DISCUSSION PAPER SERIES

No. 9524

The Dynamic Properties of Financial-Market Equilibrium with Trading Fees

Adrian Buss and Bernard J Dumas

FINANCIAL ECONOMICS

Centre for Economic Policy Research

www.cepr.org

Available online at: www.cepr.org/pubs/dps/DP9524.asp

ISSN 0265-8003

THE DYNAMIC PROPERTIES OF FINANCIAL-MARKET EQUILIBRIUM WITH TRADING FEES

Adrian Buss, INSEAD
Bernard J Dumas, INSEAD, NBER and CEPR

Discussion Paper No. 9524
July 2013

CEPR Discussion Paper No. 9524

July 2013
Revised April 2017

ABSTRACT

THE DYNAMIC PROPERTIES OF FINANCIAL-MARKET EQUILIBRIUM WITH TRADING FEES

We incorporate trading fees in a long-horizon dynamic general-equilibrium model in which traders optimally and endogenously decide when and how much to trade. A full characterization of equilibrium is provided, which allows us to study the dynamics of equilibrium trades, equilibrium asset prices and rates of return in the presence of trading fees. We exhibit the effect of trading fees on deviations from the consumption-CAPM and analyze the pricing of endogenous liquidity risk. We compare, for the same shocks, the impulse responses of this model to those of a model in which trading is infrequent because of trader inattention.

JEL Classification: G10 and G12

Adrian Buss
INSEAD
Boulevard de Constance
77305 Fontainebleau Cedex
FRANCE

Email: adrian.buss@insead.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=173126

Bernard J Dumas
INSEAD
Boulevard de Constance
77305 Fontainebleau Cedex
FRANCE

Email: bernard.dumas@insead.edu

For further Discussion Papers by this author see:
www.cepr.org/pubs/new-dps/dplist.asp?authorid=100496

*Previous versions were circulated and presented under the titles “The Equilibrium Dynamics of Liquidity and Illiquid Asset Prices” and “Financial-market Equilibrium with Friction”. Buss is with INSEAD (adrian.buss@insead.edu) and Dumas is with INSEAD, University of Torino, NBER, and CEPR (bernard.dumas@insead.edu). Work on this topic was initiated while Dumas was at the University of Lausanne and Buss was at the Goethe University of Frankfurt. Dumas’s research has been supported by the Swiss National Center for Competence in Research .FinRisk., by grant #1112 of the INSEAD research fund and by the AXA chair of the University of Torino. He is thankful to Collegio Carlo Alberto and to BI Norwegian Business School for their hospitality. The authors are grateful for useful discussions and comments to Beth Allen, Andrew Ang, Yakov Amihud, Laurent Barras, Sébastien Bétermier, Bruno Biais, John Campbell, Georgy Chabakauri, Massimiliano Croce, Magnus Dahlquist, Sanjiv Das, Xi Dong, Itamar Drechsler, Phil Dybvig, Thierry Foucault, Kenneth French, Xavier Gabaix, Nicolae Gârleanu, Stefano Giglio, Francisco Gomes, Amit Goyal, Harald Hau, John Heaton, Terrence Hendershott, Julien Hugonnier, Ravi Jagannathan, Elyès Jouini, Andrew Karolyi, Felix Kubler, David Lando, John Leahy, Francis Longsta¤, Abraham Lioui, Edith Liu, Hong Liu, Frédéric Malherbe, Ian Martin, Alexander Michaelides, Pascal Maenhout, Stefan Nagel, Stavros Panageas, Lubo. Pástor, Paolo Pasquariello, Patrice Poncet, Tarun Ramadorai, Scott Richard, Barbara Rindi, Jean-Charles Rochet, Leonid Ro¸su, Olivier Scaillet, Norman Schürhoff, Chester Spatt, Raman Uppal, Dimitri Vayanos, Pietro Veronesi, Grigory Vilkov, Vish Viswanathan, Jeffrey Wurgler, 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, the Center for Asset Pricing Research/Norwegian Finance Initiative Workshop at BI, the Indian School of Business Summer Camp, Boston University, Washington University in St Louis, ESSEC Business School, HEC Business School, McGill University, the NBER Asset Pricing Summer Institute, the University of Zurich, the University of Nantes, the Western Finance Association and the Frankfurt School of Finance and Management.

Submitted 17 June 2013

The financial sector of the economy issues and trades securities. But, more importantly, it provides a service to clients, such as the service of accessing the financial market and trading. This service is provided at a cost to cover physical as well as wage costs and the adverse selection effect of facing potentially informed customers, plus a profit. Understanding the impact of this cost on financial-market equilibrium is as important to the financial industry as the cost of producing chickpeas to the chickpea industry.

In practice, traders trade through intermediaries. However, the end users being the traders, access to a financial market is ultimately a service that traders make available to each other. As a way of constructing a simple model, we bypass intermediaries and their pricing policy, and let traders serve as dealers for, and pay trading fees to each other. We view trading fees as a metaphor for the cost (plus markup) of keeping the financial sector in operation.

We incorporate trading fees in a dynamic-equilibrium model in which traders *optimally and endogenously decide when and how much to trade*, with the purpose of increasing our understanding of the modification of trading strategies, prices, and returns associated with trading fees. We define for this market a form of Walrasian equilibrium that is the limit of a sequence of equilibria in which each trader's complementary slackness condition has been relaxed, and we invent an algorithm that takes that limit. It delivers an exact numerical equilibrium that synchronizes like clockwork the traders in the implementation of their trades and allows us to analyze the way in which trades take place and in which prices are formed and evolve. In doing this, we follow the lead of He and Modest (1995), Jouini and Kallal (1995) and Luttmer (1996), but, unlike these authors, who established bounds on asset prices, we reach a full description.

We take the cost function as given, but choose the functional form in such a way that it reflects one special and important feature of the cost of financial services: while producers of chickpeas only sell chickpeas and consumers only buy chickpeas, financial-market traders sometimes buy and, at other times, sell the very same security.[1] In both cases, they pay a positive cost to the institution that organizes the market, irrespective of the direction of the trade. For any cost function of the power type other than the square (or even-powered)

[1] Philippon (2015) defines the user cost of finance as the sum of the rate of return to a saver plus the unit cost of financial intermediation. That formulation presupposes that a saver is always a saver and a borrower always a borrower. The cost of financial intermediation is then the spread between the borrowing and the lending rates. The question of choosing between being a saver and a borrower is not on the table.

function, this implies that there is a kink at zero in the trading cost function that *induces traders not to trade* when they otherwise (without trading fees) would have. We want to draw the consequences of the traders' inaction for portfolio decisions, asset prices, and financial-market equilibrium. In particular, we assume that the cost of trading, to an approximation, is proportional to the value of the shares traded.

In our model, traders are endowed with an every-period motive for trading, over and above the long-term need to trade for lifetime planning purposes. That is, we assume that traders receive endowments that are not fully hedgeable as, even without trading fees, the financial market is incomplete. Accordingly, whenever a trader's realized endowment is above or below the amount he has previously been able to hedge, he has the desire to adjust his portfolio positions.

In the presence of proportional trading fees, traders, as is well-known from the literature on non-equilibrium portfolio choice, tolerate a deviation from their preferred holdings, the zone of tolerated deviation being called the "no-trade region." Specifically, a trader may decide not to trade, thereby preventing other traders from trading with him, which is an additional endogenous, stochastic, and, perhaps, quantitatively more important consequence of the trading fee. Liquidity begets liquidity. Conceptually and qualitatively speaking, this endogenous stochastic process of the liquidity of a security is as important to investment and valuation as is the exogenous stochastic process of its future cash flows. That is, when purchasing a security it is not sufficient for a trader to have in mind the cash flows that the security will pay into the indefinite future, he must also anticipate his, and other people's, desire and ability to resell the security in the marketplace at a later time.

We show, analytically, how this endogenous stochastic process of the liquidity affects equilibrium securities prices. That is, we compare equilibrium securities prices to traders' private valuations, which can be likened to private bid and ask prices, and explain how the gap between them triggers trades. In addition, we compare equilibrium securities prices in the presence of trading fees to those in the absence of trading fees. The differences between the two prices result from changes in the state prices, or, equivalently, consumption. Specifically, in the presence of trading fees, traders face a trade-off between smoothing consumption and smoothing holdings (to reduce trading costs).

Next, we provide quantitative results for an illustrative setting with two traded assets, one of which, the stock, is subject to trading fees. We illustrate the degree to which capital is slow-moving and how the presence of the fee intrinsically affects the traders' trading strategies. We document quantitatively the increase in consumption volatility resulting from the trading fees and its welfare implications. Finally, we study the equilibrium asset prices of both securities. The price of a risk-free bond is increasing in the trading fee of the stock because the additional consumption volatility creates a precautionary savings motive, which leads to a lower interest rate. Interestingly, the price of the stock is also increasing (slightly) in its trading fee. Particularly, the precautionary savings effect, which implies a lower discount rate for the stock, and, thus, a higher price, weakly dominates the illiquidity discount arising from stochastic (il)liquidity. We also document the implications for the return-generating processes of the two assets.

Finally, we develop three applications of our model. First, we draw the implications of our equilibrium for liquidity (risk) premia and formulate recommendations to researchers who wish to test a CAPM that takes trading fees and endogenous trading decisions into account. Second, we study an extension of our model to three traders and compare the dynamics of equilibrium to the one with two traders. Third, we construct in a proper way the responses of prices to shocks in the presence of frictions and show that the hysteresis effect of trading fees *can explain slow price reversal.*

There exists an extensive empirical literature on the impact of (il)liquidity on returns and trading activities. Particularly, it has been documented that less liquid stocks earn higher returns and are more volatile. Moreover, various liquidity (risk) premia have been studied. Our model can rationalize many of these findings. It also provides guidance for future empirical research, e.g., on endogenous liquidity (risk) premia in unconditional and conditional asset pricing model, and makes new empirical predictions, for example about the relation between trading fees and the speed of price reversal. These predictions are empirically refutable, as we verify quantitatively that the bid-ask midpoint can be used as a proxy to study the empirical implication of our model.

While we do not investigate the origin of the cost of finance, we now argue that its size is not negligible. For that, we need to be aware of what it includes. Trading costs have to be interpreted not only as bid-ask spreads or brokerage fees, but also as the opportunity cost of time devoted to portfolio selection and, above all, to information acquisition. When these activities are delegated to an intermediary, the opportunity cost becomes an effective, monetary one. Three recent papers throw light on this issue.

First, French (2008) shows that the magnitude of spreads, fees, and other trading costs, in spite of increasing competition and technological advances, is not negligible. He estimates the "the cost of active investing," defined as the difference between the aforementioned costs and the costs one would pay for holding the market portfolio, to be 0.67% per year of the aggregate market value (distinct from the size of the trade). This estimate can be interpreted as a lower bound of the actual costs for spreads and fees. He also estimates the "cost of price discovery," that is, the fees, as a percentage of managed assets, investors are willing to pay to have the portfolio selection and, above all, information acquisition done by a third party. He arrives at estimates of 1 to 2% for US mutual funds, 23 to 34 basis points for institutional investors and more than 6.5% for funds of hedge funds.

Second, Philippon (2015) provides the most comprehensive account to date of the cost of finance. He writes, "We can think of the finance industry as providing three types of services: (i) liquidity (means of payments, cash management); (ii) transfer of funds (pooling funds from savers, screening, and monitoring borrowers); (iii) information (price signals, advising on M&As)." With this definition, he shows that the total value added of the financial industry is a remarkably constant fraction of about 2% of the "total amount intermediated," in which the latter is defined as a mix of stocks of debt and equity outstanding and flows occurring in the financial industry. Hence the cost is much more than 2% of the value of trades.[2]

The third paper is by Novy-Marx and Velikov (2015), who carefully estimate the cost of implementing investment strategies based on "pricing anomalies." The costs they find range from 0.03%/month for the "Gross Profitability" strategy to 1.78%/month for the "Industry Relative Reversals" strategy. They conclude that most asset pricing anomalies could not be exploited profitably if trading costs were to be paid, so that they are no longer as puzzling. While the anomalies arise from some trader behavior that needs to be identified, trading costs do make room for, or allow, the anomalies to continue to exist.[3]

Our paper is related to the existing studies of portfolio choice under transactions costs such as Magill and Constantinides (1976), Constantinides (1976a, 1976b, 1986), Davis and Norman (1990), Dumas and Luciano (1991), Edirsinghe, Naik and Uppal (1993), Gennotte and Jung (1994), Shreve and Soner (1994), Cvitanic and Karatzas (1996), Leland (2000),

[2]The total amount intermediated, a composite of the level and flow series, is an average of the two, with flows being scaled by a factor of 8.48 to make the two series comparable.

[3]Novy-Marx and Velikov do not optimize trades, as we do here. Optimized trades may generate a reduced amount of trading costs. But, optimal policies would be of the the (s, S) type, which they do consider.

Longstaff (2001), Bouchard (2002), Obizhaeva and Wang (2013), Liu and Lowenstein (2002), Jang et al. (2007), Gerhold et al. (2011), and Gârleanu and Pedersen (2013), among others. As was noted by Dumas and Luciano, many of 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. That is, when transactions costs are linear, the portfolio strategy is of a type that recognizes the existence of a "no-trade region." Yet, portfolio-choice papers assume 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, all traders (except in Section 5.3) face a trading fee.

The papers of Heaton and Lucas (1996), Vayanos (1998), Vayanos and Vila (1999), and Lo et al. (2004) exhibit the equilibrium behavior resulting from a cost of transacting and are direct ancestors of the present one.[5] Heaton and Lucas (1996) derive a stationary equilibrium under transactions costs, but, in the neighborhood of zero trade, the cost is assumed to be quadratic, so that traders 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), a trader's only motive to trade is his finite lifetime. Transactions costs induce him to trade very little during his life. When young, he buys securities that he can resell during his old age. Here, we introduce a motive to trade that is operative at every point in time. In the paper of Lo et al. (2004), costs of trading are fixed costs, all traders have the same negative exponential utility function, and individual traders' endowments provide the motive to trade, but the amount of aggregate physical resources available is non-stochastic. In our paper, fees are proportional, preferences can be specified at will (although we present illustrative results for time-additive utility), and aggregate and individual resources are free to follow an arbitrary stochastic process. To our knowledge, ours is the first paper that

[4] Constantinides (1986), in his pioneering paper on portfolio choice under transactions costs, attempted to draw 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 very little affected by transactions costs. Liu and Lowenstein (2002), Jang et al. (2007) and Delgado et al. (2015) have shown that this is generally not true under non-IID returns. The possibility of falling in a no-trade region is obviously a violation of the IID assumption.

[5] In these papers, the cost is a physical, deadweight cost of transacting. Another predecessor is that of Milne and Neave (2003), which, however, contains few quantitative results.

obtains such an equilibrium.[6]

As far as the solution method is concerned, our analysis is closely related, in ways we explain below, to the "dual method" proposed by Bismut (1973), Cvitanic and Karatzas (1992) and used by Jouini and Kallal (1995), Cvitanic and Karatzas (1996), Cuoco (1997), Kallsen and Muhle-Karbe (2008) and Deelstra, Pham and Touzi (2002), among others.

The paper is organized in five sections. Section 1 introduces a model of financial friction. Section 2 discusses the equilibrium in an economy without trading fees, focusing particularly on the motive to trade. In Section 3, we add trading fees, define equilibrium, and compare analytically equilibrium prices to the present value of dividends and to what they would be in the absence of fees. Then, in Section 4, a numerical illustration allows to display dynamics and to discuss the impact of trading fees on trading decisions, consumption, asset prices, and the rates of return. In Section 5, we develop three applications. First, we determine theoretically what deviations from the consumption-CAPM follow from trading frictions and analyze quantitatively the endogenous behavior of liquidity premia. Second, we extend the model to three traders. Finally, we compare after an impulse the price reversals produced by our model to those that would result from infrequent trading by inattentive traders.

1 A Model of Financial Friction

We work with a generic model of a dynamic exchange economy. Time is discrete and the horizon is finite, with time being indexed by $t = 0, ..., T$. There exists a single consumption good, which is also used as a numéraire. Uncertainty is described by a tree or lattice.[7] A given node of the tree at time t is followed by K_t nodes at time $t+1$ with transition probabilities $\{\pi_{t,t+1,j}\}_{j=1}^{K_t}$.[8] The economy is populated with two traders $l = 1, 2$, who receive

[6] Longstaff (2009) studies an exogenous "blackout" period in which an asset cannot be traded, whereas in our model, the trading dates are chosen *endogenously* by the traders. In Brunnermeier and Pedersen (2008), liquidity is priced, and investors, in addition to trading with frictions, face liquidity constraints. However, some investors arrive to the market exogenously. As we do, Buss et al. (2017) derive an equilibrium in the presence of a cost of trading. But, in their paper, investors trade because of disagreement, and the focus is on the interplay between illiquidity and disagreement.

[7] Notice that the tree accommodates the exogenous state variables only. As has been noted by Dumas and Lyasoff (2012), because the tree only involves the exogenous variables, it can be chosen to be *recombining* when the endowments and payoffs are Markovian.

[8] Transition probabilities and other time-t variables depend on the current state, but, for ease of notation only, we suppress the corresponding subscript everywhere.

individual endowments and can trade multiple financial assets. However, financial markets are incomplete. The economy is further defined below.

1.1 Investment Opportunities

We model a competitive financial market with I traded securities: $i = 1, .., I$. Securities are described by their payoffs $\{\delta_{t,i};\ t = 0, ..., T\}$, which are exogenous and placed on the tree or lattice. Some securities can be short-lived, making a single payoff and being reissued immediately; for example, a risk-free, one-period bond. Other securities can be long-lived, making payoffs at all dates t. The total supply of security i is denoted by $\bar{\theta}_i$, allowing for securities in zero but also positive, net supply. Financial markets are assumed to be *incomplete*. That is, at each node of the tree, there exist fewer non-redundant securities than future nodes.

