Equilibrium and welfare in markets with financially constrained arbitrageurs

Denis Gromb and Dimitri Vayanos

This is an electronic version of an Article published in the Journal of financial economics 66 (2-3) pp. 361-407 © 2002 Elsevier Science B.V.
http://www.elsevier.com/locate/jfec

Equilibrium and welfare in markets with financially constrained arbitrageurs *

Denis Gromb
LBS and CEPR

Dimitri Vayanos
MIT and NBER

August 19, 2002

Abstract

We propose a multiperiod model in which competitive arbitrageurs exploit discrepancies between the prices of two identical risky assets traded in segmented markets. Arbitrageurs need to collateralize separately their positions in each asset, and this implies a financial constraint limiting positions as a function of wealth. In our model, arbitrage activity benefits all investors because arbitrageurs supply liquidity to the market. However, arbitrageurs might fail to take a socially optimal level of risk, in the sense that a change in their positions can make all investors better off. We characterize conditions under which arbitrageurs take too much or too little risk.

*We thank Greg Bauer, Domenico Cuoco, Peter DeMarzo, Martin Gonzalez-Eiras, Nobu Kiyotaki, Leonid Kogan, Pete Kyle, Anna Pavlova, Raghu Rajan, Roberto Rigobon, Jean-Charles Rochet, Steve Ross, David Scharfstein, Raman Uppal, Jean-Luc Vila, Jiang Wang, Greg Willard, Wei Xiong, Luigi Zingales, Jeff Zwiebel, the anonymous referee, seminar participants at Amsterdam, Boston University, Carnegie-Mellon, Columbia, Duke, Lausanne, LSE, MIT, Minnesota, Montreal, New York Fed, Northwestern, NYU, Princeton, Rice, Rochester, Santa Fe, St. Louis, Stanford, Tilburg, Toulouse, UNC, UT Austin, Utah, Warwick, Wharton, and Yale, and participants at the AFA and SAET conferences for valuable comments. We also thank Sergey Iskoz for excellent research assistance. Please address correspondence to Dimitri Vayanos, MIT Sloan School of Management, 50 Memorial Drive E52-437, Cambridge MA 02142-1347, tel 617-2532956, fax 617-2586855, e-mail `dimitriv@mit.edu`.

1. Introduction

This paper proposes a financial market model in which some investors (arbitrageurs) have better investment opportunities than others, but face financial constraints. We study the constraints' implications for arbitrageur behavior, asset prices, and welfare. In our model, arbitrage activity benefits all investors. This is because through their trading, arbitrageurs bring prices closer to fundamentals and supply liquidity to the market. Competitive arbitrageurs could, however, fail to take a socially optimal level of risk. In some cases, for example, a reduction in their positions can make all investors better off. Our analysis provides insights into possible sources of allocative inefficiency in financial markets.

While the importance of financially constrained arbitrage had been emphasized before, it was put into particularly sharp focus during the 1998 financial crisis. Prior to that crisis, many hedge funds were following arbitrage strategies, betting that prices of comparable securities would eventually converge. During the crisis, as prices instead diverged, hedge funds incurred heavy losses and had to liquidate many of their positions. That these positions were generally viewed as profitable in the long run suggests that a short-run decline in net worth severely constrained the hedge funds' investment capacity. Interestingly, different "legs" of an arbitrage position were often liquidated separately, exposing their buyers to greater risk than holding the whole position.[1] This suggests that hedge funds were uniquely able to manage complex arbitrage positions, and that other investors could not easily "jump in" and replace them.

We consider a multiperiod competitive economy with a riskless asset and two risky assets with identical payoffs. The markets for the risky assets are segmented in that some investors can only invest in one asset and some only in the other. Investors' demand for an asset is affected by endowment shocks that covary with the asset payoff. Since the covariances differ for the two types of investors, the assets' prices can differ. A third type of investor (arbitrageurs) can invest in both assets and exploit price discrepancies. Arbitrageurs act as intermediaries: by exploiting price discrepancies, they facilitate trade among the other

[1] For example, the Wall Street Journal (September 28, 1998), writing about Long-Term Capital Management, a major hedge fund and one of the worst hit during the crisis, reports, "And while Long-Term Capital ran its derivatives portfolio to offset risks and hedges from other balance-sheet investments, a bankruptcy or **liquidation** also could have thrown the entire portfolio onto the market without the dealers necessarily hedged. Such moves would have given dealers new risks as they attempted to cope with the flood of financial instruments being forced into their hands. Long-Term Capital, on its own, may have 'aggregated' the risks of both sides of a given trade to neutralize the market impact. In a bankruptcy or **liquidation**, however, these instruments would have become unbundled and spread across dealers who didn't necessarily have these positions hedged, leaving them vulnerable to market risks." (Emphasis in original.)

