under Transactions Costs," *Journal of Finance*, 46, 577-595.
- Dumas, B. and A. Lyasoff, 2010, "Incomplete-Market Equilibria Solved Recursively on an Event Tree," *Journal of Finance*, forthcoming.

- Easley, D., S. Hvidkjaer and M. O'Hara, 2002, "Is Information Risk A Determinant of Asset Prices?" *Journal of Finance*, 57, 2185-2221.
- Easley, D. S., and M. O'Hara, 1987, "Price, trade size, and information in securities markets," *Journal of Financial Economics*, 19, 69-90.
- Edirsinghe, C., Naik, V., and R. Uppal, 1993, "Optimal Replication of Options with Transactions Costs and Trading Restrictions," *The Journal of Financial and Quantitative Analysis*, 28, 117-138.
- Fong, K, C. W. Holden and C. A. Trzcinka, 2010, "Can Global Stock Liquidity Be Measured?" working paper, Indiana University.
- Foster, F. D. and S. Viswanathan, 1993, "Variations in Trading Volume, Return Volatility, and Trading Costs: Evidence on Recent Price Formation Models," *The Journal of Finance*, 1993, 48, 187-211.
- Foucault, T., 1999, "Order Flow Composition and Trading Costs in a Dynamic Limit Order Market," *Journal of Financial Markets*, 99-134.
- Foucault, T., Kadan, O., and E. Kandel, 2005. "Limit Order Book as a Market for Liquidity," *Review of Financial Studies* 18: 1171-1217.
- Garcia, C. and W. Zangwill, 1981, *Pathways to Solutions, Fixed Points and Equilibria*, Prentice-Hall, Englewood Cliffs.
- Gennotte, G. and A. Jung, 1994, "Investment Strategies under Transaction Costs: The Finite Horizon Case," *Management Science*, 40, 385-404.
- Gerhold, S., P. Guasoni, J. Muhle-Karbe and W. Schachermayer, 2011, "Transactions Costs, Trading Volume and the Liquidity Premium," working paper, Technische Universität, Wien.
- Glosten, L. R. and P. R. Milgrom, 1985, "Bid, Ask, and Transaction Prices in a Specialist Market with Heterogeneous'ly Informed Agents," *Journal of Financial Economics*, 71-100.
- Goettler, R. L., Parlour, C. A., and U. Rajan, 2005, "Equilibrium in a Dynamic Limit Order Market," *Journal of Finance*, 60: 2149-2192.
- Gromb, D. and D. Vayanos, 2002, "Equilibrium and Welfare in Markets with Constrained Arbitrageurs," *Journal of Financial Economics*. 66, 361-407.
- Grossman, S. J. and J. Stiglitz, 1980, "On the Impossibility of Informationally Efficient Markets," *American Economic review*, 70, 303-408.
- Hasbrouck, J., 1991a, "Measuring the information content of stock trades," *Journal of Finance*, 46, 179-207.
- Hasbrouck, J., 1991b, "The summary informativeness of stock trades: an econometric analysis," *Review of Financial Studies*, 4, 571-595.
- Hasbrouck, J., 2007, *Empirical Market Microstructure, The Institutions, Economics, and Econometrics of Securities Trading*, Oxford University Press.
- Heaton, J. and D. Lucas, 1996, Evaluating the Effects of Incomplete Markets on Risk Sharing and Asset Pricing, *Journal of Political Economy* 104, 443-487.
- Hellwig, M., 1980, "On the Aggregation of Information in Competitive Markets", *Journal of Economic Theory*, 22, 477 - 498.
- Herings, P. J.-J. and K. Schmedders, 2006, "Computing Equilibria in Finance Economics with Incomplete Markets and Transaction Costs," *Economic Theory*, 27, 493-512.