Financial-market transactions can entail trading fees, with the fees being calculated on the basis of the transaction price. That is, when a trader sells one unit of security i at time t, he receives in units of consumption good the transaction price multiplied by $1 - \varepsilon_{i,t}$ ($0 < \varepsilon_{i,t} < 1$) and, when he buys one unit, he must pay the transaction price times $1 + \lambda_{i,t}$ ($0 < \lambda_{i,t} < 1$). We assume that all fees are paid to a central pot, with the fees collected from one trader distributed to the other trader in the form of transfers. The transfers are taken by him to be lump-sum (but recurring) amounts, they enter his budget constraint, but do not generate an additional term in his first-order conditions. In this way, the trader remains purely competitive in that he only takes into account the cost of his own actions, not the benefits he may receive from the actions of the other trader.[9]

We consider a recursive Walrasian market for the securities.[10] Assuming all markets beyond time t are cleared, the auctioneer calls out time-t prices, which we call "posted" prices and denote as $\{S_{t,i}; i = 1, ..., I; t = 0, ..., T\}$. The posted price of a security is an effective transaction price only if and when a transaction takes place, but it is posted all the time by the Walrasian auctioneer. Traders submit to the auctioneer flow quantity schedules,

[9] At the request of a referee, we have redone all calculations with deadweight trading costs. The results are available upon request and confirm that this assumption is innocuous for price and trading behavior. The assumption, however, is convenient, as it allows to equate aggregate consumption to aggregate output. Without that, some output would be lost to deadweight trading costs, so that the sum of the consumption shares would have to be bounded away from 100%.

[10] We make no claim that the equilibrium in this market is unique.

knowing that fees will be calculated on the basis of that posted price in case of a buy or a sell transaction. If the flow demanded is positive at the price that is called out, the trader intends to buy; otherwise he intends to sell. The auctioneer clears the market by determining the intersection between the two schedules, if any, with one possible outcome of the clearing being a zero trade.

1.2 Endowments, Policies, and Preferences

Each trader is endowed with $\bar{\theta}_{l,i}$ shares of security i and a stream of exogenous, individual endowments $\{e_{l,t} \in \mathbb{R}_{++}; l = 1, 2; t = 0, ..., T\}$.[11] He chooses consumption and investment policies to maximize the expected utility over his lifetime consumption. For trader l, denote his consumption at time t by $c_{l,t}$, and the number of units of security i in his hands after all transactions of time t by $\theta_{l,t,i}$, so that $\{c_{l,t}; t = 0, ..., T\}$ and $\{\theta_{l,t,i}; t = 0, ..., T; i = 1, ..., I\}$ describe his consumption and investment policies.

We assume that all traders have expected utility of the form:[12]

$$\mathbb{E}_0 \sum_{t=0}^{T} u_l \left(c_{l,t}, \cdot, t\right). \tag{1}$$

While we write the utility function (1) in the additive form, given the recursive technique to be used, it would be easy to handle recursive utility, especially in the isoelastic case. Also, the utility function may contain other arguments than a trader's time-t consumption (hence the $\cdot$ as an argument), for example, past consumption in the case of habit formation.

While the equilibrium construction is based on a finite horizon T, we are able to increase T indefinitely, until such point at which the horizon no longer changes anything in the behavior of the equilibrium we found.[13]

[11] The initial holdings satisfy the restriction that $\sum_{l=1}^{2} \bar{\theta}_{l,i} = \bar{\theta}_i$.

[12] We assume that utility functions are strictly increasing, strictly concave, and differentiable to the first order with respect to consumption.

[13] See Appendix A.3 for more details.

1.3 The Trading Motive

In the model, traders trade because they receive stochastic endowments while the financial market is incomplete. They are "liquidity traders." That is, at each time t they trade or hedge a marketable component of their endowments. Thereafter, they trade again whenever the realized endowment is above or below the amount they have previously been able to hedge.[14] Particularly, in the absence of frictions, the trading motive is completely straightforward: *the trader who receives an endowment that exceeds the amount previously being hedged uses some of his funds to consume an extra amount and uses the other part to save by means of the available securities.* Frictions will impede that trading motive somewhat. How much of the "extra" endowment he consumes and how he allocates the remainder across the securities is determined endogenously.

1.4 Illustrative Setting

Throughout the paper, we will use a specific setting to illustrate the predictions of our model. Here, we quickly introduce this setting, which is described in detail in Appendix A.

In particular, we consider an economy with two traded securities: a short-lived riskless security (the "bond"), $i = 1$, in zero net supply that is not subject to trading fees; and, $i = 2$, a long-lived claim (the "stock") in unit net supply that pays out dividend $\delta_{t,2}$. Trading the stock entails trading fees. Both traders have the same preferences of the external-habit type, implemented as surplus consumption, similar to Campbell and Cochrane (1999).[15]

Aggregate output is assumed to follow a binomial tree, with the expected value of growth and its volatility matching their empirical counterparts. The dividends of the stock are simply modeled as constant fraction of aggregate output. The remainder is distributed as endowments, with the endowment shares of the traders following an independent, two-state Markov chain.[16] Particularly, we assume that the endowment shares are symmetric

[14] We leave for future or ongoing research two other motives for trading that are obviously present in the real world, such as the sharing of risk between two investors of differing risk aversions and the speculative motive arising from informed trading, private signals, or differences of opinion.

[15] External habit is introduced solely for the purpose of being able to illustrate the effects of trading fees in the presence of a realistic behavior of stock prices (equity premium, return volatility, etc.).

[16] One can demonstrate a property of scale invariance that is valid without and with trading fees: All the nodes of a given point in time, which differ only by their value of the aggregate output, are isomorphic to each other. In this way, we do not need to perform a new set of calculations for each and every node of a

and persistent and set the parameters of the Markov chain in such a way that the traders' endowments match empirical labor income dynamics. Accordingly, each trader faces $K_{t+1} = 2 \times 2 = 4$ states of nature for the immediate future: positive vs. negative growth in aggregate output and a high vs. a low share of aggregate endowment. With only two securities, the financial market is incomplete and traders must trade in response to the endowment shocks they receive.

We solve the economy recursively, increasing the horizon until such point at which it no longer changes the behavior of the equilibrium. We then simulate the equilibrium quantities on $500,000$ paths for 300 periods, after which the frequency distribution of the endogenous state variable(s) is invariant. More details are provided in Appendix A.3.

2 Equilibrium in the Absence of Trading Fees

In a first step, we analyze the equilibrium in the absence of trading fees. This case can be obtained from the general model, described in Section 1, by setting $\varepsilon_{i,t}$ and $\lambda_{i,t}$ to zero for all securities i and all dates t. This will allow us to illustrate the trading behavior in a frictionless setting and can be used to contrast the results for the case with trading fees.

2.1 Optimization Problem and Equilibrium

In order to derive an equilibrium for the economy without trading fees, we begin by stating each trader's budget constraint and optimization problem under a given stock price process. Given initial holdings $\theta_{l,-1,i} = \bar{\theta}_{l,i}$, each trader chooses consumption $\{c_{l,t}\}$ and holdings of the securities $\{\theta_{l,t,i}\}$, so as to maximize his expected utility, given in (1), subject to a sequence of flow-of-funds budget constraints for $t = 0, .., T$:

$$c_{l,t} + \sum_{i=1}^{I} (\theta_{l,t,i} - \theta_{l,t-1,i}) \times S_{t,i} = e_{l,t} + \sum_{i=1}^{I} \theta_{l,t-1,i}\delta_{t,i}. \tag{2}$$

Each trader will try to use the available assets to smooth his consumption across time and states. Particularly, in reaction to a realized endowment that exceeds the amount that a trader has hedged previously, he will consume more but also save by means of the available

given point in time. We prove this property in the Internet Appendix.

securities. The amount allocated to consumption and each of the securities is determined endogenously.

An equilibrium is defined as a process for the allocation of consumption $\{c_{l,t}\}$ of both traders, a process for portfolio choices $\{\theta_{l,t,i}\}$ of both traders, and a process for securities prices $\{S_{t,i}\}$ such that the supremum of (1) subject to the budget set is reached for all l, i and t, and the market-clearing conditions are satisfied with probability 1 at all times:

$$\sum_{l=1,2} \theta_{l,t,i} = \bar{\theta}_i; \quad i = 1, ..., I; \; t = 0, ..., T-1. \tag{3}$$

2.2 Asset Pricing

Equilibrium asset prices $S^*_{t,i}$ in the economy without trading fees can be easily derived from the individual trader's first-order conditions with respect to consumption and holdings

$$S^*_{t,i} = \mathbb{E}_t \left[\frac{\phi^*_{l,t+1}}{\phi^*_{l,t}} \times \left(\delta_{t+1,i} + S^*_{t+1,i} \right) \right], \tag{4}$$

where $\phi^*_{l,t}$ denotes the Lagrange multiplier associated with time-t budget constraint (2) and is, in equilibrium, equal to marginal utility of consumption. That is, the price of security i is simply given by the time-t value of the security's future dividends $\delta_{t+1,i}$ and future price $S^*_{t+1,i}$, discounted using a trader's stochastic discount factor $\phi^*_{l,t+1}/\phi^*_{l,t}$.

Because financial markets are incomplete, the individual traders' stochastic discount factors will not be equated. Also, the traders' individual consumption growth rates are not perfectly correlated.

2.3 Numerical Illustration

To be more specific, we now focus on the numerical illustration, introduced in Section 1.4. Table 1 reports, conditional on a high or low *realized* endowment share for Trader 1, the endowment and dividend income received by the trader as well as his endogenous consumption and security trading decisions. When the trader receives a high endowment share, he has, in total, 0.6153 units of the consumption good available, from which he allocates 0.5282 units to consumption and 0.0870 units to savings. The case of a low endowment share is

	High endow. share	**Low endow. share**
Endowment (exogenous)	0.5313	0.3188
Dividend	0.0840	0.0660
Consumption	0.5282	0.4718
Change in stock holdings	+0.1046	−0.1046
Change in bond holdings	−0.0176	+0.0176
Prob. of increase in stock holdings	100%	0%
Prob. of increase in bond holdings	33.84%.	66.16%

Table 1: **Trades in the absence of trading fees.** The table reports the consumption and investment decisions of Trader 1, conditional on his realized endowment share. In particular, it shows the size of the trader's endowment and dividend income, his consumption choice as well as the (dollar) values of his securities trades (number of shares×price) – all normalized by aggregate output. The last two rows show the probability of an increase in the stock's and bond's holdings. The table is based on the numerical illustration described in Section 1.4 and averages are computed across 500,000 simulation paths.

symmetric, with an available income of 0.3848 units, a consumption of 0.4718 units, and a disinvestment of 0.0870 units to enhance consumption.

The difference in the amount consumed between the two states is a reflection of the degree of consumption smoothing that the trader has been able to achieve. Most apparent, however, is the simple trading pattern in the stock market: *the trader always (with probability* 1*) increases his stock holdings if he receives a high share of aggregate endowment,* i.e., he buys additional shares of the stock. Knowing that his endowment shock is persistent – a positive shock today announces further positive shocks in the future – he borrows (more often than not) a modest amount through the bond in order to buy even more stock. Symmetric trading decisions can be observed for the case of a low share of aggregate endowment.

This simple trading pattern is also illustrated in Figure 2 on page 22 which shows a single sample path of the economy. It is apparent from the black, dashed line in Panel (a) that for all periods in which Trader 1 receives a high share of aggregate endowment (highlighted by shaded grey), he increases his investment in the stock – here shown in terms of number of shares. The corresponding trading decisions for the bond are depicted in Panel (c).[17]

[17] An increase (decrease) in the number of shares of the bond, as shown in Panel (c) of Figure 2, is not (always) equivalent to an increase (decrease) in the dollar bond holdings. Particularly, due to the short-term nature of the bond, the change in bond holdings is given by $\theta_{l,t,1} \times S_{t,1} - \theta_{l,t-1,1}$.

3 Equilibrium with Trading Fees

We now turn to the dynamic properties of equilibrium in the *presence of trading fees*. This allows us to study how trading fees affect traders' consumption and trading decisions as well as equilibrium asset prices and returns.

3.1 Optimization Problem and Equilibrium

We begin by stating each trader's budget constraint and optimization problem under a given stock price process. As in the case of no trading fees, each trader chooses consumption $\{c_{l,t}\}$ and holdings of the securities $\{\theta_{l,t,i}\}$, so as to maximize his expected utility (1). The only difference is that this optimization is subject to a sequence of flow-of-funds budget constraints for $t = 0, .., T$ that now take into account the fact that transactions entail trading fees:

$$c_{l,t} + \sum_{i=1}^{I} \max\left[0, \theta_{l,t,i} - \theta_{l,t-1,i}\right] \times S_{t,i} \times (1 + \lambda_{i,t}) +$$
$$\sum_{i=1}^{I} \min\left[0, \theta_{l,t,i} - \theta_{l,t-1,i}\right] \times S_{t,i} \times (1 - \varepsilon_{i,t}) = e_{l,t} + \sum_{i=1}^{I} \theta_{l,t-1,i}\delta_{t,i} + \zeta_{l,t}, \tag{5}$$

where the three terms on the left-hand side reflect consumption, net cost of purchases, and net cost of sales of securities (i.e., proceeds of sales with a negative sign), and the three terms on the right-hand side reflect endowment and dividend income as well as the transfer received from the central pot, $\zeta_{l,t}$, which is given by

$$\zeta_{l,t} \triangleq \sum_{l' \neq l} \sum_{i=1}^{I} \max\left[0, \theta_{l',t,i} - \theta_{l',t-1,i}\right] \times S_{t,i}\lambda_{i,t} - \sum_{i=1}^{I} \min\left[0, \theta_{l',t,i} - \theta_{l',t-1,i}\right] \times S_{t,i}\varepsilon_{i,t}. \tag{6}$$

A change of notation reformulates the problem in the form of an optimization *under inequality constraints*, which is more suitable for mathematical programming. That is, writing the purchases $\max\left[0, \theta_{l,t,i} - \theta_{l,t-1,i}\right]$ and sales $\min\left[0, \theta_{l,t,i} - \theta_{l,t-1,i}\right]$ (a negative number) of securities as

$$\hat{\theta}_{l,t,i} - \theta_{l,t-1,i} \triangleq \max\left[0, \theta_{l,t,i} - \theta_{l,t-1,i}\right]$$

and

$$\check{\theta}_{l,t,i} - \theta_{l,t-1,i} \triangleq \min\left[0, \theta_{l,t,i} - \theta_{l,t-1,i}\right],$$

the time-t recursive dynamic-programming formulation of the trader's problem is[18]

$$J_{l,t}\left(\{\theta_{l,t-1,i}\}, \cdot, e_{l,t}\right) = \sup_{c_{l,t}, \left\{\hat{\theta}_{l,t,i}, \check{\theta}_{l,t,i}\right\}} u_l\left(c_{l,t}, \cdot, t\right) + \mathbb{E}_t J_{l,t+1}\left(\left\{\hat{\theta}_{l,t,i} + \check{\theta}_{l,t,i} - \theta_{l,t-1,i}\right\}, \cdot, e_{l,t+1}\right), \tag{7}$$

subject to the budget constraint (5), for time t only, and to the inequality conditions:

$$c_{l,t} + \sum_{i=1}^{I}\left(\hat{\theta}_{l,t,i} - \theta_{l,t-1,i}\right) \times S_{t,i} \times (1 + \lambda_{i,t}) + \sum_{i=1}^{I}\left(\check{\theta}_{l,t,i} - \theta_{l,t-1,i}\right) \times S_{t,i} \times (1 - \varepsilon_{i,t})$$

$$= e_{l,t} + \sum_{i=1}^{I} \theta_{l,t-1,i}\delta_{t,i} + \zeta_{l,t}, \tag{8}$$

$$\check{\theta}_{l,t,i} \leq \theta_{l,t-1,i} \leq \hat{\theta}_{l,t,i}. \tag{9}$$

Under standard concavity assumptions on utility functions, the maximization of (7) subject to (8) and (9) is a convex problem. First-order conditions of optimality (including terminal conditions $\theta_{l,T,i} = 0$) are necessary and sufficient for the optimum to be reached. In Appendix B we derive the system of first-order conditions (which is system (10) to (16) below with $\eta = 0$). To obtain an equilibrium, one then usually combines the first-order conditions of both traders with the market-clearing conditions, and solves the resulting equation system. Here, for reasons explained below, we define a sequence of η-equilibria.

Definition 1 *An η-equilibrium is defined as a process for the allocation of consumption $\{c_{l,t}\}$ of both traders, a process for trading decisions $\left\{\hat{\theta}_{l,t,i}, \check{\theta}_{l,t,i}\right\}$ of both traders, a process for posted securities prices $\{S_{t,i}\}$, state prices $\{\phi_{l,t}\}$, and shadow prices $\{R_{l,t,i}\}$ that solve the following system of equations for all l, i, and t:*

$$u_l'\left(c_{l,t}, \cdot, t\right) = \phi_{l,t} \tag{10}$$

$$e_{l,t} + \sum_{i=1}^{I} \theta_{l,t-1,i}\delta_{t,i} - c_{l,t} - \sum_{i=1}^{I}\left(\hat{\theta}_{l,t,i} + \check{\theta}_{l,t,i} - 2 \times \theta_{l,t-1,i}\right) \times R_{l,t,i} \times S_{t,i} + \zeta_{l,t} = 0 \tag{11}$$

[18]The value function $J_l\left(\{\theta_{l,t-1,i}\}, \cdot, e_{l,t}, t\right)$ refers explicitly only to Trader l's individual state variables. The complete set of state variables actually used in the backward induction is chosen below.

$$\sum_{j=1}^{K_t} \pi_{t,t+1,j} \times \phi_{l,t+1,j} \times (\delta_{t+1,i,j} + R_{l,t+1,i,j} \times S_{t+1,i,j}) = \phi_{l,t} \times R_{l,t,i} \times S_{t,i} \tag{12}$$

$$\check{\theta}_{l,t,i} \le \theta_{l,t-1,i} \le \hat{\theta}_{l,t,i} \tag{13}$$

$$1 - \varepsilon_{i,t} \le R_{l,t,i} \le 1 + \lambda_{i,t}; \tag{14}$$

$$(-R_{l,t,i} + 1 + \lambda_{i,t}) \times \left(\hat{\theta}_{l,t,i} - \theta_{l,t-1,i}\right) = \eta \tag{15}$$

$$(R_{l,t,i} - (1 - \varepsilon_{i,t})) \times \left(\theta_{l,t-1,i} - \check{\theta}_{l,t,i}\right) = \eta \tag{16}$$

while the market-clearing conditions (3) are also satisfied with probability 1*.*

Definition 2 *An equilibrium is the limit (if it exists) of an η-equilibrium as $\eta \to 0$.*

In the equation system described by (10) to (16) and (3), the unknown variables $R_{l,t,i}$ (defined in Appendix B) represent the shadow price of a "paper security" valued at the posted price, in units of consumption. Whenever Trader l's inventory of security i is large (small) at time t, the value of $R_{l,t,i}$ is smaller (greater) than 1. In particular, when the shadow price of a trader reaches the value of one plus the cost of buying, the trader buys, and when the shadow price reaches the value of one minus the cost of selling, he sells. The exact opposite happens for the other trader.