investors, in effect providing liquidity to them.[2]

We model the financial constraints as follows. First, arbitrageurs have one margin account for each risky asset, consisting of positions in that asset and in the riskless asset. Second, the overall position in the account has to be such that the account's value remains positive until the next period. Requiring each account to be collateralized separately (i.e., ruling out cross-margining) implies that arbitrageurs' wealth constrains the positions they can take. Intuitively, arbitrageurs must have enough wealth to cover variations in the values of the two accounts, even though these variations cancel out eventually. The no cross-margining assumption captures the notion that the custodians of the arbitrageurs' margin accounts in one market might not accept a position in the other as collateral.[3]

We show that if arbitrageurs' wealth is insufficient, they may be unable to eliminate price discrepancies between the risky assets. The resulting price wedge increases with the relative demand of the two types of investors, and decreases with the arbitrageurs' wealth. Arbitrageurs exploit the price wedge by holding opposite positions in the two assets. Interestingly, if the capital gain on the arbitrage opportunity until the next period is risky (because the relative demand of the two types of investors can vary), arbitrageurs could choose not to invest up to the financial constraint. This is for risk management reasons: arbitrageurs realize capital losses when the price wedge widens, which deprives them of funds when they have the best use for them. Although arbitrageurs engage in risk management, they can exacerbate price volatility: when prices diverge, they may have to liquidate some of their positions in each market, further widening the price wedge. These results are consistent with earlier papers, particularly Shleifer and Vishny (1997), and capture some features of the 1998 crisis.

We next move to the welfare analysis, which we view as this paper's main contribution. Understanding the welfare implications of investors' financial constraints is important, as they underlie many policy debates. For example, during the 1998 crisis, it was feared that the positions of Long-Term Capital Management (LTCM), a major hedge fund and one of the worst hit during the crisis, were so large that their forced liquidation would

[2]While market segmentation is exogenous in our model, it could result from frictions such as asymmetric information or institutional constraints. For example, prior to the 1998 crisis, British government bonds were significantly more expensive than comparable German bonds. Some hedge funds attempted to exploit this price discrepancy, which they viewed as arising from the fact that many British institutional investors were constrained to hold British securities.

[3]This assumption is thus related to that of market segmentation. Indeed, the same friction that prevents investors in one asset from investing in the other can also prevent the custodians of the arbitrageurs' accounts in one market from accepting a position in the other as collateral. Returning to the government bond example, many British bond dealers did not deal in German securities. Therefore, a hedge fund shorting British bonds through these dealers could not post German bonds as collateral.

depress prices. This could disrupt markets and possibly jeopardize the financial system, with consequences reaching far beyond LTCM's investors. Such concerns were behind the Federal Reserve's controversial decision to orchestrate LTCM's rescue.[4] Our model provides a framework for conducting a welfare analysis of markets with financial constraints, which to our knowledge has not been done before.

In our model, arbitrage activity is beneficial to all investors, because arbitrageurs supply liquidity to the market. We show, however, that arbitrageurs can fail to take a socially optimal level of risk. When, in equilibrium, arbitrageurs are heavily invested in the arbitrage opportunity, a reduction in their positions can make all investors better off. Conversely, when, in equilibrium, arbitrageurs do not invest much in the arbitrage opportunity, an increase in their positions can be Pareto improving.

How can a change in the arbitrageurs' positions be Pareto improving? The intuition is that (i) competitive arbitrageurs fail to internalize that changing their positions affects prices, and (ii) due to market segmentation and financial constraints, agents' marginal rates of substitution differ, and so a redistribution of wealth induced by a change in prices can be Pareto improving. This mechanism was first pointed out by Geanakoplos and Polemarchakis (1986) in a general incomplete markets setting. Our contribution is to explore this mechanism when incompleteness is created by market segmentation and financial constraints.[5]

To illustrate how the mechanism operates in our setting, suppose that after arbitrageurs have chosen their positions, the other investors' relative demand increases. These investors are then eager for liquidity, which arbitrageurs are eager to provide because the price wedge is wide. However, the arbitrageurs are hampered by the capital losses on their positions. Reducing the positions would limit the losses, and enable arbitrageurs to provide more liquidity. Of course, arbitrageurs internalize that liquidity provision is profitable, which is why they might choose positions below the financial constraint. However, what competitive arbitrageurs fail to internalize is that reducing their positions affects prices, i.e., that with