- Ho, T. and R. Macris, 1994, "Dealer Bid-ask Quotes and Transactions Prices: an Empirical Study of Some AMEX Options," *Journal of Finance*, 39, 23-45.
- Ho, T. and H. R. Stoll, 1980, "On Dealer Markets under Competition," *Journal of Finance*, 35, 259-67.
- Ho, T. and H. R. Stoll, 1983, "The Dynamics of Dealer Markets under Competition," *Journal of Finance*, 38, 1053-1074.
- Holmström, B. and J. Tirole, 2001, "LAPM: A Liquidit-Based Asset Pricing Model," *The Journal of Finance*, LVI, 1837-1867
- Huang, J. and J. Wang, 2010, "Market Liquidity, Asset Prices, and Welfare," *Journal of Financial Economics*, 95, 107-127.
- Jang, B.-G., H. K. Koo, H. Liu and M. Lowenstein, 2007, "Liquidity Premia and Transactions Costs," *Journal of Finance*, 62, 2329-2366.
- Jouini, E. and H. Kallal, 1995, "Martingales and Arbitrage in Securities Markets with Transactions Costs," *Journal of Economic Theory*, 66? 178-197.
- Kallsen, J., J. Muhle-Karbe, 2008, "On Using Shadow Prices in Portfolio Optimization with Transaction Costs," working paper, Christian-Albrechts-Universität zu Kiel.
- Kühn C. and M. Stroh, 2010, "Optimal Portfolios of a Small Investor in a Limit-order Market: A Shadow-Price Approach," *Mathematics and Financial Economics*, 3, 45-72.
- Kyle, A. S., 1985, "Continuous auctions and insider trading," *Econometrica*, 53, 1315-1336.
- Leland, H. E., 2000, "Optimal Portfolio Implementation with Transactions Costs and Capital Gains Taxes," working paper, University of California, Berkeley.
- Lesmond, D., Ogden J., Trzcinka C., 1999, "A new estimate of transaction costs," *Review of Financial Studies*, 12, 1113-1141.
- Lettau, M. and S. C. Ludvigson, 2005, "Expected returns and expected dividend growth," *Journal of Financial Economics*, 76, 583-626.
- Liu, H., 2004, "Optimal Consumption and Investment with Transactions Costs and Multiple Risky Assets," *Journal of Finance*, 59, 289-338.
- Liu, H. and M. Loewenstein, 2002, "Optimal Portfolio Selection with Transactions Costs nad Finite Horizons," *Review of Financial Studies*, 15, 805-835.
- Llorente, G., R. Michaely, G. Saar and J. Wang, 2002, "Dynamic Volume-Return Relation of Individual Stocks," *The Review of Financial Studies*, 15, 1005-1047.
- Lo, A. W., H. Mamaysky and J. Wang, 2004, "Asset Prices and Trading Volume under Fixed Transactions Costs," *Journal of Political Economy*, 112, 1054-1090.
- Longstaff, F, 2001, "Optimal Portfolio Choice and the Valuation of Illiquid Securities," *Review of Financial Studies*, 14, 407-431.
- Longstaff, F, 2009, "Portfolio Claustrophobia: Asset Pricing in Markets with Illiquid Assets," *The American Economic Review*, 99, 1119-1144.
- Lucas, R. E. Jr., 1982, "Interest rates and currency prices in a two-country world," *Journal of Monetary Economics*, 10, 335-359.

- Madhavan, A. and S. Smidt, 1991, "A Bayesian Model of Intraday Specialist Trading," *Journal of Financial Economics*, 30, 99-134.
- Milne, F., E. Neave, 2003, "A General Equilibrium Financial Asset Economy with Transaction Costs and Trading Constraints," working paper Queen's University.
- Nazareth, M., 2002, "Portfolio Selection with Stochastic Transaction Costs," working paper.
- O'Hara, 2003, "Presidential Address: Liquidity and Price Discovery," *Journal of Finance*, 58, 1335-1354.
- Obizhaeva, A. and J. Wang, 2005, "Optimal Trading Strategy and Supply/demand Dynamics," NBER Working Paper 11444.
- Parlour, C. A., 1998, "Price Dynamics in Limit Order Markets," *Review of Financial Studies*, 11, 789-816.
- Pástor, L. and R. F. Stambaugh, 2003, "Liquidity Risk and Expected Stock Returns," *The Journal of Political Economy*, 111, 642-685.
- Peress, J., 2005, "Information vs. Entry Costs: What Explains U.S. Stock Market Evolution", *The Journal of Financial and Quantitative Analysis*, 40, 563-594.
- Roll, R., 1984, "A Simple Implicit Measure of the Effective Bid-ask Spread in an Efficient Market," *Journal of Finance*, 39, 1127-1139.
- Roşu, I., 2009, "A Dynamic Model of the Limit Order Book," *Review of Financial Studies*, 22, 4601-4641.
- Shreve, S. E. and M. H. M. Soner, 1994, "Optimal Investment and Consumption with Transactions Costs," *Annals of Applied Probability*, 4, 609-692.
- Stoll, H. R., 1989, "Inferring the components of the bid-ask spread: Theory and empirical tests," *Journal of Finance*, 44, 115-134.
- Stoll, H. R., 2000, "Friction," *Journal of Finance*, 55, 1479-1514.
- Tuckman, B. and J.-L. Vila, 2010, "Holding Costs And Equilibrium Arbitrage," bikibook.com.
- Van Binsbergen, J. H. and R. S. J. Koijen, 2010, "Predictive Regressions: A Present-Value Approach," *The Journal of Finance*, 65, 1439-1471.
- Vayanos, D., 1998, "Transaction Costs and Asset Prices: A Dynamic Equilibrium Model," *The Review of Financial Studies*, 11, 1-58.
- Vayanos, D. and J.-L. Vila, 1999, "Equilibrium interest rate and liquidity premium with transaction costs," *Economic Theory*, 13, 509-539.