For $\eta = 0$, the last two equations (15) and (16) are the familiar complementary-slackness conditions of Karush, Kuhn and Tucker (KKT). In contrast, in an η-equilibrium with η finite, the complementary slackness conditions have been relaxed and traders are conducting a suboptimal policy. In defining equilibrium in our economy, we found it necessary to introduce the limit of a sequence of η-equilibria because, literally speaking, at $\eta = 0$, the posted prices $\{S_{t,i}\}$ and the shadow prices $\{R_{l,t,i}\}$, at times other than transaction times, become indeterminate.[19] We illustrate that point by means of Figure 1.

The figure shows the traders' demands for the stock, aggregate demand for the stock, and the traders' demands for the bond, plotted against the posted price of the stock (on the horizontal axis), for a small, positive but finite value of η and for trading fees of 0.5%

[19] We thank a referee for this remark. Their product remains determinate as equations (12) and (17) below make clear. We should still stress that at equilibrium, the consumption allocation – at time t and $t+1$ – and the securities demands are constant over the no-trade region, irrespective of the posted price, so that we can regard the equilibrium allocations that we have reached as being fully determinate.

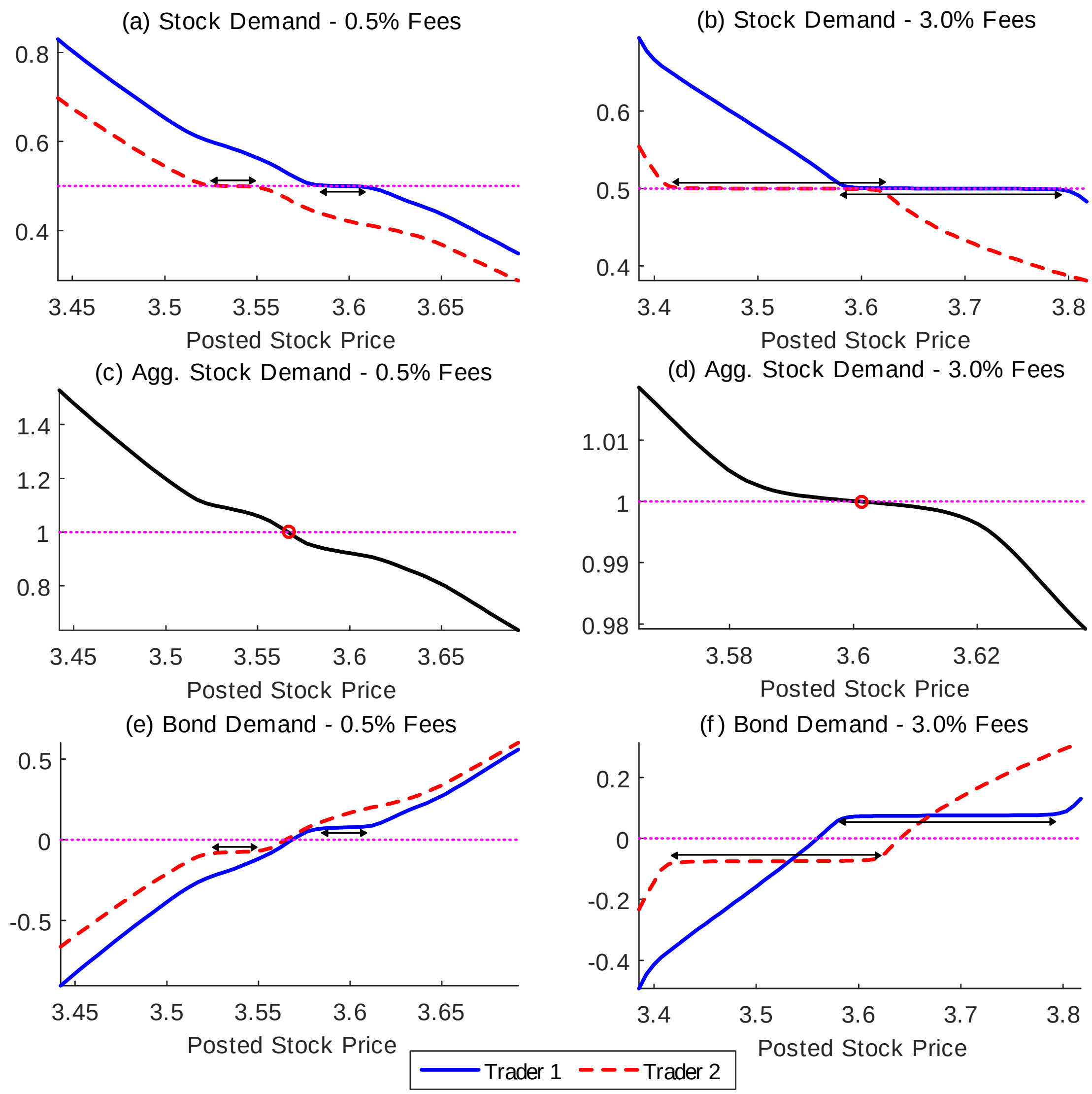

Figure 1: **Securities demand curves**. The figure shows the traders' demands for the stock, aggregate demand for the stock, and the traders' demands for the bond, plotted against the posted price of the stock (on the horizontal axis). The figure is drawn conditional on a high realized endowment shock for Trader 1, a small, positive but finite, value of η and for tradings fees of 0.5% and 3.0%. Note that in Panel (d), the scale of the x-axis is magnified to better show the intersection point. The no-trade regions of the two traders are indicated by double-headed arrows. The figure is based on the numerical illustration described in Section 1.4.

and 3.0%.[20] In particular, as illustrated in Panels (a) and (b), both traders' demands for the stock exhibit some flat regions (highlighted by double-headed arrows) in which it would be optimal not to trade. For low trading fees (0.5%) the traders' no-trade regions do not overlap, so that aggregate stock demand uniquely determines the equilibrium (transaction) price (cf. Panel (c)). In contrast, for high trading fees (3.0%) there exists a *joint no-trade zone*. For $\eta = 0$, this implies that the aggregate demand curve has a flat part over the domain of the joint no-trade region, thus leaving the posted stock price indeterminate. However, just before the limit, the aggregate-demand curve intersects the supply at a *single point*, which is the *equilibrium posted price* (highlighted by a red circle in the magnified graph in Panel (d)). Panels (e) and (f) show the demand for the bond, which is flat over the same region as the stock demand. The same holds for consumption (not shown).

Also, in the equation system described by (10) to (16) and (3), the unknown variables $\phi_{l,t}$ are the customary state prices for the state prevailing at time t. They are specific to Trader l in part because the market is incomplete and in part because of the presence of trading fees. They are, of course, different from the state prices $\phi^*_{l,t}$ that prevailed in the frictionless economy. Since state prices are also marginal utilities of consumption, it follows that the same statements can be made about individual consumption behavior.

Remark 1 *Because aggregate output (consumption) volatility is exogenously given, the probability distribution of aggregate consumption is, of course, unaffected by trading fees. But the conditional joint distribution of the individual consumptions of the two traders reflects asset holdings, which are very much affected because trading fees create impediments to trade. That is, in the presence of trading fees, traders smooth their consumption across states less effectively than they do in the frictionless economy. In other words, traders face a trade-off between the goal of smoothing consumption and the goal of smoothing holdings, with the latter being due to the desire to reduce trading fees. This will lead to an increase in the individual trader's consumption growth volatility. Since aggregate consumption volatility is unchanged, the increased individual consumption volatility must be matched with a reduced correlation of individual consumptions. The illustration of Section 4 will further discuss these effects (quantitatively).*

[20] These are demand curves drawn at time t for two traders who assume prices and wealth at time $t+1$ which are those of equilibrium. When solving for equilibrium we do not make use of demand curves; instead we solve the system of equations characterizing equilibrium (see Section 3.3 and Appendix E). Demand curves are used in Figure 1 for illustrative purposes and to discuss the point at hand.

3.2 Asset Pricing: Two Comparisons

Equilibrium asset prices in the economy with trading fees can be derived from the traders' first-order conditions. In particular, it follows directly from the "kernel condition" (12) that the securities' posted prices $S_{t,i}$ are given by:

$$S_{t,i} = \mathbb{E}_t \left[\frac{1}{R_{l,t,i}} \frac{\phi_{l,t+1}}{\phi_{l,t}} \times (\delta_{t+1,i} + R_{l,t+1,i} \times S_{t+1,i}) \right]; \quad S_{T,i} = 0, \tag{17}$$

where the shadow prices $R_{l,t,i}$ and $R_{l,t+1,i}$, which are bounded between $1 - \varepsilon_{i,t}$ and $1 + \lambda_{i,t}$, capture the effect of current and anticipated future trading fees, respectively.

The dual variables $R_{l,t+1,i}$, in addition to the intertemporal marginal rates of substitution $\phi_{l,t+1}$, drive the prices of assets that are subject to trading fees, as do, in the "LAPM" of Holmström and Tirole (2001), the shadow prices of the liquidity constraints.[21] In effect, there are two distinct pricing kernels: one, $\phi_{l,t+1}$, applies to the time-$t+1$ payoffs paid in consumption units, the other, $\phi_{l,t+1} \times R_{l,t+1,i}$, applies to the time-$t+1$ posted price.

We now present two comparisons. First, we compare equilibrium posted prices, $S_{t,i}$, to the private valuation $\hat{S}_{l,t,i}$, defined as the present value of dividends on security i calculated at Trader l's equilibrium *state prices as they are under trading fees*:

Definition 3

$$\hat{S}_{l,t,i} \triangleq \mathbb{E}_t \left[\frac{\phi_{l,t+1}}{\phi_{l,t}} \times \left(\delta_{t+1,i} + \hat{S}_{l,t+1,i} \right) \right]; \quad \hat{S}_{l,T,i} = 0.$$

In Appendix C, we show that, by induction, the present value of all future payouts, discounted using the "normal" pricing kernel only, gives the private valuation $R_{l,t,i} \times S_{t,i}$

Proposition 1

$$R_{l,t,i} \times S_{t,i} = \hat{S}_{l,t,i}. \tag{18}$$

This means that the posted prices can *at most* differ from the private valuation of their dividends by the amount of the potential *one-way trading fee at the current date only.*[22] The

[21] Holmström and Tirole (2001) make assumptions such that their liquidity constraint is always binding. Here, the inequality constraints (9) bind whenever it is optimal for them to do so.

[22] Our proposition makes the statement of Vayanos (1998) more precise, who writes (page 26): "Second, the effect of transaction costs is *smaller* than the present value of transaction costs incurred by a sequence of marginal investors." Emphasis added.

posted price is, in fact, some form of average of the two private valuations.[23]

Second, we compare equilibrium securities prices that prevail in the presence of trading fees to those that would prevail in a frictionless economy, that is, to prices based on *state prices that would obtain under zero trading fees*, as defined in (4). Denoting all quantities in the zero-trading fees economy with an asterisk *, we show in Appendix D the following proposition:

Proposition 2

$$R_{l,t,i} \times S_{t,i} = S^*_{t,i} + \mathbb{E}_t \left[\sum_{\tau=t+1}^{T} \frac{\phi_{l,\tau-1}}{\phi_{l,t}} \times \left(\frac{\phi_{l,\tau}}{\phi_{l,\tau-1}} - \frac{\phi^*_{l,\tau}}{\phi^*_{l,\tau-1}} \right) \times \left(\delta_{\tau,i} + S^*_{\tau,i} \right) \right]. \tag{19}$$

That is, the two asset prices, $S_{t,i}$ and $S^*_{t,i}$, 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 trading fees; (ii) the present value of all *future price differences* arising from the differences in state prices $\phi_{l,\tau}/\phi_{l,\tau-1} - \phi^*_{l,\tau}/\phi^*_{l,\tau-1}$.

The differences in consumption schemes pointed out in Remark 1 influence the future state prices and explain the differences in prices, so that the following is proposed:

Proposition 3 *The reason for any effect of anticipated trading fees on prices is that traders do not hold the optimal, frictionless portfolios and, therefore, also have consumption schemes that differ from those that would prevail in the absence of trading fees.*

Particularly, the increased volatility of individual consumption plays a role in setting the price because of the marginal utilities, and the reduced correlation of individual consumption also plays a role via the term $\Delta\phi_{l,\tau} \times \left(\delta_{\tau,i} + S^*_{\tau,i}\right)$. Indeed, as the payoff δ is a fraction of total output and whatever part of total output one group of traders is not consuming because of trading fees, the other group is consuming, trading fees affect the correlation between $\Delta\phi_{l,\tau}$ and $\left(\delta_{\tau,i} + S^*_{\tau,i}\right)$.

Proposition 3 is in marked contrast with the proposition stated by Amihud and Mendelson (1986a),[24] who attribute the effect of trading fees on a securities price to the future-fee cash

[23] Below we introduce private bid and ask prices that are based on the two traders' private valuations, and we claim that the posted price is extremely close to the bid-ask midpoint.

[24] See also Vayanos and Vila (1999, page 519, equation (5.12)).

expense itself. The difference between their proposition and ours is ascribed to the fact that these authors exogenously force investors to trade, whereas in our setting traders *trade optimally*. The difference is *not* to be ascribed to our assumption that the fees are refunded. If the fees had been a deadweight loss, the effect of that expense would still have been felt on consumption only. It would *not* have appeared directly in the present-value formula in the form of altered future cash flows on the security being priced, and Proposition 2 would have been equally valid.[25]

3.3 Solution Algorithm

The method used to obtain an equilibrium blends in an original fashion a shift of equations that has been proposed by Dumas and Lyasoff (2012) to facilitate backward induction with the Interior-Point algorithm, which is an optimization technique based on Karush-Kuhn-Tucker first-order conditions for optimization under non-negativity constraints.

The "time-shift" of Dumas and Lyasoff (2012) implies shifting all first-order conditions, except the kernel and market clearing conditions, forward in time and letting traders at time t plan their time-$t+1$ consumption but choose their time-t portfolio (which will, in turn, finance the time-$t+1$ consumption).

The Interior-Point algorithm amounts to replacing the above equation system (consisting of equations (10) to (16) and (3)) by a sequence of equation systems in each of which the complementary-slackness conditions are relaxed.[26] This corresponds closely to our definition of the η-equilibria. In one very convenient implementation, Armand et al. (2008) show a way to add to the system a single equation that drives η toward zero progressively with each Newton step of the solver. More details are provided in Appendix E.

[25] See footnote 9.

[26] Parenthetically, the Interior-Point method should be of great interest to microeconomists who study choice problems with inequality constraints. That is, the comparative statics of the solution can be obtained by total differentiation of the first-order conditions, for a given value of η, in the same way as is done in Microeconomics textbooks to derive Slutzky's equation. In cases in which limits can be interchanged, these comparative-statics properties are close to those that would obtain in the original system of first-order conditions with $\eta = 0$. Our approach is more closely connected to microeconomic theory than other optimization techniques, such as steepest-ascent. This remark was made by Dimitri Vayanos in a private conversation.

4 Simulation Results

To further describe the equilibrium in the presence of trading fees, we now come back to the numerical illustration introduced in Section 1.4.

4.1 Dynamics in Equilibrium

First, we describe the mechanics over time of the equilibrium we found and the transactions that take place. Particularly, in the presence of trading fees, a key concept is that of a "no-trade zone," which is the area of the state space where both traders prefer not to adjust their portfolios.

Figure 2 displays a simulated sample path that illustrates how our financial market with trading fees operates over time, with periods of a high endowment share for Trader 1 highlighted by shaded grey, and transaction dates highlighted by a red circle. Specifically, Panels (a) and (b) show a sample path of: (i) the stock holdings (expressed as a fraction of the security's supply, not as a dollar value) as they would be in a zero-trading fee economy; (ii) the actual stock-holdings with a 3% trading fee; and (iii) the boundaries of the optimal no-trade zone. Note that the boundaries of the no-trade zone fluctuate in tango with the optimal frictionless holdings, except that they allow a tunnel of deviations on each side. Within that tunnel, the traders' logic is apparent: the actual holdings move up or down whenever they are pushed up or down by the movement of the boundaries, with a view to reduce the amount of trading fees paid and making sure that *there occur as few wasteful round trips as possible*, leading to a trade-off between the desire to smooth consumption and the desire to smooth holdings. Panel (a) viewed in parallel with Panel (b) illustrates how the two traders are wonderfully synchronized by the algorithm; they are made to trade exactly opposite amounts exactly at the same time.

The figure also illustrates the degree to which *capital is slow-moving*, an issue we return to in Section 5.3 That is, the optimal stock holdings in the presence of trading fees are a delayed version of the frictionless holdings, but with the length of the delay being stochastic. To the opposite, as shown in Panels (c) and (d), holdings of the riskless bond, which is assumed not to entail trading fees, fluctuate more than they would in a frictionless economy. Specifically, when traders receive their endowments, they use the cost-free, riskless bond as a holding tank and trade it much more than they would if the stock were also cost-free.