[4] According to Alan Greenspan's testimony before Congress: "[T]he act of unwinding LTCM's portfolio in a forced liquidation would not only have a significant distorting effect on market prices but also in the process would produce large losses, or worse, for a number of creditors and counterparties, and for other market participants who were not directly involved with LTCM....Had the failure of LTCM triggered the seizing up of markets, substantial damage could have been inflicted on many market participants...and could have potentially impaired the economies of many nations, including our own." (Quoted from Edwards, 1999.)

[5] We should note that in our model the extent of market incompleteness is endogenous, and depends on the arbitrageurs' wealth. Indeed, if wealth is large, then arbitrageurs can close the arbitrage opportunity, and markets are effectively complete. In this sense, arbitrageurs can be interpreted as "financial innovators," and our model is related to the financial innovation literature. For surveys of this literature, see Allen and Gale (1994) and Duffie and Rahi (1995).

smaller losses, they can invest more aggressively, thus attenuating the widening of the price wedge. This would further reduce their losses, allowing them to further exploit the wide price wedge. Hence, they can be better off. For the other investors, the benefit of increased liquidity when they need it most can dominate the cost of the initial reduction in liquidity. Hence, they too can be better off.

Critical to this Pareto improvement is that the change in prices induces a redistribution of wealth, and that agents' marginal rates of substitution differ. The wealth redistribution benefits the arbitrageurs in the early periods and in states where the price wedge widens, and benefits the other investors in later periods (through the arbitrageurs' increased ability to provide liquidity). This can be Pareto improving because arbitrageurs have a stronger preference, relative to the other investors, for receiving funds in the early periods (since they have a greater return on wealth) and in states where the price wedge widens (since only arbitrageurs can exploit the price wedge). Note that the differences in marginal rates of substitution arise naturally from market segmentation and financial constraints.

The sources of allocative inefficiency in our model seem quite realistic. That a liquidation of arbitrageurs' positions can reduce other arbitrageurs' net worth through price effects was an important feature of the 1998 crisis. Indeed, during the crisis, there were concerns that hedge funds were imposing negative externalities on each other, precisely through price effects. A liquidation of arbitrageurs' positions can also be detrimental to other investors through a reduction in market liquidity. Indeed, the Federal Reserve's concerns about market disruption can partly be interpreted as concerns about market liquidity. (In fact, liquidity dried up in many markets during the crisis.)

This paper is related to several strands of the literature, in addition to that on general equilibrium with incomplete markets. The notion that some investors have better investment opportunities than others underlies all models of financial markets with asymmetric information. In most of these models, however, financial constraints are very limited. For example, it is generally assumed that investors can freely borrow at the riskless rate. We should note that asymmetric information models assume some implicit financial constraints. For example, informed investors are generally prevented from issuing "equity," i.e., forming a mutual fund, although Admati and Pfleiderer (1990) represent an exception.

Some papers focus on financial constraints more explicitly. In Tuckman and Vila (1992, 1993), financial constraints arise from holding costs, and they prevent arbitrageurs from eliminating mispricings.[6] In Dow and Gorton (1994), financial constraints take the form of a short horizon and a trading cost, and again mispricings can arise. Yuan (1999, 2001) considers a model where arbitrageurs face a borrowing constraint.

Shleifer and Vishny (1997) are the first to emphasize the intertemporal wealth effects of financial constraints: the arbitrageurs' ability to invest is constrained by their wealth, which itself depends on the past performance of the arbitrageurs' investments. In their model, arbitrageurs rely on external funds and face the constraint that the inflow of funds is sensitive to performance.[7] They show that arbitrageurs can choose not to invest up to the financial constraint and, through their trading, can amplify the effects of noise shocks on prices. Intertemporal wealth effects are also central to our analysis. The main difference with Shleifer and Vishny is that we model more explicitly the arbitrageurs' advantage over the other investors (through market segmentation) and the mechanics of the financial constraint (through the margin accounts). This allows us to conduct a welfare analysis.