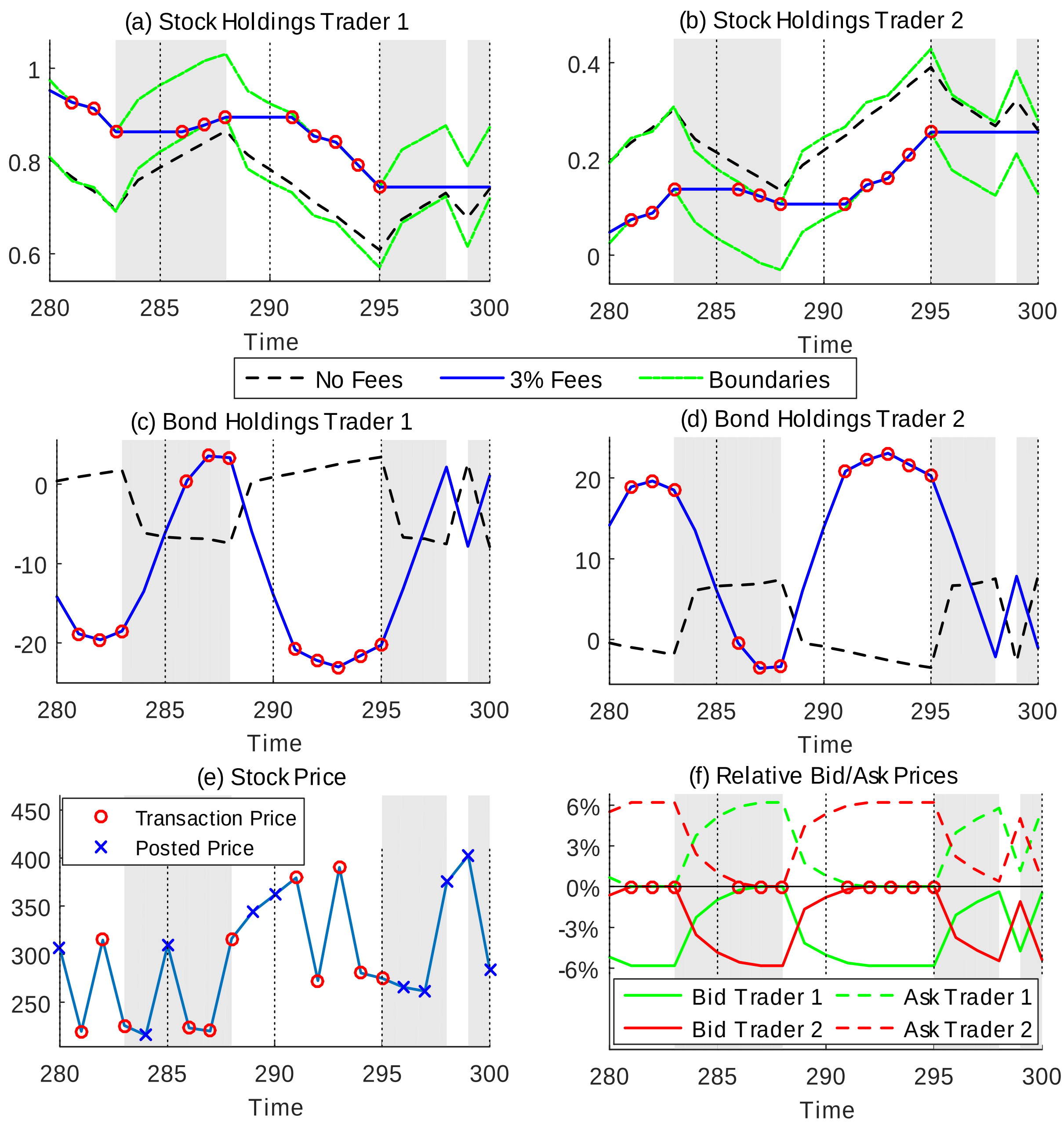

Figure 2: **A sample path**. Panels (a) and (b) show a sample path of: (i) the stock holdings as they would be in a zero-trading fee economy; (ii) the actual stock-holdings with a 3% trading fee; and (iii) the boundaries of the no-trade zone. Panels (c) and (d) show the holdings of the riskless security. Panel (e) shows the behavior of the stock price, and Panel (f) displays the bid and ask prices of both traders, computed as a percentage difference from the posted price. In all panels, periods during which Trader 1 receives a high endowment share are highlighted in shaded grey, and transaction dates are highlighted by a red circle. The figure is based on the numerical illustration described in Section 1.4.

Panel (e) shows the stock's posted price. While the posted price forms a stochastic process with realizations at each point in time, transactions prices materialize as a "marked point process" with realizations at random times only.[27] The simultaneous observation of Panels (a), (b), and (e) shows the way in which the algorithm has synchronized the trades of the two traders. After an extension to more traders, the properties of this process could be confronted empirically with those of illiquid-market prices.

Even though ours is a Walrasian market and neither a limit-order nor a dealer market, one can define a virtual concept of bid and ask prices. In Definition 3, we defined the traders' private valuations of dividends. The *bid price* of a trader can then be defined as being equal to the trader's private valuation of dividends divided by one plus the trading fee to be paid in case the person buys. Similarly, the *ask price* is defined as the trader's private valuation divided by one minus the sell fee. When the two private valuations differ by the sum of the one-way trading fees for the two traders, a transaction takes place. Equivalently, a trade occurs when the bid price of one trader is equal to the ask price of the other trader. That mechanism is displayed in Panel (f). Defining the *effective spread* as the difference between the higher of two bid prices and the lower of the two ask prices, one could also say that a transaction takes place when the effective spread becomes equal to zero.

The posted price can thus be interpreted as some form of average of the two private valuations, or some form of average of the higher bid and the lower ask price (cf. Proposition 1). We have verified quantitatively that the posted price differs very little from the bid-ask midpoint. As far as levels are concerned, the mean absolute difference between the posted price and the bid-ask midpoint divided by the effective spread is equal to 0.62bp, 1.03bp, and 1.52bp for the three cases of 1%, 2%, and 3% trading fees, respectively. Also the mean absolute difference between the rates of return on the posted price and the bid-ask midpoint are 0.00bp, 0.01bp, and 0.04bp, respectively.[28] Hence the empirical counterpart of

[27] For an application to the bond market of this type of process, see Björk et al. (1997).

[28] We have also varied the degree of symmetry between the two traders in terms of risk aversion, endowment shares, and endowment persistence. None of these invalidated the quasi-equality between the posted price and the bid-ask midpoint. One might surmise that an asymmetry in trading fees would invalidate it. We have, therefore, also computed a setting with asymmetric trading fees. In it, Trader 1 pays fees of 3% and Trader 2 pays fees of 1%. A comparison of that setting with a setting in which both traders pay 2% trading fees reveals that the level difference between posted price and bid-ask midpoint is a bit bigger, but the differences in expected returns, return volatility, Sharpe ratio, and other return statistics computed on the posted price and on the midpoint are completely negligible.

any statement we make below regarding the behavior of the posted price or rates of returns on it is a *testable statement involving the bid-ask midpoint*, viewed as a very close proxy.

The figure also illustrates that after transitioning to a high endowment state, Trader 1 may at first not buy the stock, buying the bond instead and only buys the stock if the high endowment persists. This happens when his prior holding is already high. Instead, he may buy the stock right away if his prior holding is less high. To generalize the intuition provided by a single path and to give a systematic, probabilistic representation of the pattern of trading, we depict, in Figure 3, the transition probabilities for the "shadow price ratio" $R_1/(R_1+R_2)$ of the stock for the case in which Trader 1 is currently in his high endowment state.[29] A shadow price ratio of 1.03 means that Trader 1 buys shares of the stock, and a ratio of 0.97 means that he sells. The figure shows the probability of a value of that ratio at time $t+1$ conditional on a value of it at time t.[30]

For any value of the shadow price ratio at time t, the figure makes clear that the probability of a mid-level value of the ratio occurring at $t+1$ is equal to zero or nearly so. This is the result of the trading fee being proportional: when a trader needs to trade in one direction, he trades as little as possible, knowing that he can trade repeatedly the next few times at no greater cost than he would have incurred if he had traded in a lump. The ratio, therefore, transitions to either a high value of the shadow price ratio near the buy boundary (left-hand ridge in the diagram) if Trader 1 remains in the high-endowment state, or it transitions directly to a very low value of the ratio at the sell boundary (right-hand ridge) if his endowment shifts to the low level.[31] The smaller ridge close to the large one on the left-hand side results from the combination of a high consumption share of Trader 1 at time t coupled with a negative output shock at $t+1$. Because of the high consumption share, the trader holds a large fraction of wealth in the stock, so that the negative output (and, therefore, dividend) shock implies a negative wealth shock and, accordingly, he is less willing to buy more of the stock. In summary, it follows that the equilibrium system is most of the time near a trading boundary, or, equivalently, that the steady-state probability distribution of the shadow price ratio is U-shaped.

[29] The figure for the low endowment state is symmetric, i.e., one just needs to switch the "roles" of the two traders in this figure to arrive at the figure for the low endowment share.

[30] At both times, the probability is integrated over the remaining state variables: aggregate output shock and consumption share. In those two dimensions, the probabilities shown are marginal probabilities.

[31] Intuitively, for low time-t shadow price ratios, he does not always sell after a negative shock.

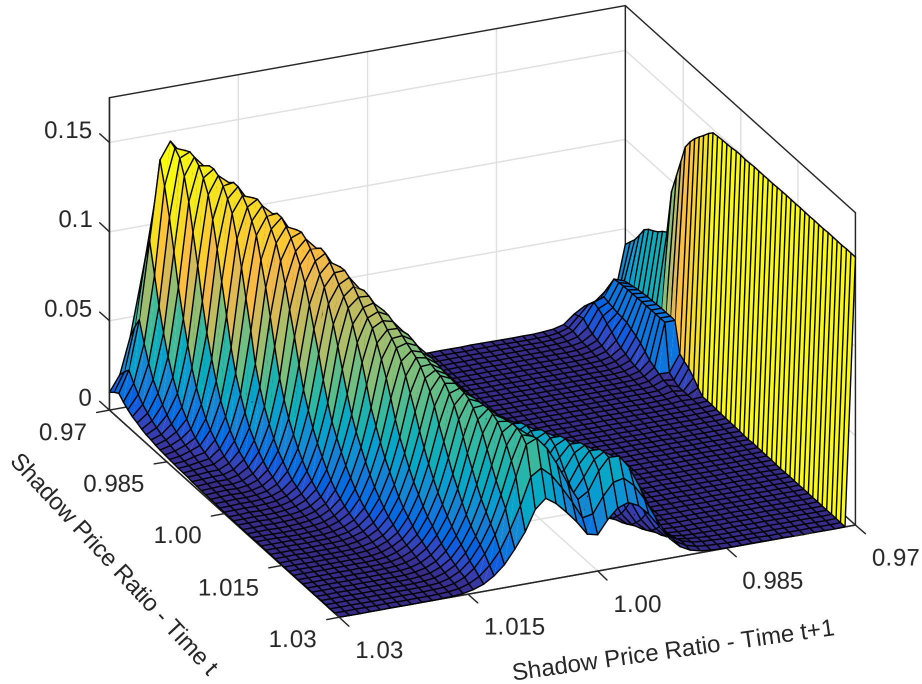

Figure 3: **Transition probabilities of the shadow price ratio** $R_1/(R_1+R_2)$. The figure shows the probability of a value of the shadow price ratio at time $t+1$ conditional on a value of it at time t, with, at both times, the probability being integrated over the remaining state variables. The figure is based on the numerical illustration described in Section 1.4 and drawn for the case in which Trader 1 is currently in his high endowment state. The trading fee is set at 3%. Thus, a ratio of 1.03 means that Trader 1 buys, and a ratio of 0.97 means that he sells.

4.2 Average Effects of Trading Fees

We now display some summary statistics showing the average effect of trading fees on the equilibrium. Our first order of business is to illustrate the statements we have already made in Remark 1 and in Propositions 1 and 2.

4.2.1 Consumption

As discussed in Remark 1, in the presence of trading fees, the traders face a trade-off between smoothing consumption and smoothing holdings, with the latter being due to the desire to reduce trading fees. We now display the endogenous consumption choices of the two traders for the numerical illustration.

Figure 4: **Optimal consumption behavior and welfare loss**. Panels (a) to (c) show the conditional consumption growth volatility, conditional consumption growth correlation and welfare changes of the two traders for different levels of trading fees. Welfare is expressed in terms of an equivalent permanent drop in output for the no-fee case. The figure is based on the numerical illustration described in Section 1.4 and averages are computed across 500,000 simulation paths. All curves are bracketed by dotted lines showing the two-sigma confidence intervals for the estimate of the mean.

Panel (a) of Figure 4, plots, against the rate of trading fees, the average conditional volatility of individual consumption growth. Panel (b) plots their correlation.[32] In summary, we have the following proposition:

Proposition 4 *Trading fees have the effect of increasing the volatility of the consumption of both traders and of reducing their correlation.*[33]

In Panel (c) of Figure 4, we also document the impact of trading fees on traders' welfare, measured by the permanent drop in output in the frictionless economy that would lead to the same welfare as in the economy with trading fees. In general, higher trading fees lead to a reduction in the traders' welfare. For example, the equilibrium with a trading fee of 3% is equivalent in terms of welfare to a 0.2% permanent drop in output for the frictionless economy even though there is no loss of aggregate consumption (such as would occur in the presence of deadweight costs). For comparison, Barro (2009) reports welfare gains, in a model without habit formation, of 0.73% to 1.65% for eliminating *all* business cycle risk, that is, an output volatility of zero.

4.2.2 Asset Prices

In Section, 3.2, we had derived analytical expressions that allow for a comparison between securities prices with and without trading fees. In the following, we quantify the impact of trading fees for our numerical illustration.

In general, trading fees on the stock have two effects. First, they increase the risk associated with holding the stock, because a trader might not be able to resell the security in the next period, causing the traders to reduce their demand. This will lead to a price discount for the stock that is subject to trading fees. It is a *liquidity effect.* Second, fees increase the traders' consumption growth volatility, which will affect the discount rate, and, thus, all traded securities. In particular, the effect of the increased volatility of consumption, resulting from the trading fees, can be likened to the effect of increased volatility resulting from more volatile endowment shocks in the absence of trading fees, which would be *a*

[32] Recall that, even at level zero of trading fees, the market is incomplete. That is why the correlation of consumption is never equal to 1.

[33] While this proposition is fully in line with microeconomic theory, it is based on numerical experiments.

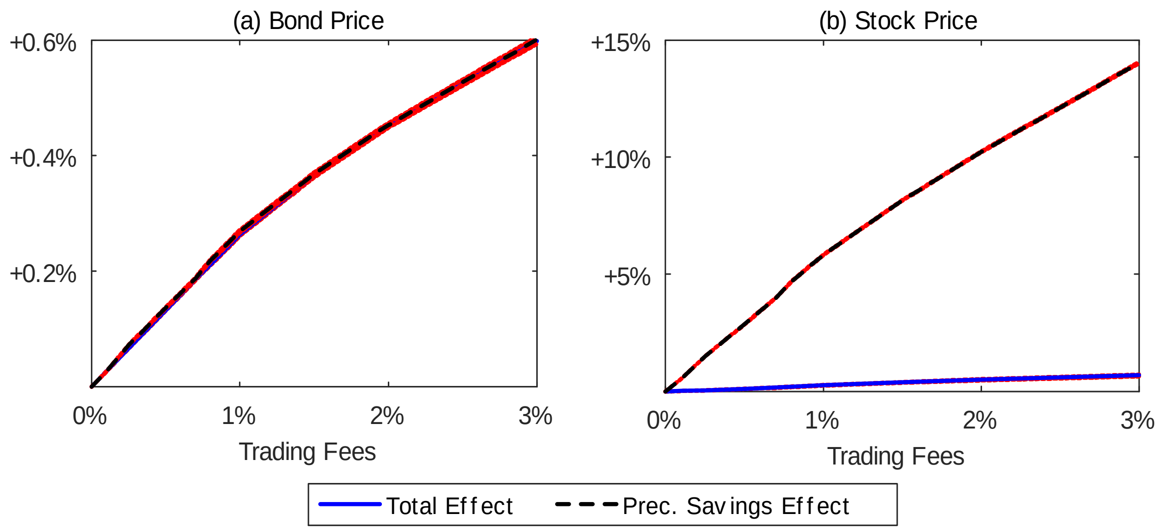

Figure 5: **Asset prices**. The solid lines in Panels (a) and (b) show the bond and stock price for different levels of the trading fee, respectively. The dashed curve shows the prices for frictionless economies with the same consumption growth volatility as for the case with trading fees. The figure is based on the numerical illustration described in Section 1.4 and averages are computed across 500,000 simulation paths. All curves are bracketed by dotted lines showing the two-sigma confidence intervals for the estimate of the mean.

precautionary-savings effect. It is well known that, with positive prudence, this effect encourages saving, brings down the rate of interest, and, all else equal, reduces the discount rate and increases asset prices. The actual variation in the price will be determined endogenously by the interaction of the two effects.

To illustrate the impact of the two effects, Figure 5, Panels (a) and (b) show the bond and stock prices for different levels of trading fees. Moreover, the dashed curves in the figure depict the prices that arise in a frictionless economy in which traders have the same consumption growth volatility as in the corresponding economy with trading fees, thereby capturing exclusively the precautionary savings effect.[34]

Panel (a) shows that for the bond, which is not subject to trading fees, the precautionary savings effect *fully explains* the change in price. Particularly, the bond price is increasing in the trading fee on the stock, due to the lower discount rate. This makes sense intuitively, as the liquidity effect is not present for the bond. Interestingly, the price of the stock is also

[34]This is achieved by slightly modifying the volatility of the endowment shocks.

(slightly) increasing in the trading fee.[35] As Panel (b) shows, the precautionary savings effect alone would lead to a strong increase in the price of the stock. In contrast, the liquidity effect leads to a price discount that counteracts and almost offsets the precautionary savings effect. Particularly, the downward difference between the solid and the dashed curves is accounted for by *the drop in the correlation between individual consumptions*, which implies a drop in the correlation between the aggregate dividend and individual consumption.

Proposition 5 *The greater consumption volatility that arises with trading fees drives precautionary savings higher. This lowers the interest rate, which tends to raise asset prices. For securities subject to trading fees, this effect is counteracted by the drop in the correlation between dividends and individual consumption, which tends to reduce asset prices.*

4.2.3 Trading Strategies

Next, we examine the changes in trading strategies induced by trading fees. Similar to the discussion in Section 1.4, we focus on the first trader's trading strategies in reaction to an endowment shock. In particular, Panels (a) and (b) of Figure 6 depict the (dollar) change in the stock holdings as well as in the bond holdings (the value of the bond purchased or sold minus the redemption value of the bond having matured; see footnote 17), conditional on the realized endowment share for Trader 1 and normalized by aggregate output.

The intercepts of the curves, for zero trading fees, are identical to the numbers reported in Table 1. For instance, in case of a high endowment share and no trading fees, Trader 1 invests 0.1046 units into the stock and borrows 0.0176 units through the bond. As trading fees increase, the change in the stock holdings is gradually reduced (Panel (a)). The change in the bond holdings, shown in Panel (b), is more striking. Whereas, at zero trading fees, when receiving a high endowment, the trader borrows against future endowments for the purpose of leveraging his investment into the stock; he gradually gives up this strategy when fees are larger (approximately greater than 0.625%) and starts using the bond, on which there is zero fee, as the primary investment vehicle, that is, the bond is used as a "substitute."

Proposition 6 *There exists a level of trading fees below which the bond, which can be traded*

[35] Vayanos (1998) has noted that prices can be increased by the presence of transactions costs. Gârleanu (2009) draws a similar conclusion in a limited-trading context. However, in both of these papers, the rate of interest is exogenous.

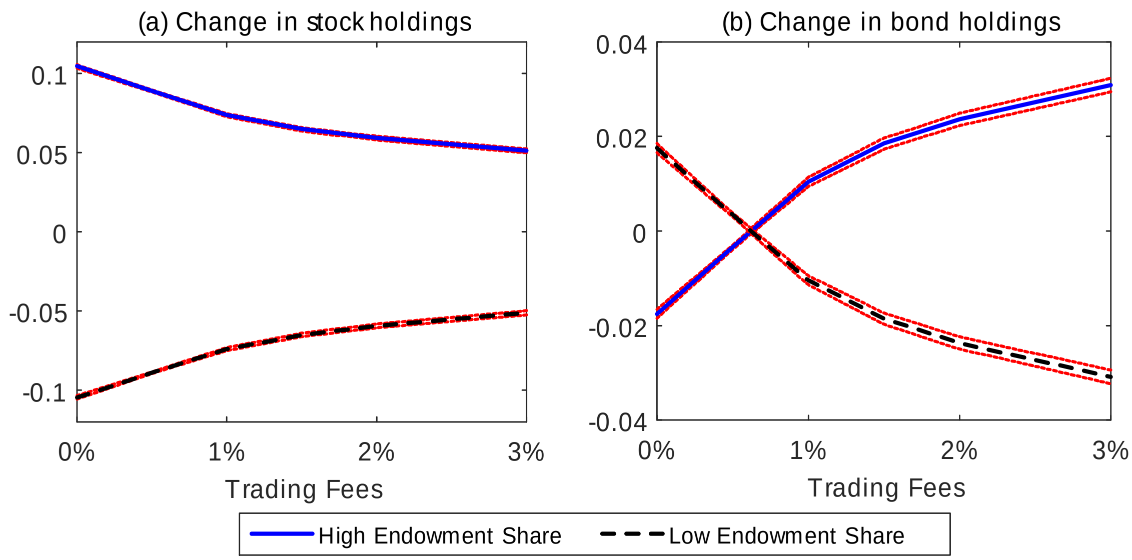

Figure 6: **Trading strategies**. Panels (a) and (b) depict the (dollar) change in the stock holdings as well as in the bond holdings (the value of the bond purchased or sold minus the redemption value of the bond having matured), conditional on the realized endowment share for Trader 1 and normalized by aggregate output. The figure is based on the numerical illustration described in Section 1.4, and averages are computed across 500,000 simulation paths. Curves are bracketed by dotted lines showing the two-sigma confidence intervals for the estimate of the mean.

without fee, serves to enhance the investment into the stock, and above which it partially replaces the investment into the stock as the means to optimize consumption over time.[36]

4.2.4 Rates of Return

The same pricing mechanism reported in Section 4.2.2, as it affects rates of return, is illustrated in Figure 7. Particularly, the figure depicts the effect of trading fees on the return-generating processes and, for comparison, the rates of return for economies with an increase in endowment volatility that artificially mimics the consumption risk added by trading fees (dashed line).

As expected, the rate of interest is reduced due to the precautionary savings effect, matching the bond price result. In contrast, the expected stock return is basically left unchanged by trading fees, while it would be reduced by the precautionary savings effect by

[36] This proposition is based on numerical experiments and is really a conjecture.

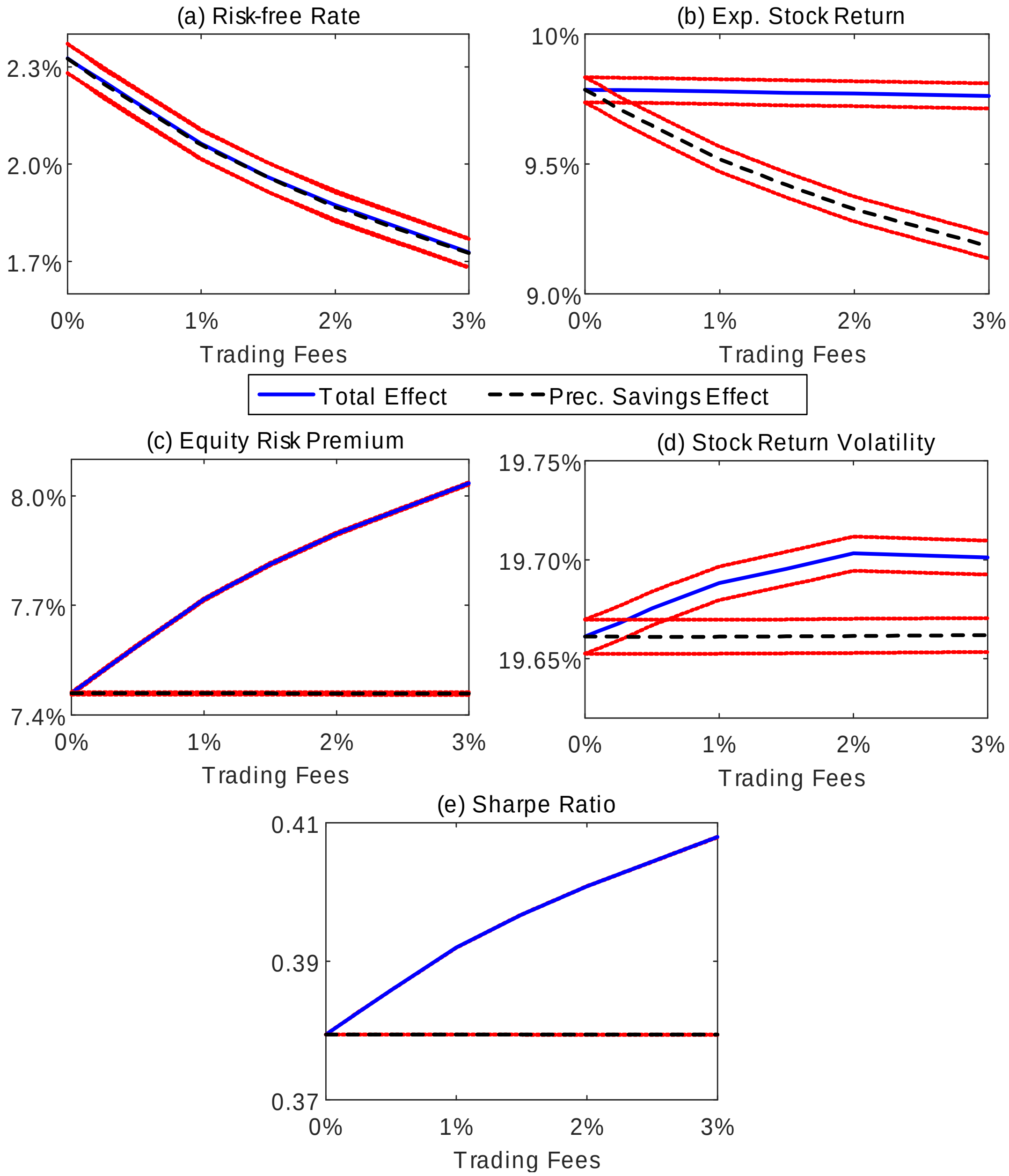

Figure 7: **Rates of Return.** Panels (a) to (e) show the risk-free rate, the conditional expected stock return, the conditional equity premium, the conditional volatility and the conditional Sharpe ratio of equity, respectively, for different trading fees. The solid curves represent averages resulting from trading fees, and the dashed curves show the precautionary savings effect created by endowment shocks that would induce the same consumption volatility as do trading fees. The figure is based on the numerical illustration described in Section 1.4, and averages are computed across 500,000 simulation paths. All curves are bracketed by dotted lines showing the two-sigma confidence intervals for the estimate of the mean.

exactly the same amount as the reduction of the rate of interest. The effect on the equity premium follows; it is increased quite markedly by trading fees. The volatility of stock returns is increased somewhat by fees, in line with the empirical findings of Hau (2006) and, finally, the net effect on the Sharpe ratio is an increase. As shown, these latter quantities would be unchanged by an increase in endowment risk capturing the precautionary savings effect.

In the next section, we decompose the rates of return into premia and, returning to dynamics, show how the premia behave over time.

5 Applications

We now develop three applications of our model. First, we study the pricing of liquidity risk, particularly, the consumption-CAPM (CCAPM) that arises in the presence of trading fees and endogenous trading. Second, we consider an extension to three traders. Third, we study the reaction of asset prices to shocks in models with frictions.

5.1 The Pricing of Liquidity Risk

In the equilibrium with trading fees, the capital-asset pricing model is described by equation (17). It is specific to each trader; we make no attempt at aggregation across traders.[37] This expression can be used to show how, in the presence of trading fees, various premia arise, relative to the classic CCAPM.

5.1.1 Deviations from the classic Consumption-CAPM

Define the gross rate of return on asset i as

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

and, for simplicity, assume that the first security, $i = 1$, is a risk-free bond that is not subject to trading fees. Thus, $r_{t+1,1}$ is conditionally riskless at time t. A CCAPM in the presence of

[37] The CCAPM could be aggregated across traders, using weights of our choice. But, even in the absence of frictions, *aggregate consumption* cannot become exactly the basis for pricing in discrete time. Traders being symmetric, however, the behavior of their risk premia is symmetric around the center of the state space.

trading fees can then easily be derived from equation (17):

$$\mathbb{E}_t\left[r_{t+1,i}\right] = r_{t+1,1} - \mathrm{cov}_t\left(r_{t+1,i}, \frac{\phi_{l,t+1}}{\mathbb{E}_t\left[\phi_{l,t+1}\right]}\right) + \mathbb{E}_t\left[\tau_{l,t+1,i}\right] + \mathrm{cov}_t\left(\tau_{l,t+1,i}, \frac{\phi_{l,t+1}}{\mathbb{E}_t\left[\phi_{l,t+1}\right]}\right), \quad (20)$$

where $\tau_{l,t+1,i}$ denotes the change in liquidity, defined as:

Definition 4 (Liquidity change)

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

Note that $1 - R_{l,t+1,i}$ is a shadow trading fee rate applying to asset i at time $t+1$, from the point of view of Trader l, so that $(1 - R_{l,t+1,i}) \times S_{t+1,i}$ is a future shadow dollar amount of the trading fee. Accordingly, $((1 - R_{l,t+1,i}) \times S_{t+1,i})/S_{t,i}$ is the drag on the asset's rate of return, created by future trading, or the "dollar cost per dollar invested" in the words of Acharya and Pedersen (2005).

Although we refer to the key variables as "liquidity change," notice that the *level* of the liquidity variables R also play a role in the CCAPM deviation. Indeed, supposing it were known that $R_{l,t+1,i,j} = R_{l,t,i}\ \forall j$, then the CCAPM deviation would be equal to

$$(1 - R_{l,t,i}) \times \left\{\mathbb{E}_t\left[\frac{S_{t+1,i}}{S_{t,i}}\right] - r_{t+1,1} + \mathrm{cov}_t\left(\frac{S_{t+1,i}}{S_{t,i}}, \frac{\phi_{l,t+1}}{\mathbb{E}_t\left[\phi_{l,t+1}\right]}\right)\right\},$$

which is still not equal to zero. Specifically, even if the shadow fee rate were known, a stochastic dollar amount of the trading fee would still have to be charged when transacting because the security price itself is uncertain.[38]

The first part of equation (20) is exactly the CCAPM expression of a frictionless market, that is, the risk-free rate minus the covariance between an asset's return and the (normalized) pricing kernel. The remainder captures a deviation from the CCAPM, which we can split into two components:

[38] Only if the liquidity variable tomorrow were fixed at the level 1 at both points in time, $R_{l,t+1,i,j} = R_{l,t,i} = 1$, would the liquidity change be equal to zero.

Definition 5 (Components of CCAPM deviation)

$$\textit{Expected liquidity change} \triangleq \mathbb{E}_t\left[\tau_{l,t+1,i}\right] \tag{21}$$

$$\textit{Liquidity-risk premium} \triangleq cov_t\left(\tau_{l,t+1,i}, \frac{\phi_{l,t+1}}{\mathbb{E}_t\left[\phi_{l,t+1}\right]}\right) \tag{22}$$

The deviation due to the expected liquidity change (21) captures the fact that security i is potentially purchased or sold tomorrow minus current liquidity, $(1 - R_{l,t,i}) \times r_{t+1,1}$, capturing that one dollar of the asset is potentially purchased or sold today against the riskless asset, interest on the fee being included. The deviation in the form of the liquidity-risk premium (22), in contrast, reflects the fact that the dollar fee to be paid upon potential resale is uncertain.

Equation (20) has given us a decomposition 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 traders to trade or not to trade and all quantities have explicit and *endogenous dynamics.*

5.1.2 Endogenous Liquidity-Risk Premia

With these deviations from the classic consumption-CAPM in mind, we can now study endogenous liquidity-risk premia for the numerical illustration introduced in Section 1.4. Specifically, Panels (a) and (b) of Figure 8 depict the total CCAPM deviation as well as the two components separately, for different trading fees.

As expected, the absolute CCAPM deviation is increasing in trading fees. Amihud and Mendelson (1986a) explain that the total premium should be concave in the size of trading fees. For that reason, Amihud and Mendelson (1986b) fit the cross section of equity portfolio returns to the log of the bid-ask spread of the previous period and find a highly significant relationship. Our figure does exhibit that concavity property.[39] For trading fees of 3%, the total deviation reaches about 60bp, which is less than the trading fees themselves, measured as a percentage of the value of each trade. That deviation is too small to be able to account for the several percentage points of returns that empirical researchers commonly attribute

[39] See also Figure 3.1 in Amihud et al. (2005). Note, however, that the analogy is not perfect, as they empirically display a cross-section of firms affected differently by transactions costs and we display a single premium for different levels of trading fees. But the underlying rationale is identical.

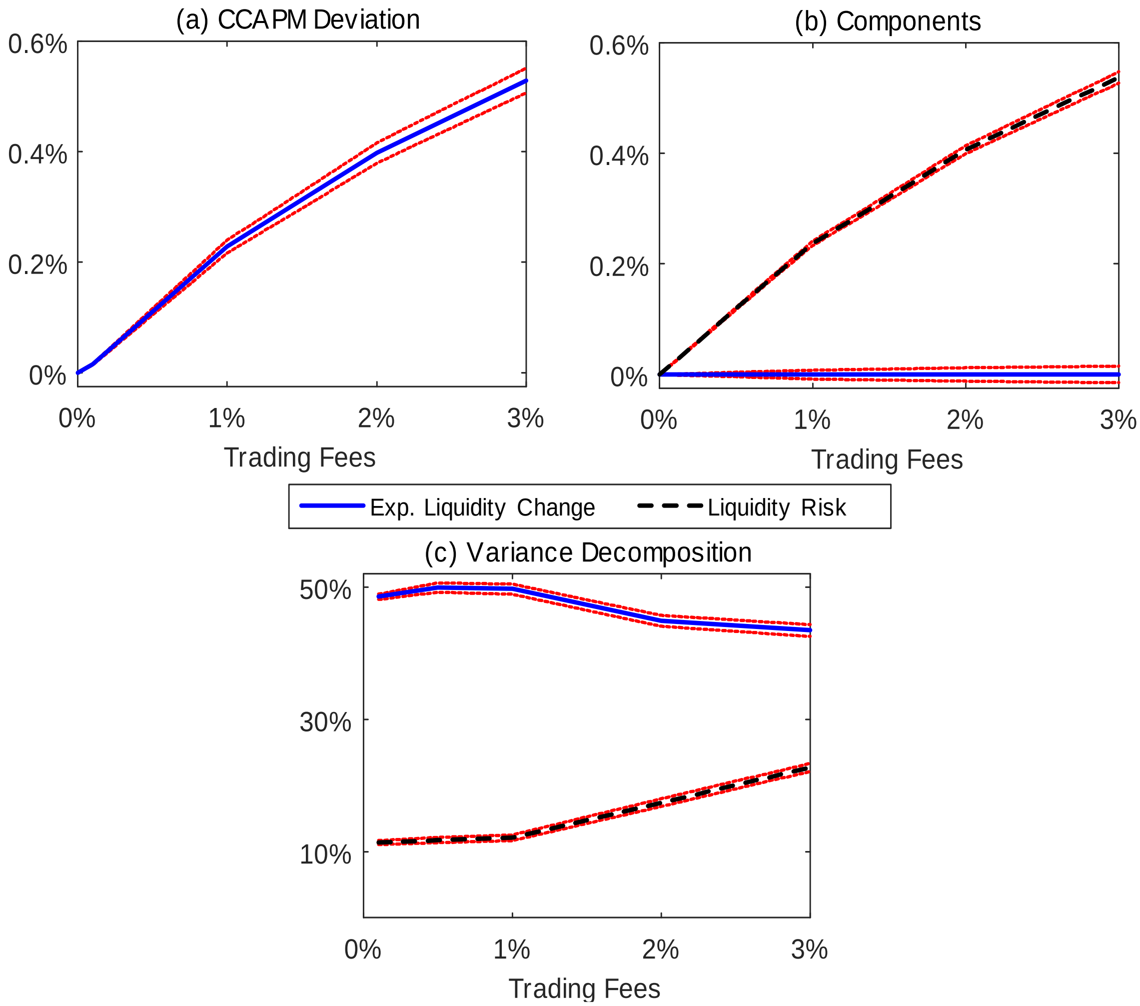

Figure 8: **CCAPM deviations**. Panels (a) and (b) show the unconditional deviations from the classic CCAPM and the two components of the deviation – the expected liquidity change and liquidity risk, for different trading fees. Panel (c) shows the decomposition of the unconditional variance (across paths) of the deviations from the conditional CCAPM. The figure is based on the numerical illustration described in Section 1.4 and 500,000 simulation paths. All curves are bracketed by dotted lines showing the two-sigma confidence intervals for the estimate of the mean.

to liquidity premia.[40] But, it does illustrate the role played by trading frictions when we try to explain empirical deviations from classic asset pricing models. In particular, Panel (b) reveals that *the unconditional average value of the CCAPM deviation is mostly due to the liquidity-risk premium* (as presupposed by Pástor and Stambaugh (2003)), while the expected liquidity change is comparatively small, unconditionally speaking.[41]

Panel (c) of Figure 8 shows, however, that the unconditional variance of the conditionally expected liquidity term is the larger one and is, therefore, mostly responsible for the fluctuations over time of the CCAPM deviations.[42] Thus, it is an important benefit of our model, in which the *liquidity variable is endogenized*, that we can study the variation of each of the terms over time.