Intertemporal wealth effects are studied in several other recent papers. Xiong (2001) considers a model in which arbitrageurs and "long-term" investors invest in a single risky asset. He shows that arbitrageurs can amplify the effects of noise shocks on prices. Aiyagari and Gertler (1999) and Sodini (2001) obtain similar amplification effects in models with margin constraints. See also Attari and Mello (2001) for a model with a monopolistic arbitrageur. Kyle and Xiong (2001) assume two risky assets, and show that arbitrageurs can induce financial contagion in the form of an increased correlation between asset prices. These results are driven not by financial constraints but by the fact that arbitrageurs have logarithmic utility, and thus their demand for risky assets is increasing in wealth.

In Loewenstein and Willard (2001), arbitrageurs with long horizons provide liquidity to overlapping generations of short-horizon investors, by holding the assets with short-term price risk. Arbitrageurs face the constraint that wealth must be nonnegative at any time. In Basak and Croitoru (2000), agents hold heterogeneous beliefs, and trade a risky asset and a financial derivative, under portfolio constraints. In both papers, arbitrage opportunities can exist.

A recent strand of the literature studies the optimal policy of an investor facing exoge-

[6] In Tuckman and Vila, the arbitrageurs' advantage over the other investors derives not from better information, but from the arbitrageurs' ability to participate in a larger set of markets. For other models of limited market participation see, for example, Basak and Cuoco (1998) and Zigrand (2001).

[7] Allen and Gale (1999) and Holmström and Tirole (2001) also consider equilibrium models where investors rely on external funds and are facing financial constraints. For models exploring the macroeconomic implications of financial constraints see, for example, Kiyotaki and Moore (1998), Krishnamurthy (2000), and Mian (2002).

nous portfolio constraints.[8] In Liu and Longstaff (2001), an arbitrageur can invest in a stochastic "spread," known to converge at some fixed time, under the constraint that his position cannot exceed a given function of wealth. Because of this constraint the arbitrage strategy becomes risky, and the arbitrageur could choose not to invest up to the constraint.

This paper proceeds as follows. In Section 2 we present the model. In Section 3 we derive a competitive equilibrium, and in Section 4 we study its welfare properties. In Section 5 we compute the equilibrium and the welfare effects in closed form, in the "continuous-time" case, where the time between consecutive periods goes to zero. In Section 6 we conclude, and discuss some possible extensions and policy implications. All proofs are in the Appendix.

2. The model

There are $T+1$ periods, $t = 0, 1, .., T$, with $T \geq 2$. The universe of assets consists of a riskless asset and two risky assets, A and B. All agents can invest in the riskless asset. However, only some agents (arbitrageurs) can invest in both risky assets, while the other agents (A- and B-investors) can invest in only one risky asset.

2.1. Assets

The riskless asset has an exogenous return equal to one. Assets A and B are in zero net supply, and pay off only in period T. Their payoffs are identical and equal to $\sum_{t=0}^{T} \delta_t$, where δ_t is a random variable revealed in period t. We assume that the δ_t are independent and identically distributed, and that the distribution is symmetric around zero on the bounded support $[-\overline{\delta}, \overline{\delta}]$. We denote by $p_{i,t}$ the price of asset $i = A, B$ in period t, and set

$$\phi_{i,t} = E_t\left(\sum_{s=0}^{T} \delta_s\right) - p_{i,t} = \sum_{s=0}^{t} \delta_s - p_{i,t}.$$

The variable $\phi_{i,t}$ represents the expected excess return per share of asset i and, for simplicity, we refer to it as asset i's risk premium.

The assumptions of exogenous riskless return, zero net supply assets, and identical asset payoffs are for simplicity. In particular, the zero net supply assumption ensures that arbitrageurs hold opposite positions in the two risky assets, and do not bear any aggregate risk. The bounded support assumption plays a role for the financial constraint (see below).

[8]See, for example, Cvitanic and Karatzas (1992), Grossman and Vila (1992), and Cuoco (1997). See also Kogan and Uppal (2001), who perform both a partial and a general equilibrium analysis.

2.2. A- and B-investors

The markets for assets A and B are segmented in that some agents, A-investors, can only invest in asset A and the riskless asset, while others, B-investors, can only invest in asset B and the riskless asset.

We take market segmentation as given. We simply assume that A-investors face large transaction costs for investing in asset B, and so do B-investors for asset A. These costs can be due to "physical" factors, such as distance. Alternatively, they may be a reduced form for information asymmetries or institutional constraints. Market segmentation is a realistic assumption in many contexts. In an international context, for example, it is well known that there is "home bias," i.e., investors mainly hold domestic rather than foreign assets. One simple story in the spirit of information asymmetry could run as follows. Assets A and B are "certificates," written in different languages, A and B. A-investors understand language A but not language B. Hence, A-investors will not hold an asset written in language B, for fear of holding a worthless piece of paper. The reverse holds for B-investors.