This theoretical contrast between the conditional and the unconditional pictures should provide guidance for empirical researchers working on illiquid markets and trying to decide which of the two terms is more important. Bongaerts, De Jong and Driessen (2016), for instance, study 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." Our model shows that, here especially, empirical conclusions could vary a lot depending on conditioning.

5.2 Extension to Three Traders

It might be argued that deviations from the frictionless equilibrium will be reduced when there are more traders in the market.[43] That is, if one trader wants to trade, he is more likely to find a counterparty to his trade when there are many traders. The comparative dynamics of trades between economies with two and three traders can serve to cope with that issue. This is the first reason for which we now study an extension to the case of three traders. The second reason is that we would like to have an answer to a tantalizing question: will

[40] Furthermore, the terms being of opposite signs for the two traders, their values would be even smaller in any CCAPM that would be somehow aggregated across traders.

[41] The term "unconditional mean" is used here for the first time. It has the same meaning as the term "average (across paths)" that we have used so far. We alter the language slightly at this point in order to conform with the distinction, which is traditional in the asset-pricing literature, between tests of the CCAPM in its "unconditional" vs. its "conditional" form.

[42] We are grateful to Luboš Pástor for suggesting this distinction to us.

[43] We are thankful to one referee for doing just that.

most trades be bilateral trades – an order matching another – or will most be trilateral and "centralized"?

When extending the economy to three traders, we keep the exogenous process for aggregate output unchanged and continue to assume that the three traders are symmetric.[44] Because there are now three, rather than two, traders, individual endowments are no longer perfectly negatively correlated. Instead, individual endowments have an idiosyncratic component.[45]

Panel (a) of Figure 9 shows that, for all levels of the trading fee (including zero fees), trading volume for the stock is higher in the three-trader economy than in the two-trader economy. Naturally, as the trading fees now impact a higher volume, *each trader's welfare loss due to trading fees is actually increased* when three symmetric traders are present (Panel (b)). For example, while the two-trader equilibrium with a trading fee of 3% was equivalent in terms of welfare to a 0.2% permanent drop in output for the frictionless economy, with three traders it is equivalent to a drop of 0.35%. Intuitively, while with more traders in a market, there are more people potentially ready to trade, there are also more people who need to trade (for partially idiosyncratic reasons). In summary, although trading takes place over a wider part of the state space, *it is not true that the effect of trading fees goes away as one increases the number of traders,* as long as it is increased in a way that preserves symmetry between traders (with attendant rising idiosyncratic risk).[46]

While it is true that each trader now faces two candidate trading counterparts, the counterparts' endowments are now less correlated with his own than before. While the first effect increases the probability of stock trading, the second effect reduces it. Panel (c) shows that the first effect dominates and, in line with the increased trading volume, the probability of a trade in the stock market occurring increases with the introduction of a third trader. Interestingly, Panel (d) shows that even though the probability of *any* trade occurring is higher in the three-trader economy, the average time each trader waits for a trade in the stock is increased when passing from two to three traders.

[44] For more details, cf. Appendix A.2.

[45] We emphasize at the outset that of all the two-trader, homogeneous-probability beliefs settings one could have assumed, the one we have picked was such as to generate *a low frequency of no trading*, because the endowment shocks of the two traders are perfectly negatively correlated.

[46] It would be conceivable to take that experiment to the limit of an infinite number of traders. However, that would require a completely different kind of algorithm.

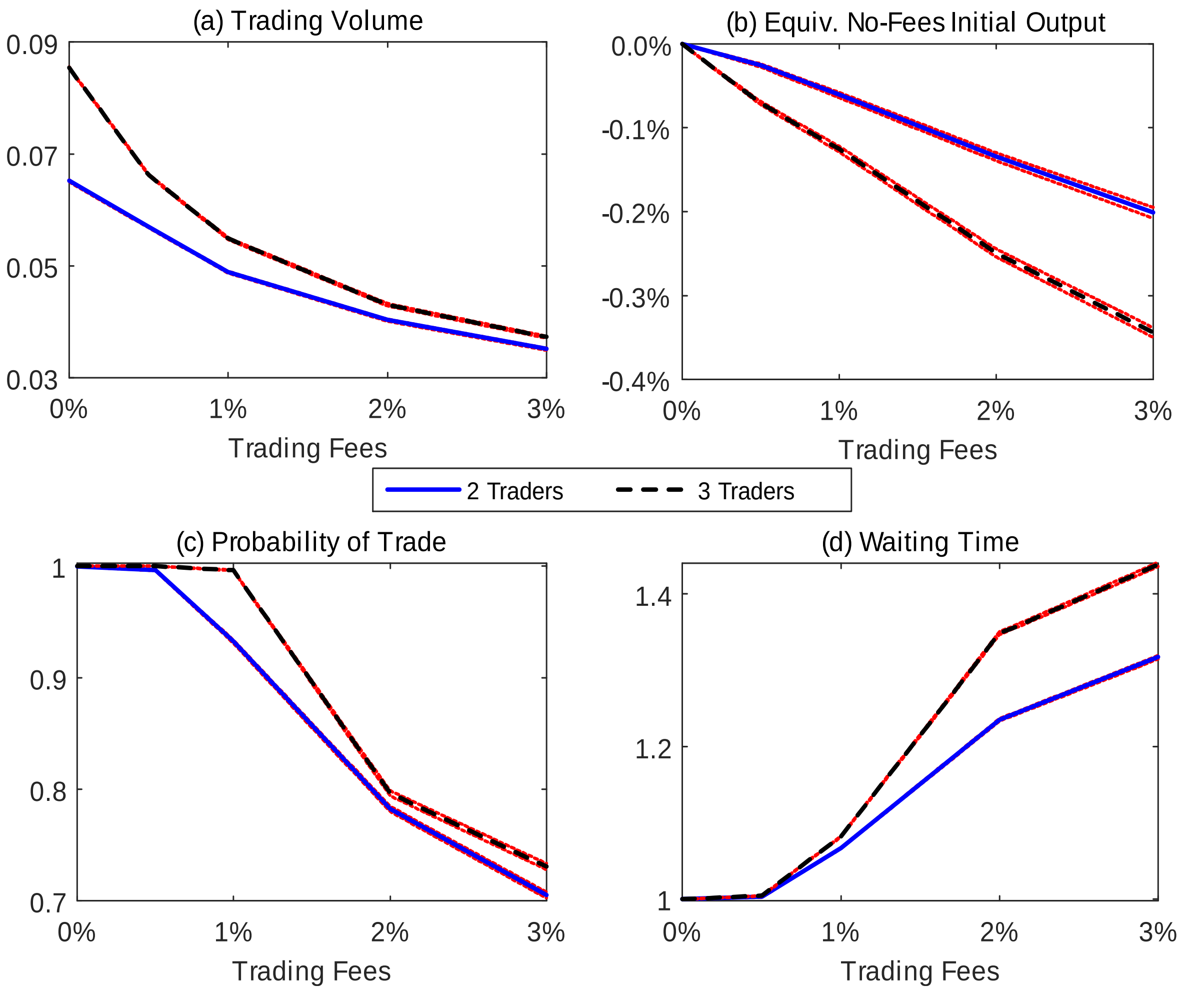

Figure 9: **Two- vs. three-trader economy**. Panels (a) to (d) show the trading volume for the stock, welfare changes, the probability of a trade in the stock market occurring and the waiting time between stock trades, for two- vs. three-traders economies and different trading fees. Welfare is expressed in terms of an equivalent permanent drop in output for the corresponding no-fee economy. The figure is based on the numerical illustrations described in Appendix A, and averages are computed across $500,000$ simulation paths. All curves are bracketed by dotted lines showing the two-sigma confidence intervals for the estimate of the mean.

	Minority Threshold (as % of total trade volume at path)										
Fee	0%	2.5%	5%	7.5%	10%	12.5%	15%	17.5%	20%	22.5%	25%
0.0%	0.00	0.00	0.00	0.01	0.04	0.10	0.18	0.29	0.44	0.68	1.00
0.5%	0.01	0.01	0.02	0.03	0.04	0.08	0.14	0.25	0.40	0.64	1.00
1,0%	0.25	0.25	0.26	0.28	0.31	0.36	0.42	0.49	0.59	0.75	1.00
2.0%	0.37	0.37	0.39	0.41	0.44	0.47	0.51	0.56	0.63	0.74	1.00
3.0%	0.43	0.43	0.45	0.48	0.51	0.55	0.59	0.64	0.70	0.79	1.00

Table 2: **Trade matching.** The table shows the relative frequency of paths for which the minority trade is below a given threshold, defined as a percentage of total volume at the simulated path – for different levels of trading fee. The table is based on the numerical illustration with three symmetric traders, as presented in Appendix A.2, and 500,000 simulated paths.

These two observations can only be reconciled if, in the three-trader setting, a sizeable fraction of trades in the stock market take place between two traders only – a bilateral trade. Specifically, when three traders trade, the trade imbalance between the two largest traders is balanced by the smallest one, which might be called the "minority trade." Table 2 shows the relative frequency for which the minority trade in the stock market is below a given threshold, defined as a percentage of total stock volume at the simulated path – for different levels of the trading fee. In particular, for a threshold of zero, it reports the frequency of bilateral trades, in which the orders of *two* traders only match exactly. While for a fee level of 1%, a quarter of the trades are purely bilateral, the frequency increases to 43% for trading fees of 3%. After an extension to more than three traders, these numbers could be confronted with the actual market-clearing data that are used in the empirical Microstructure literature.

5.3 Slow-Moving Investment Capital

Duffie (2010) gives numerous examples – with supporting empirical evidence – of situations in which investment capital does not adjust immediately and rather seems to move slowly toward profitable trades. That is, when a shock occurs, the price of a security reacts first before the quantities adjust. When they do, the price movement is reversed. Duffie's examples include additions and deletions from the S&P 500 index, arrival of a new order in the book, natural disasters impacting insurance markets, defaults affecting CDS spreads, and issuance of U.S. Treasury securities affecting yields, as well as many other "price-pressure" situations.[47]

One may think of several approaches to the modelling of slow-moving capital. Duffie himself offers as an illustration a model in which the attention of traders is limited. In a similar spirit, Duffie and coauthors (2002, 2007, 2012, 2014) assume that trading requires the physical encounter of one trader with another and that searching for encounters is costly or requires some time. In the Microstructure literature (which does not provide a general-equilibrium model), imperfect, non-competitive intermediation or asymmetric information is the reason for the slow movement of capital. Here, we offer a third, more basic microfoundation for sluggish capital. That is, in our model, capital moves slowly simply because the financial industry operates at a cost, which can be interpreted as a trading cost or fee. As compared to limited attention, our theory has the advantage of simplicity and absence of irrationality. It only makes use of a very basic, in fact, classic, economic principle, namely that access to a service is not free.

In Duffie (2010), in response to an aggregate supply shock, capital moves slowly because some traders are inattentive at the time of the shock, so that a "thin subset of traders [has] to absorb the shocks." Only when the inattentive traders re-enter the market, the price movement is reversed. He displays the effect in the form of impulse-response functions for the prices of securities.[48] In the following, we evaluate similarly the equilibrium price response to traders' endowment shocks and compare the price responses for the case of trading fees and the case of stochastic inattention.

In most published work, an impulse-response function is defined as the path followed by an endogenous variable after an exogenous shock of arbitrary size occurs at a specific time, *followed by a complete absence of shocks.* That is, after the shock, the exogenous variables of the economy remain at the same level as they are directly after the shock. In essence, the economy becomes deterministic. However, the probability of occurrence of that path is

[47] See the discussion in Duffie (2010) and the references therein.

[48] The theoretical literature on infrequent trading is burgeoning. Bacchetta and van Wincoop (2010) calibrate a two-country model in which agents make infrequent portfolio decisions. Chien, Cole and Lustig (2012) set up an equilibrium model in which a large mass of investors do not rebalance their portfolio shares in response to aggregate shocks. Hendershott, Li, Menkveld, and Seasholes (2014) expand the Duffie (2010) slow-moving capital model to analyze multiple groups of investors. Rachedi (2014), as Peress (2005) had done, introduces an observation cost in a production economy with heterogeneous agents, incomplete markets, and idiosyncratic risk. Bogousslavsky (2016) shows that inattention can explain return autocorrelation patterns for intraday returns.

equal to zero. Therefore, it is not representative of what one would observe if the economy is treated as an ongoing entity. A different definition of an impulse-response function is called for, to reflect the concept of a *shock occurring along the way.* In Appendix F, we explain the concept of an impulse-response function that is adapted to an economy with ongoing shocks. Specifically, one has to compare two conditionally expected paths depending on the shock at a given impulse time.

To get as close to empirical work as possible, we seek to have available *a transaction price at all times.* For that reason, we rely on our three-trader model. First, we consider the case in which one of the three traders only (Trader 1) has to pay trading fees, while the other two do not face trading fees. Then, we consider the case in which one of the traders (Trader 1) might become (stochastically) inattentive, while the other two are always free to trade. In both settings, the traders 2 and 3 effectively trade all the time, delivering a transaction price.

5.3.1 Response to an Endowment Shock: the Case of Trading Fees

The impulse responses are shown in Figure 10 for the two alternative impulse response functions. By way of benchmark, the price response in a frictionless market is also shown (as the solid line) and is perfectly flat in both cases because the three traders are free to trade and are able to stabilize the price. This is true irrespective of the fact that the endowment impulse is followed by other shocks.

In contrast, in the case of 2% trading fees, relative to the frictionless price, the stock price is depressed by about 30 basis points when the fee-paying trader receives a positive impulse. Intuitively, in the absence of a trading fee, he would have invested (at least) some of his endowment income into the stock. Since he does so in a smaller amount in the presence of trading fees, the average price is lower. Only over time, the *price reverts back.*

Note that it is a common belief in the profession that trading fees could not produce such a price reversal, because fee-paying traders react instantaneously, albeit in smaller quantities. For example, Duffie (2010) writes:

> At the time of a supply or demand shock, the entire population of investors would stand ready to absorb the quantity of the asset supplied or demanded, with an excess price concession relative to a neoclassical model that is bounded

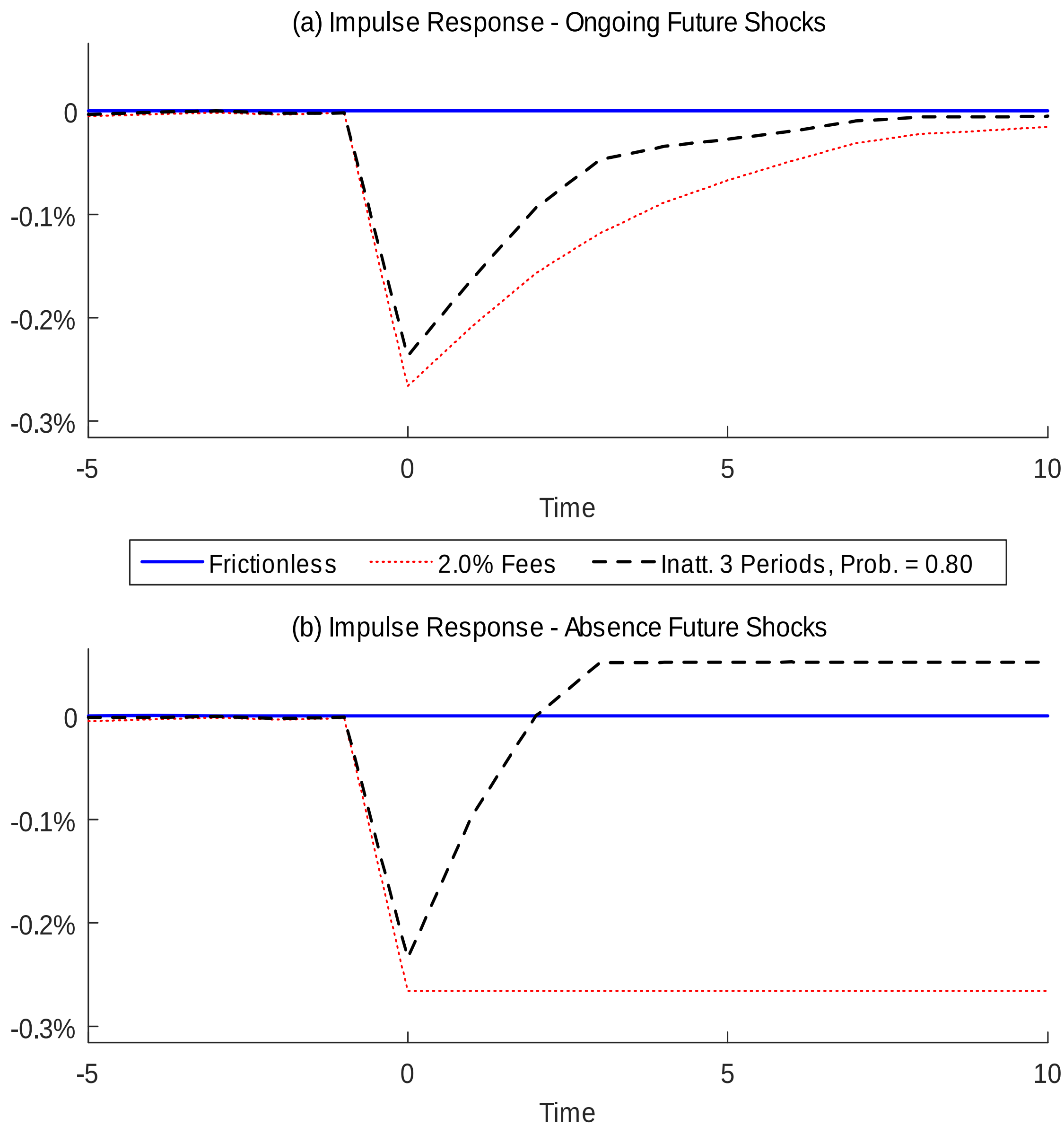

Figure 10: **Impulse-response functions**. The figure shows the difference in the price of the stock between two sets of paths, normalized by the stock price in the frictionless economy. The first set of paths is selected conditional on a positive endowment shock for Trader 1 at the impulse time, and the second set is selected conditional on a positive endowment shock for Trader 2 at the impulse time. Panel (a) depicts the impulse response for the case of ongoing future shocks, i.e., it compares *conditionally expected paths* depending on the shock at a given impulse time. Panel (b) depicts the impulse response in the absence of future shocks. The figure is based on the numerical illustrations described in Section 5.3.

by marginal trading costs. After the associated price shock, price reversals would not be required to clear the market.