The i-investors, $i = A, B$, are competitive, form a continuum with measure 1, and have initial wealth $w_{i,0}$. They maximize the expected utility of period T wealth, $w_{i,T}$. We assume that utility is exponential, i.e.,

$$U_i(w_{i,T}) = -\exp\left(-\alpha w_{i,T}\right).$$

In each period $t \geq 1$, investors receive an endowment that is correlated with the information δ_t on the asset payoffs. We assume that the endowment of i-investors in period t is $u_{i,t-1}\delta_t$. The coefficient $u_{i,t-1}$ measures the extent to which the endowment covaries with δ_t. If $u_{i,t-1}$ is high, the covariance is high, and thus the willingness of i-investors to hold asset i in period $t-1$ is low. We refer to $u_{i,t-1}$ as the "supply shock" of i-investors in period $t-1$, to emphasize that it negatively affects investor demand in that period. To be consistent with the zero net supply assumption, the endowments can be interpreted as positions in a different but correlated asset. Our specification of endowments is quite standard in the market microstructure literature (see O'Hara, 1995).

We assume that the supply shocks are opposites for the A- and B-investors, i.e.,

$$u_{A,t} = -u_{B,t} = u_t, \quad \text{for} \quad t = 0, .., T-1.$$

We assume opposite shocks for simplicity. What is critical for our model is that A- and B-investors incur different shocks. Different shocks, together with market segmentation, create a role for the arbitrageurs. Indeed, arbitrageurs exploit price discrepancies between assets A and B, which can arise because A- and B-investors have different propensities to hold the assets (due to the different supply shocks) but cannot trade with each other (due to market segmentation). Note that arbitrageurs act as intermediaries. Suppose, for example, that A-investors receive a positive supply shock, in which case B-investors receive a negative shock. Then arbitrageurs buy asset A from the A-investors, who are willing to sell, and sell asset B to the B-investors, who are willing to buy. Through this transaction arbitrageurs make a profit, while at the same time providing liquidity to the other investors.

We consider two cases for the supply shock u_t. The first is the *certainty case* where u_t is deterministic and, for simplicity, identical in all periods, i.e., $u_t = u_0$ for $t = 0, .., T-1$. The certainty case is a useful benchmark, and illustrates the mechanics of the model. Second, we consider the *uncertainty case* where u_t is stochastic. For simplicity, we assume that all uncertainty is resolved in period 1, and u_t is identical in all subsequent periods, i.e., $u_t = u_1$ for $t = 1, .., T-1$.[9] In both cases, we assume that $u_0 > 0$. Furthermore, in the uncertainty case we assume that u_1 has positive and bounded support $[\underline{u}_1, \overline{u}_1]$ and is independent of δ_1.

The i-investors choose holdings of asset i in period t, $y_{i,t}$, to maximize expected utility of period T wealth. Their optimization problem, $\mathcal{P}_i$, is

$$\max_{\substack{y_{i,t} \\ t=0,..,T-1}} \quad -E_0 \exp\left(-\alpha w_{i,T}\right),$$

subject to the dynamic budget constraint

$$w_{i,t+1} = w_{i,t} + y_{i,t}(p_{i,t+1} - p_{i,t}) + u_{i,t}\delta_{t+1} \qquad \text{for} \quad t = 0, .., T-1. \tag{1}$$

Eq. (1) states that period $t+1$ wealth equals period t wealth plus the capital gains and endowment received between periods t and $t+1$.

2.3. Arbitrageurs

Arbitrageurs can invest in both assets A and B. They are competitive, form a continuum with measure μ, and have initial wealth w_0.[10] They maximize expected utility of period T wealth, w_T. We denote the arbitrageurs' utility by $U(w_T)$.

[9] A more natural assumption would be that uncertainty is resolved gradually over periods $1, .., T-1$. Gradual resolution of uncertainty would, however, complicate the analysis, while assuming that $T = 2$ would eliminate some interesting economic effects (as explained in Section 5).

[10] We should note that by fixing the measure of the arbitrageurs, we do not allow for entry into the arbitrage industry, which seems a realistic assumption for understanding short-run market behavior. For example, during the 1998 crisis, when prices of securities involved in arbitrage strategies diverged, there was little inflow of new capital to correct the divergence.

Arbitrageurs are subject not only to the budget constraint, as are the A- and B-investors, but also to a financial constraint, that we model as follows. First, arbitrageurs have one margin account for each risky asset, consisting of positions in the asset and in the riskless asset. Second, the overall position in the account has to be such that the account's value remains positive until the next period. Denoting by $x_{i,t}$ the position in asset $i = A, B$ in period t, and by $V_{i,t}$ the value of the margin account, we have

$$V_{i,t+1} = V_{i,t} + x_{i,t}(p_{i,t+1} - p_{i,t}).$$

Requiring that $V_{i,t+1} \geq 0$ implies that

$$V_{i,t} \geq \max_{p_{i,t+1}} \{x_{i,t}\,(p_{i,t} - p_{i,t+1})\}.$$

This in turn implies the financial constraint

$$w_t = \sum_{i=A,B} V_{i,t} \geq \sum_{i=A,B} \max_{p_{i,t+1}} \{x_{i,t}\,(p_{i,t} - p_{i,t+1})\},$$

where w_t denotes the arbitrageurs' wealth in period t. The financial constraint requires arbitrageurs to have enough wealth to cover the maximum loss that each margin account can incur. This implies that arbitrageurs' wealth constrains the positions they can take. In particular, arbitrageurs may be unable to eliminate a price discrepancy in a given period, even if it is known that the discrepancy will disappear in the next period. Of course, arbitrageurs would always be able to eliminate such a discrepancy if they were subject only to the standard constraint that wealth be nonnegative in each period.

Requiring each margin account to be collateralized separately (i.e., ruling out cross-margining) means that arbitrageurs cannot use a position in one asset as collateral for a position in the other. Suppose, for example, that arbitrageurs short asset B. Then they must deposit as collateral in their B-account both the cash proceeds from selling asset B and some additional cash, to cover the cost of buying asset B next period. They cannot, however, deposit asset A.

The no cross-margining assumption is, in fact, related to that of market segmentation. Indeed, the same frictions (e.g., institutional constraints, etc.) that prevent B-investors from investing in asset A can also prevent the custodians of arbitrageurs' B-accounts from accepting asset A as collateral.[11] Returning to our language story, the custodians of the

[11] The custodians can be the financial exchanges if the assets are futures contracts, or the brokers/dealers through whom the arbitrageurs are trading if the assets are stocks or bonds. The no cross-margining assumption is quite realistic in both cases. For example, futures exchanges generally accept as collateral only positions in contracts traded within the exchange, and dealers generally accept only positions in assets they are dealing in. In practice, arbitrageurs sometimes avoid cross-margining even when it is allowed, for fear of revealing all their information to a single counterparty and then being front-run (see Ko, 2000).

B-accounts do accept asset A as collateral because they do not understand language A.

Requiring each margin account to be fully collateralized ensures that arbitrageurs never default. Ruling out default allows us to avoid modeling the custodians of the arbitrageurs' margin accounts, and having to consider their welfare. (For example, A- and B-investors can serve as custodians.) Note that it is because of the full collateralization assumption that we need to consider probability distributions with bounded support.[12] [13]

The arbitrageurs' optimization problem, $\mathcal{P}$, is

$$\max_{\substack{x_{A,t}, x_{B,t} \\ t=0,..,T-1}} E_0 U\left(w_T\right),$$

subject to the dynamic budget constraint

$$w_{t+1} = w_t + \sum_{i=A,B} x_{i,t}(p_{i,t+1} - p_{i,t}) \quad \text{for} \quad t = 0, .., T-1,$$

and the financial constraint

$$w_t \geq \sum_{i=A,B} \max_{p_{i,t+1}} \{x_{i,t}\left(p_{i,t} - p_{i,t+1}\right)\} \quad \text{for} \quad t = 0, .., T-1.$$

2.4. Equilibrium

We define competitive equilibrium as follows.