It is true that, when the shock hits, all traders adjust immediately, and less so than they would in the absence of fees. Then, with the common concept of impulse response criticized above, there is no need for further adjustment and there is no reversal, since there is no more shock after the one shock of the impulse. This is illustrated in Panel (b), which shows an impulse response function for a shock, followed by a complete absence of future shocks.

However, with our definition, which is closer to empirical work because it compares conditionally expected paths, after the impulse, the effect gradually disappears: *there is a reversal* (Panel (a)).[49] Particularly, when the shock hits, all traders adjust immediately, but on an equilibrium path with ongoing shocks, the traders will also react later on. That is true *because of hysteresis*. Indeed, the impulse has moved the fee-paying trader closer to a trade boundary, so that when later shocks arrive in the same direction, *he will act*, more so than he would have acted in the absence of the impulse. This causes the price reversal.

5.3.2 Response to an Endowment Shock: the Case of Inattention

Consider now the case in which Trader 1 may randomly become inattentive for n periods. Note that the trader still optimizes his decisions intertemporally when he is attentive, and rationally anticipates becoming inattentive again.[50] As in Duffie (2010), we have to choose the probability of becoming inattentive and the length n of periods of inattention. For the illustration in Figure 10, we set the probability of the trader becoming inattentive to 0.8 and assume that he becomes inattentive for three periods. With that choice, the two economies – the one with trading fees and the one with stochastic inattention – have about the same aggregate trading volume.

Panel (a) displays the impulse response function under limited attention for the case with

[49]Note, the paths were segregated based on draws from a uniform distribution, which is not persistent. See Appendix F.

[50]We solve the model recursively, using as an additional endogenous state variables last period's stock holdings of the potentially inattentive trader. For each combination of the state variables, we solve a system of equations similar to the one without trading fees with the small difference that, in case the potentially inattentive trader is inattentive, he does not agree with the other traders on the price of the stock, because he cannot trade the stock. We keep track of his "private valuation," which we carry backwards until the trader is attentive again. When that happens, all traders agree on the price of the stock.

ongoing future shocks. Here again, the reaction of the price to the impulse is immediate and approximately equal to 25 basis points and we observe a slow reversal as in Duffie (2010).

Note that for inattention, slow price reversal can also be observed in the case of a complete absence of future shocks (Panel (b)). Even if there are no future shocks, the price is changing because the inattentive trader re-enters the market.[51]

In summary, it is clear that in the case of ongoing shocks the time path is extremely similar whether we consider the trading-fee model or the limited-attention model. As far as responses to endowment shocks are concerned, they are empirically indistinguishable.

6 Conclusion

In this paper, we develop a general-equilibrium model of a financial friction to describe the properties of financial-market equilibrium, that is, trading strategies, asset prices and returns, in the presence of trading fees. We define a concept of Walrasian equilibrium for this market and invent an algorithm that delivers an exact numerical equilibrium. The algorithm synchronizes like clockwork the traders in the implementation of their trades and allows us to analyze the way in which trades take place and in which prices are formed and evolve.

We analytically compare the equilibrium securities prices in the presence of trading fees to those without trading fees as well as to the traders' private valuations, and explain how the gap between them triggers trades.

Using a numerical setting with two traded securities – a bond and a stock – we then study the impact of trading fees quantitatively. We find that the trade-off between smoothing consumption and smoothing holdings, leads to a higher volatility of individual consumption, a lower correlation between individual consumptions and a drop in welfare. The prices of securities are actually increased slightly by the presence of fees. The price increase of the bond is related to the increased volatility of individual consumption, which produces a drop in the rate of interest because of precautionary saving. The same effect would lead to a marked increase in the price of stock. However, the effect is largely offset by an illiquidity

[51] The small "over-reaction" of the price after the reversal is due to changes in traders' consumption (wealth) shares, created by the impulse. Particularly, if Trader 1 is inattentive and therefore unable to react to the impulse, the impulse will trigger a permanent reduction in his consumption (wealth) share. This dislocation leads to a higher price because it pushes him closer to the habit subsistence level. In a model with many traders this effect would be negligible.

discount because the stock's liquidity is now endogenously varying over time. As for rates of return, we show that the equity premium, stock return volatility, and the Sharpe ratio of equity are increased.

Finally, we develop three applications of our model. First, we obtain endogenously the behavior over time of the various components of a CCAPM that incorporates frictions. That is, we identify the risk factors and display their relative sizes and movements over time. Second, we study an extension of our model to the case of three traders and show that the impact of trading fees does not decrease with the introduction of more (symmetric) traders. Third, we compare the impulse responses of this model to those of a model in which trading is infrequent because of trader inattention. Contrary to what has been asserted by some authors, limited-attention models and trading-fees models produce very similar price responses if one considers the realistic case of an economy with ongoing future shocks. Thus, they are, so far, equally good contenders as representations of slow-moving capital.

The model generates a rich set of predictions, which are empirically refutable, as we verify quantitatively that the bid-ask midpoint can be used as a proxy to study the empirical implication of our model. For example, similar to our results, many empirical papers have found a "liquidity premium" for less liquid stocks, though a quantitatively bigger one. The theoretical nature of our paper allows to comment on the adequacy of extant empirical tests of CCAPMs that include a premium for liquidity risk, thereby guiding future empirical work. Another point of contact with empirics is the paper by Hau (2006), who already showed that volatility does increase with trading fees, as predicted by our model. By empirically studying the price reversal for stocks with different levels of liquidity, one should also be able to shed more light on the underlying explanation for price reversal. Particularly, our model predicts that for a more liquid stock the price reverts back at a faster rate.

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 traders would submit limit and market orders. The behavior of the limit-order book would be obtained. This work would be similar to that of Parlour (1998), Foucault (1999), Foucault et al. (2005), Goettler et al. (2005) and Roşu (2009), except that trades would arrive at the time and in quantities of the traders' choice, and would not be driven by an exogenous process.[52]

[52] Recently, 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 Numerical Illustrations

Here, we describe the settings that we use to illustrate the dynamics of equilibrium in the presence of trading fees. Note, they are only meant to illustrate, in a stylized fashion, the workings of the model. They cannot be seen as being calibrated to a real-world economy because we have two (three) traders, not millions; two securities, not tens of thousands; and a trading frequency of one year. Thus, although we incorporate a motive to trade that is present at all times, the volume of trading does not come anywhere close to market data.

A.1 Two Traders

We assume that aggregate output, O_t, follows a binomial tree with expected growth, $\mu_O =$ 1.8% and a volatility, $\sigma_O = 3.2\%$, matching their empirical counterparts. There exist two traders with preferences of the additive external-habit type, implemented as surplus consumption, similar to Campbell and Cochrane (1999):

$$\mathbb{E}\left[\sum_{t=0}^{T} \beta^t \times \frac{(c_{l,t} - h \times C_{t-1})^{1-\gamma}}{1-\gamma}\right],$$

where C_{t-1} denotes aggregate last period consumption. Traders have homogeneous preferences, that is, the same time-preference β, risk appetite γ and habit parameter h.

There exist two securities: $i = 1$, a short-lived riskless security (the "bond") in zero net supply that is not subject to trading fees; $i = 2$, a long-lived claim (the "stock") for which trading entails trading fees. The stock is in unit net supply and pays out dividend $\delta_{t,2}$, which is modeled as a constant fraction, χ, of aggregate output.

The remainder, total output minus dividend, $(1 - \chi) \times O_t$, is distributed as endowments, $e_{l,t}$, to the two traders. Particularly, we assume that the fractions of aggregate endowment, $v_{l,t}$, follow a simple, symmetric two-state Markov chain, with realizations of 62.5% and 37.5%. We set the probability of transitioning from a high (low) state today to a high (low) state in the next period to 0.85. These parameters imply a volatility of 20% for the endowment shocks, which is comparable to the volatility of labor income shocks of about 24% (see Gourinchas and Parker (2002)).

The parameters β, h, γ and χ are chosen to match the empirical risk-free rate, equity risk premium, stock market volatility, and wealth-income ratio. The resulting values are

Moment	Data	Model
Agg. cons. growth mean	1.79%	1.82%
Agg. cons. growth volatility	3.22%	3.26%
Risk-free rate	2.02%	2.32%
Equity premium	6.73%	7.47%
Stock return volatility	18.60%	19.65%
Sharpe ratio	0.36	0.38
Price-dividend ratio	23.75	21.01
Volatility of log P/D ratio	0.32	0.16

Table 3: **Return moments without friction.** The data is based on Campbell (2003) with a sample period spanning from 1891-1998. Consumption growth denotes real per capital consumption growth of non-durables and services for the United States. The stock return data are based on the S&P500 index, and the risk-free rate is based on the 6-month U.S. Treasury bill rate.

$\beta = 0.98$, $h = 0.2$, $\gamma = 7.5$, $\chi = 0.15$. Note that 85% of total output being distributed as endowments, implies that the average wealth-income ratio in our economy is 3.81, comparable to the ratios documented by the Survey of Consumer Finances (2014) of 0.37 (age < 35) to 5.52 (age ≥ 75). The resulting return moments are shown in Table 3, demonstrating that our quantitative experiments are conducted in a realistic financial-market setting.

A.2 Three Traders

The three-trader economy that we study in Section 5.2, is a simple extension of the two-trader economy presented above. Particularly, we again model a symmetric, Markov chain, now with three states to accommodate the third trader, and with endowment share realizations of 45%, 27.5%, and 27.5%. The probability of one trader transitioning from the high endowment share today to the high endowment share in the next period is set to 0.7 and the probability of transitioning to a state in which one of the other two traders receives a high endowment share is set to 0.15 each. On average each trader gets 1/3 of aggregate endowment. Finally, to keep the surplus consumption ratio unchanged, we set the habit parameter to 0.1333.

In the case of the three-trader economy discussed in Section 5.3, we also use a habit parameter of 0.1333. However, to preserve perfect *conditional* symmetry between Traders 1 and 2, we consider a compound of two 2×2 Markov chains for endowments. While in the first chain, Trader 3 gets a share of either 41.68% or 24.98%, in the second chain, the remainder is distributed to the other two traders, with each one getting either 39% or 61% of what is left. Both separate Markov chains are persistent, with the probability of staying in a state being 0.85. Again, on average each trader gets 1/3 of aggregate endowment.

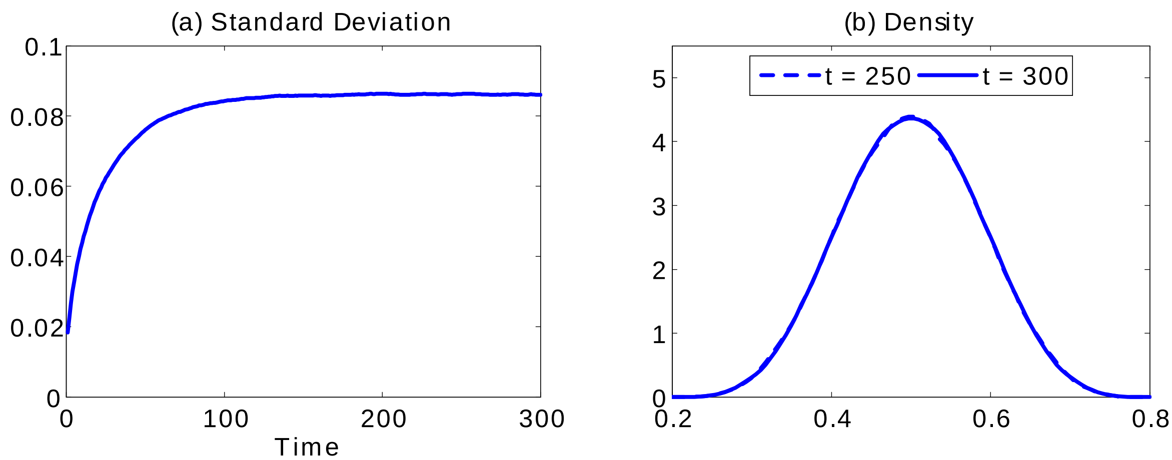

Figure 11: **Dynamics of state variable**. The picture illustrates the existence of a time window starting some time before $t = 250$ within which the probability distribution of the state variable (consumption share) is unchanged. Panel (a) shows the standard deviation of the consumption share of Trader 1 over time. Panel (b) shows the simulated density of the consumption share for two dates. The figure is based on the numerical illustration described in 1.4 and on 500,000 simulation paths.

A.3 The Horizon and the Steady State

For all our numerical illustrations we run the algorithm backward from a fixed horizon date until there is no change in all the functions being carried backward, thereby obtaining an equilibrium of an economy where traders are very long-lived.[53] Besides displaying features that hold for a very long horizon, we also want to make sure that those features do not depend on initial conditions. For that purpose, we simulate the long-horizon economy forward and we keep track of the frequency distribution of the state variable(s) across simulated paths. We only stop the simulations when the distribution of the state variable(s) has converged. For all results reported in the paper, we rely only on dates from this "steady state."

Figure 11 illustrates this for the economy with two traders in the absence of trading fees. The steady-state probability distribution of consumption shares obtains after about 250 periods. After this the distribution does not change anymore. That is, this many years is a long enough "burn-in" history. In this specific case, our results are based on equilibrium quantities at $t = 300$.

[53] The criterion for stopping the backward calculation is the mean absolute relative difference from one time step to the next of all iterated functions. We stop when the value of that criterion is below 0.01%. Further refining the criterion has virtually no effects.

B First-order Conditions

The Karush-Kuhn-Tucker first-order conditions, associated with maximization problem (7) subject to (8) and (9) are

$$
\begin{gathered}
u_l'\left(c_{l,t},\cdot,t\right)=\phi_{l,t} \\
e_{l,t}+\sum_{i=1}^{I}\theta_{l,t-1,i}\delta_{t,i}-c_{l,t}+\zeta_{l,t} \\
-\sum_{i=1}^{I}\left(\hat{\theta}_{l,t,i}-\theta_{l,t-1,i}\right)S_{t,i}\left(1+\lambda_{i,t}\right)-\sum_{i=1}^{I}\left(\check{\theta}_{l,t,i}-\theta_{l,t-1,i}\right)S_{t,i}\left(1-\varepsilon_{i,t}\right)=0 \\
\sum_{j=1}^{K_t}\pi_{t,t+1,j}\frac{\partial J_{l,t+1,j}}{\partial\theta_{l,t,i}}\left(\left\{\hat{\theta}_{l,t,i}+\check{\theta}_{l,t,i}-\theta_{l,t-1,i}\right\},\cdot,e_{l,t+1,j},t+1\right) \\
=\phi_{l,t}\times S_{t,i}\times\left(1+\lambda_{i,t}\right)-\mu_{1,l,t,i} \\
\sum_{j=1}^{K_t}\pi_{t,t+1,j}\frac{\partial J_{l,t+1,j}}{\partial\theta_{l,t,i}}\left(\left\{\hat{\theta}_{l,t,i}+\check{\theta}_{l,t,i}-\theta_{l,t-1,i}\right\},\cdot,e_{l,t+1,j},t+1\right) \\
=\phi_{l,t}\times S_{t,i}\times\left(1-\varepsilon_{i,t}\right)+\mu_{2,l,t,i} \\
\check{\theta}_{l,t,i}\leq\theta_{l,t-1,i}\leq\hat{\theta}_{l,t,i};\mu_{1,l,t,i}\geq 0;\mu_{2,l,t,i}\geq 0 \\
\mu_{1,l,t,i}\times\left(\hat{\theta}_{l,t,i}-\theta_{l,t-1,i}\right)=0;\mu_{2,l,t,i}\times\left(\theta_{l,t-1,i}-\check{\theta}_{l,t,i}\right)=0,
\end{gathered}
$$

where $\phi_{l,t}$ is the Lagrange multiplier attached to the flow budget constraint (8) and $\mu_{1,l,t,i}$ and $\mu_{2,l,t,i}$ are the Lagrange multipliers attached to the inequality constraints (9). The last two equations are usually referred to as the "complementary-slackness" conditions.