Definition 1 *A competitive equilibrium consists of prices* $\{p_{i,t}\}_{\substack{i=A,B \\ t=0,..,T}}$*, asset holdings of the i-investors* $\{y_{i,t}\}_{t=0,..,T-1}$*, for* $i = A, B$*, and of the arbitrageurs* $\{x_{i,t}\}_{\substack{i=A,B \\ t=0,..,T-1}}$*, such that*

- *given the prices,* $\{y_{i,t}\}_{t=0,..,T-1}$ *solve problem* $\mathcal{P}_i$*, for* $i = A, B$*, and* $\{x_{i,t}\}_{\substack{i=A,B \\ t=0,..,T-1}}$ *solve problem* $\mathcal{P}$*, and*

- *for* $i = A, B$*,* $t = 0, .., T-1$*, markets clear:*

$$y_{i,t} + \mu x_{i,t} = 0.$$

[12] In one sense, our financial constraint is endogenous in that it depends on the properties of the price process. The notion that margin requirements are endogenously chosen to prevent default has appeared in recent general equilibrium literature (see, e.g., Geanakoplos, 2001 and the references therein).

[13] The financial constraint is imposed only on the arbitrageurs and not on the A- and B-investors. The constraint will not be binding for these investors if their initial wealth is large enough. Indeed, since utility is exponential, optimal holdings of the risky asset are independent of wealth, and so are capital gains. Moreover, since asset payoffs and supply shocks have bounded support, so do capital gains. Therefore, for large enough initial wealth, capital losses are always smaller than wealth, and the financial constraint is not binding. Note that the initial wealth of the A- and B-investors does not have to be larger than that of the arbitrageurs. Indeed, if the measure μ of the arbitrageurs is small enough, the arbitrageurs' positions are much larger than those of the A- and B-investors, and thus require more collateral.

3. Equilibrium

In this section, we derive a competitive equilibrium. We look for an equilibrium that satisfies two properties. First, the risk premia of assets A and B are opposites, i.e., $\phi_{B,t} = -\phi_{A,t}$, because the assets are in zero net supply and the supply shocks of the A- and B-investors are opposites. Second, the arbitrageurs' positions in assets A and B are also opposites, i.e., $x_{B,t} = -x_{A,t}$, because the risk premia of the two assets are opposites. (Note that since the arbitrageurs' positions are opposites, the same is true for the positions of the A- and B-investors.) We refer to an equilibrium satisfying these properties as symmetric.

A symmetric equilibrium is characterized by the risk premium of asset A, $\phi_{A,t}$, and the arbitrageurs' position in that asset, $x_{A,t}$, $t = 0, .., T-1$ (for $t = T$, $\phi_{A,T} = 0$). In what follows we drop the subscript A from $\phi_{A,t}$, $x_{A,t}$, and $y_{A,t}$. Note that the risk premium ϕ_t is one-half of the price wedge between assets A and B, since

$$p_{B,t} - p_{A,t} = \left(\sum_{s=0}^{t} \delta_s + \phi_t\right) - \left(\sum_{s=0}^{t} \delta_s - \phi_t\right) = 2\phi_t.$$

In a symmetric equilibrium, ϕ_t and x_t do not depend on δ_t, the asset payoff information. This is because the arbitrageurs' positions in assets A and B are opposites, and thus the arbitrageurs' wealth does not depend on δ_t. Therefore, ϕ_t and x_t can be stochastic only because of the supply shock u_t. In the certainty case, where u_t is deterministic, ϕ_t and x_t are thus deterministic. Likewise, in the uncertainty case, where u_t is deterministic from period 1 on, so are ϕ_t and x_t. We study the certainty case in Section 3.1, and the uncertainty case in Section 3.2.

3.1. *The certainty case*

We first study the optimization problem $\mathcal{P}_A$ of the A-investors. Since $u_{A,t} = u_t = u_0$, we can write the dynamic budget constraint (1) as

$$\begin{aligned} w_{A,t+1} &= w_{A,t} + y_t(p_{A,t+1} - p_{A,t}) + u_0\delta_{t+1} \\ &= w_{A,t} + y_t\left[\left(\sum_{s=0}^{t+1} \delta_s - \phi_{t+1}\right) - \left(\sum_{s=0}^{t} \delta_s - \phi_t\right)\right] + u_0\delta_{t+1} \\ &= w_{A,t} + y_t(\phi_t - \phi_{t+1}) + (y_t + u_0)\delta_{t+1}. \end{aligned}$$

The term $y_t(\phi_t - \phi_{t+1})$ is the expected capital gain of the A-investors between periods t and $t+1$. It is proportional to the difference between the risk premia in these periods. The term $(y_t + u_0)\delta_{t+1}$ represents the risk borne by the A-investors between periods t and $t+1$. It is the sum of the unexpected capital gain and the period $t+1$ endowment.