Two of the first-order conditions imply that

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

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

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

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

$$\sum_{j=1}^{K_t}\pi_{t,t+1,j}\frac{\partial J_{l,t+1,j}}{\partial\theta_{l,t,i}}\left(\left\{\hat{\theta}_{l,t,i}+\check{\theta}_{l,t,i}-\theta_{l,t-1,i}\right\},\cdot,e_{l,t+1,j},t+1\right)=\phi_{l,t}\times R_{l,t,i}\times S_{t,i}. \tag{23}$$

In order to eliminate the value function from the first-order conditions, we differentiate the Lagrangian with respect to $\theta_{l,t-1,i}$ (invoking the Envelope theorem) and use (23):

$$\begin{aligned}\frac{\partial J_{l,t}}{\partial \theta_{l,t-1,i}} = \frac{\partial L_{l,t}}{\partial \theta_{l,t-1,i}} &= -\sum_{j=1}^{K_t} \pi_{t,t+1,j} \frac{\partial J_{l,t+1,j}}{\partial \theta_{l,t,i}} \left(\left\{ \hat{\theta}_{l,t,i} + \check{\theta}_{l,t,i} - \theta_{l,t-1,i} \right\}, \cdot, e_{l,t+1,j}, t+1 \right) \\ &+\phi_{l,t} \left[\delta_{t,i} + S_{t,i} \times (1 + \lambda_{i,t}) + S_{t,i} \times (1 - \varepsilon_{i,t}) \right] - \mu_{1,l,t,i} + \mu_{2,l,t,i} \\ &= -\sum_{j=1}^{K_t} \pi_{t,t+1,j} \frac{\partial J_{l,t+1,j}}{\partial \theta_{l,t,i}} \left(\left\{ \hat{\theta}_{l,t,i} + \check{\theta}_{l,t,i} - \theta_{l,t-1,i} \right\}, \cdot, e_{l,t+1,j}, t+1 \right) \\ &+\phi_{l,t}\delta_{t,i} + 2\phi_{l,t} \times R_{l,t,i} \times S_{t,i} = \phi_{l,t} \times \left(\delta_{t,i} + R_{l,t,i} \times S_{t,i} \right),\end{aligned}$$

so that the first-order conditions can be written as equations (10) to (16).

C Proof of Proposition 1

At $t = T-1$, the present value of dividends δ for a generic asset from the point of view of Trader l is given by

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

whereas Equation (17) applied to time $T-1$ is

$$R_{l,T-1} \times S_{T-1} = \mathbb{E}_{T-1} \left[\frac{\phi_{l,T}}{\phi_{l,T-1}} \times \delta_T \right] = \hat{S}_{l,T-1}. \tag{24}$$

At $t = T-2$, the present value of dividends is

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

whereas Equation (17) applied to time $T-2$ is

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

where we used equation (24) to replace $R_{l,T-1} \times S_{T-1}$.

By an induction argument one obtains the final result (18).

D Proof of Proposition 2

At $t = T - 1$, the price of a generic asset without trading 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 (17) applied to time $T - 1$ is

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

This can be rewritten as

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

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-trading fees economy, S_{T-1}^{*}, and the stock price in an economy with trading fees, S_{T-1}

$$R_{l,T-1} \times S_{T-1} - S_{T-1}^{*} = \mathbb{E}_{T-1}\left[\Delta\phi_{l,T}\delta_T\right]. \tag{25}$$

At $t = T - 2$, the stock price in an economy without trading fees is given by

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

whereas Equation (17) applied to time $T - 2$ is

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

Replacing $R_{l,T-1} \times S_{T-1}$ with expression (25), this can be rewritten as

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

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

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

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

By an induction argument one reaches the final result (19).

E The Algorithm

As has been noted by Dumas and Lyasoff (2012) in a different context, the system made of (10) to (16) and (3) has a drawback. It must be solved simultaneously (or globally) for all nodes of all times. As written, it cannot be solved recursively in a backward way because the unknowns at time t include consumption at time t, $c_{l,t}$, whereas equation (12) if rewritten as

$$\sum_{j=1}^{K_t} \pi_{t,t+1,j} \times u'_l\left(c_{l,t+1,j}, \cdot, t\right) \times \left[\delta_{t+1,i,j} + R_{l,t+1,i,j} \times S_{t+1,i,j}\right] = \phi_{l,t} \times R_{l,t,i} \times S_{t,i}$$

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 kernel and market clearing conditions, forward in time, and, second, we no longer make explicit use of the trader's positions $\theta_{l,t-1,i}$ held when entering time t, focusing instead on the positions $\theta_{l,t+1,i,j}$ held when exiting time $t+1$, which are carried backward. Regrouping equations in that way, substituting the rewritten definition (6) of the pot ζ and appending market-clearing conditions (3) leads to the equation system of Section E.1.

E.1 A Shift of Equations

An equilibrium can then be calculated by means of a single backward-induction procedure. Given endogenous state variables, which are the dual variables $\left\{\phi_{l,t}, R_{l,t,i}\right\}$, one solves the following equation system. Note that the shift of equations amounts, from a computational standpoint, to letting traders 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, in turn, finance the time-$t+1$ consumption).

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

$$u_l'\left(c_{l,t+1,j}, \cdot, t+1\right) = \phi_{l,t+1,j}.$$

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

$$c_{l,t+1,j} + \sum_{i=1}^{I}\left(\theta_{l,t+1,i,j} - \theta_{l,t,i}\right) S_{t+1,i,j} R_{l,t+1,i,j} = e_{l,t+1,j} + \sum_{i=1}^{I} \theta_{l,t,i}\delta_{t+1,i,j} + \zeta_{l,t+1,j},$$

where the central pot $\zeta_{l,t+1,j}$ is defined in (6).

3. The third subset of equations ("kernel conditions") says that, when they trade, all traders must agree on the prices of traded securities and, more generally, they must agree on the "posted prices" inclusive of the shadow prices R that make units of paper securities more or less valuable than units of consumption.

$$\begin{aligned} &\frac{1}{R_{1,t,i} \times \phi_{1,t}} \sum_{j=1}^{K_t} \pi_{t,t+1,j} \times \phi_{1,t+1,j} \times \left(\delta_{t+1,i,j} + R_{1,t+1,i,j} \times S_{t+1,i,j}\right) \\ = &\frac{1}{R_{2,t,i} \times \phi_{2,t}} \sum_{j=1}^{K_t} \pi_{t,t+1,j} \times \phi_{2,t+1,j} \times \left(\delta_{t+1,i,j} + R_{2,t+1,i,j} \times S_{t+1,i,j}\right) \end{aligned}$$

4. Definitions:

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

5. Complementary-slackness conditions:

$$\left(-R_{l,t+1,i,j}+1+\lambda_{i,t+1,j}\right)\times\left(\hat{\theta}_{l,t+1,i,j}-\theta_{l,t,i}\right) = 0; \tag{26}$$
$$\left(R_{l,t+1,i,j}-(1-\varepsilon_{i,t+1,j})\right)\times\left(\theta_{l,t,i}-\check{\theta}_{l,t+1,i,j}\right) = 0. \tag{27}$$

6. Market-clearing restrictions:

$$\sum_{l=1,2}\theta_{l,t,i}=\bar{\theta}_i.$$

7. Inequalities:

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

This is a system of $2K_t+2K_t+I+2K_tI+2K_tI+2K_tI+I$ equations (not counting the inequalities) with $2K_t+2K_t+2K_tI+2I+2K_tI+2K_t+I$ unknowns $\Big\{c_{l,t+1,j},\phi_{l,t+1,j},R_{l,t+1,i,j},\theta_{l,t,i},$ $\hat{\theta}_{l,t+1,i,j},\check{\theta}_{l,t+1,i,j};l=1,2;i=1,...,I;j=1,...,K_t\Big\}$.[54]

Besides the exogenous endowments $e_{l,t+1,j}$ and dividends $\delta_{t+1,i,j}$, the "givens" are the time-t trader-specific shadow prices of consumption $\{\phi_{l,t};l=1,2\}$ and of paper securities $\{R_{l,t,i};l=1,2;i=1,...,I\}$, 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 $\frac{R_{2,t,i}}{R_{1,t,i}}$ 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 $0\leq\frac{\phi_{1,t}}{\phi_{1,t}+\phi_{2,t}}\leq 1$.[55]

In addition, the given securities' price functions $S_{t+1,i,j}$ and the given future position functions $\theta_{l,t+1,i,j}$ are obtained by backward induction of the previous solution of the above system. All the functions carried backward are interpolated by means of quadratic interpolation based on the modified Shepard method.

Moving back through time until $t=0$, the last portfolio holdings we calculate are $\theta_{l,0,i}$. These are the post-trade portfolios held by the traders 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 $\bar{\theta}_{l,i}$. This is done by solving a time-0 system of equations that consists of the two traders' budget constraints, definitions, complementary slackness, and market clearing

[54] The size of the system is reduced when some securities do not carry trading fees.

[55] In a given exogenous node, the two variables $\phi_{1,t}$ and $\phi_{2,t}$ are one-to-one related to the consumption shares of the two traders, so that consumption scales are actually used as state variables. Consumption shares of the two traders add up to 1 because the trading fees are refunded in a lump-sum.

conditions for time 0. However, the equation system does not contain the kernel conditions, which have already been solved. In essence, it includes all equations from the global system that had not been solved yet because of the time-shift.

E.2 The Interior-Point Algorithm

The system of equations described above can be solved numerically by Newton iterations. However, the iterations can run into indeterminacy because of the KKT complementary-slackness conditions (26, 27), which contain a product of unknowns equated to zero. Indeed, if a Newton step produces, for instance, a value $-R_{l,t+1,i,j}+1+\lambda_{i,t+1,j}$ on the boundary, where it is equal to zero, then the requirement placed on $\hat{\theta}_{l,t+1,i,j} - \theta_{l,t,i}$ drops out of the system and indeterminacy follows. The Interior-Point algorithm is a solution to that problem. It amounts to replacing the above equation system by a sequence of equation systems in each of which the complementary-slackness conditions are relaxed, as shown in equations (15) and (16) where η is a small number, which is made to approach zero as the algorithm progresses. In this way, the indeterminacy is avoided.

F Impulse-response Functions

A new definition of an impulse-response function is called for to reflect the concept of a shock occurring along the way.

We generate 500,000 paths of a simulation, at each point in time drawing three $[0,1]$ uniform random numbers – and transforming them through cumulative probability distributions as needed – to determine (i) whether total output goes up or down, (ii) whether Trader 3 gets a high or a low share (first Markov chain), and (iii) whether Trader 1 or Trader 2 gets a high or a low endowment share. Importantly, note that the draws from the uniform, as opposed to the output and endowment realizations themselves, make up *purely transient processes.*

Then, we segregate the 500,000 paths into two subsets depending on the third draw from the uniform distribution (setting the endowment share between traders 1 and 2) is above or below 0.5 – "the impulse." We compute the average of each of these two subsets of paths and take the difference between them. This is the difference between two sets of paths, both of which are expected conditional on two levels of the impulse. They represent the effect of the impulse that an observer would actually witness on average. Empirical event studies à la Fama, Fisher, Jensen and Roll (1969) plot an average path for cumulative abnormal return (CAR) that is defined exactly the same way.

References

Acharya, V. and L. H. Pedersen, 2005, "Asset Pricing with Liquidity Risk," Journal of Financial Economics, 77, 375-4010.

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 Analyst Journal, 42, 43-48.

Amihud, Y., H. Mendelson and L. Pedersen, 2005, "Liquidity and Asset Prices," Foundations and Trends in Finance, 4, 269-364.

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.

Bacchetta, P. and E. van Wincoop, 2010, "Infrequent Portfolio Decisions: A Solution to the Forward Discount Puzzle," American Economic Review, 100, 870–904.

Barro, Robert J., 2009, "Rare Disasters, Asset Prices, and Welfare Costs," American Economic Review, 99, 243-64.

Bismut, J. M., 1973, "Conjugate Convex Functions in Optimal Stochastic Control," Journal of Mathematical Analysis and Applications, 44, 384-404.

Björk, T., Y. Kabanov and W. Runggaldier, 1997, "Bond Market Structure in the Presence of Marked Point Processes," Mathematical Finance, 7, 211–223.

Bongaerts, D., J. Driessen and F. de Jong, 2016,"An Asset Pricing Approach to Liquidity Effects in Corporate Bond Markets," Review of Financial Studies, forthcoming.

Bouchard, B., 2002, "Utility Maximization on the Real Line under Proportional Transactions Costs," Finance and Stochastics, 6, 495-516.

Bogousslavsky, V., 2016, "Infrequent Rebalancing, Return Autocorrelation, and Seasonality," Journal of Finance, forthcoming.

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, 2017, "Where Experience Matters: Asset Allocation and Asset Pricing with Opaque and Illiquid Assets," Working paper, INSEAD.

Campbell, J. Y., 2003, "Consumption-based asset pricing, " Handbook of the Economics of Finance, 1, Part B, 803–887.

Campbell, J. Y., and J. H. Cochrane, 1999, "By Force of Habit: A Consumption-Based Explanation of Aggregate Stock Market Behavior," Journal of Political Economy, 107, 205 -251.

Chien, Y., H. Cole and H. Lustig, 2012, "Is the Volatility of the Market Price of Risk Due to Intermittent Portfolio Rebalancing?", American Economic Review, 102, 2859–2896.

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.

Cuoco, D., 1997, "Optimal Consumption and Equilibrium Prices with Portfolio Constraints and Stochastic Income," Journal of Economic Theory, 72, 33-73.

Cvitanic J. and I. Karatzas, 1992, "Convex Duality in Constrained Portfolio Optimization," The Annals of Applied Probability, 2, 767-818.

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.

Delgado, F., Dumas, B. and G. Puopolo, 2015, "Hysteresis Bands on Returns, Holding Period and Transactions Costs," Journal of Banking and Finance, 57, 86-100.

Demsetz, H., 1968, "The Cost of Transacting," Quarterly Journal of Economics, 82, 33-53.

Duffie, D., 2010, "Presidential Address: Asset Price Dynamics with Slow-Moving Capital," The Journal of Finance, 55, 1237-167.

Duffie, D., N. Gârleanu, and L. H. Pedersen, 2002, "Securties Lending, Shorting, and Pricing," Journal of Financial Economics, 66, 307–339.

Duffie, D., N. Gârleanu, and L. H. Pedersen, 2007, "Valuation in Over-the-counter Markets," Review of Financial Studies, 20, 1865–1900.

Duffie, D., G. Manso, and S. Malamud, 2014, "Information Percolation with Segmented Markets," Journal of Economic Theory, 153, 1–32.

Duffie, D., and B. Strulovici, 2012, "Capital Mobility and Asset Pricing," Econometrica, 80, 2469–2509.

Dumas, B. and E. Luciano, 1991, "An Exact Solution to a Dynamic Portfolio Choice Problem under Transactions Costs," The Journal of Finance, 46, 577-595.

Dumas, B. and A. Lyasoff, 2012, "Incomplete-Market Equilibria Solved Recursively on an Event Tree," The Journal of Finance, 67, 1897-1940.

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.

Fama, E. F., L. Fisher, M. C. Jensen and R. Roll, 1969, "The Adjustment of Stock Prices to New Information," International Economic Review, 10, 1-21.

Foucault, T., 1999, "Order Flow Composition and Trading Cost in a Dynamic Limit Order Market,", Journal of Financial Markets, 2, 99-134.

Foucault, T., O. Kadan and E. Kandel, 2005, "Limit Order Book as a Market for Liquidity," Review of Financial Studies, 18, 1171-1217.

French, K., 2008, "The Cost of Active Investing," Journal of Finance, 63, 1537–1573.

Gârleanu, N., 2009, "Portfolio Choice and Pricing in Illiquid Markets," Journal of Economic Theory, 144, 532–564

Gârleanu, N. and L. Pedersen, 2013, "Dynamic Trading with Predictable Returns and Transactions Costs," The Journal of Finance, 68, 2309–2340.

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.

Goettler, R., C. Parlour, and U. Rajan, 2005, "Equilibrium in a Dynamic Limit Order Market,", Journal of Finance, 60, 2149–2192.

Gourinchas, P.-O., and J. Parker, 2002, "Consumption over the Life-Cycle," Econometrica, 70, 47–89.

Hau, H., 2006, "The Role of Transaction Costs for Financial Volatility: Evidence from the Paris Bourse," Journal of the European Economic Association, 4, 862-890.

He, H. and D. M. Modest, 1995, "Market Frictions and Consumption-Based Asset Pricing," Journal of Political Economy, 103, 94-117.

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.

Hendershott, P., S. X. Li, A. Menkveld and M. Seasholes, 2014, "Asset Price Dynamics with Limited Attention," Working paper, VU University of Amsterdam.

Holmström, B., J. Tirole, 2001, "LAPM: A Liquidity-Based Asset Pricing Model," Journal of Finance, 56, 1837-1867.

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.

Jang, B.-G., H. K. Koo, H. Liu and M. Lowenstein, 2007, "Liquidity Premia and Transactions Costs," The 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.

Leland, H. E., 2000, "Optimal Portfolio Implementation with Transactions Costs and Capital Gains Taxes," Working paper, University of California, Berkeley.

Liu, H. and M. Loewenstein, 2002, "Optimal Portfolio Selection with Transactions Costs and Finite Horizons," Review of Financial Studies, 15, 805-835.

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.

Luttmer, E. G. J., 1996, "Asset Pricing in Economies with Frictions," Econometrica, 64, 1439-1467.

Magill, M. and G. M. Constantinides, 1976, "Portfolio Selection with Transactions Costs," Journal of Economic Theory, 13, 245-263.

Milne, F., E. Neave, 2003, "A General Equilibrium Financial Asset Economy with Transaction Costs and Trading Constraints," Working paper, Queen's University.

Novy-Marx, R. and M. Velikov, 2015, "A Taxonomy of Anomalies and their Trading Costs," Review of Financial Studies, 29, 104-147.

Obizhaeva, A. and J. Wang, 2013, "Optimal Trading Strategy and Supply/demand Dynamics," Journal of Financial Markets, 16, 1–32.

Parlour, C. 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.

Philippon, T., 2015. "Has the US Finance Industry Become Less Efficient? On the Theory and Measurement of Financial Intermediation," American Economic Review, 105, 1408-38.

Rachedi, O., 2014, "Asset Pricing with Heterogeneous Inattention," Working paper, Universidad Carlos III de Madrid.

Roşu, I., 2009, "A Dynamic Model of the Limit Order Book," Review of Financial Studies, 22, 4601-4641.

Shreve, S. E. and H. M. Soner, 1994, "Optimal Investment and Consumption with Transactions Costs," Annals of Applied Probability, 4, 609-692.

Stoll, H. R., 2000, "Friction," The Journal of Finance, 55, 1479-1514.

Vayanos, D., 1998, "Transaction Costs and Asset Prices: A Dynamic Equilibrium Model," 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.