Expected utility is

$$-E\exp\left(-\alpha w_{A,T}\right) = -E\exp\left[-\alpha\left(w_{A,0} + \sum_{t=0}^{T-1}\left(y_t(\phi_t - \phi_{t+1}) + (y_t + u_0)\delta_{t+1}\right)\right)\right].$$

To compute expected utility, we need to compute

$$E\exp(-\alpha(y_t + u_0)\delta_{t+1}).$$

This expectation depends on the probability distribution of δ_{t+1}. We do not assume a specific distribution, but rather define the function f by

$$E\exp(-\alpha y\delta) \equiv \exp(\alpha f(y)).$$

Some useful properties of f are summarized in the following Lemma.

Lemma 1 *The function f is positive, strictly convex, and satisfies $f(y) = f(-y)$ and $\lim_{y\to\infty} f'(y) = \overline{\delta}$.*

Using f, we can write expected utility as

$$-\exp\left[-\alpha\left(w_{A,0} + \sum_{t=0}^{T-1}\left(y_t(\phi_t - \phi_{t+1}) - f(y_t + u_0)\right)\right)\right],$$

and the optimization problem of the A-investors as

$$\max_{\substack{y_t \\ t=0,..,T-1}} \sum_{t=0}^{T-1}\left(y_t(\phi_t - \phi_{t+1}) - f(y_t + u_0)\right).$$

The optimization problem takes a simple form. We can interpret $f(y_t + u_0)$ as a cost of bearing risk between periods t and $t+1$. This "inventory" cost depends on the position y_t in asset A and on the supply shock u_0. (The inventory cost would be quadratic if the probability distribution of δ_t were normal; a normal distribution is, however, ruled out by the bounded support requirement.) The optimization problem consists in maximizing the sum of expected capital gains, minus the sum of inventory costs. At the optimum, the expected capital gain per unit of asset A equals the marginal inventory cost, i.e.,

$$\phi_t - \phi_{t+1} = f'(y_t + u_0). \tag{2}$$

The optimization problem $\mathcal{P}_B$ of the B-investors also yields Eq. (2), since the risk premium, the supply shock, and investors' positions for assets A and B are opposites, and $f'(y) = -f'(-y)$.

We next study the optimization problem $\mathcal{P}$ of the arbitrageurs. The arbitrageurs' financial constraint is

$$
\begin{aligned}
w_t &\geq \sum_{i=A,B} \max_{p_{i,t+1}} \{x_{i,t}\,(p_{i,t} - p_{i,t+1})\} \\
&\geq \max_{\delta_{t+1}} \{x_t(-\phi_t + \phi_{t+1} - \delta_{t+1})\} + \max_{\delta_{t+1}} \{-x_t(\phi_t - \phi_{t+1} - \delta_{t+1})\} \\
&\geq 2\max_{\delta_{t+1}} \{x_t(-\phi_t + \phi_{t+1} - \delta_{t+1})\} \\
&\geq 2|x_t|\overline{\delta} - 2x_t(\phi_t - \phi_{t+1}),
\end{aligned}
$$

where the last two steps follow from the symmetry of the support of δ_{t+1} around zero. For $x_t \geq 0$ (which will be the case in equilibrium) we can write the financial constraint as

$$
x_t \leq \frac{w_t}{2\left(\overline{\delta} - (\phi_t - \phi_{t+1})\right)}.
$$

The constraint becomes more severe when the arbitrageurs' wealth w_t decreases. It also becomes more severe when the bound $\overline{\delta}$ increases, because more volatile asset payoffs increase the maximum loss of a position. Finally, it becomes less severe when $\phi_t - \phi_{t+1}$ increases, i.e., the price wedge in period $t+1$ becomes narrower relative to that in period t. This is because the maximum loss of a position decreases.

The dynamic budget constraint is

$$
\begin{aligned}
w_{t+1} &= w_t + \sum_{i=A,B} x_{i,t}(p_{i,t+1} - p_{i,t}) \\
&= w_t + x_t(\phi_t - \phi_{t+1} + \delta_{t+1}) - x_t(-\phi_t + \phi_{t+1} + \delta_{t+1}) \\
&= w_t + 2x_t(\phi_t - \phi_{t+1}). \qquad (3)
\end{aligned}
$$

The term $2x_t(\phi_t - \phi_{t+1})$ is the arbitrageurs' capital gain between periods t and $t+1$. It is independent of δ_{t+1}, and thus riskless, since the arbitrageurs' positions in assets A and B are opposites.

The arbitrageurs' optimization problem consists in maximizing the sum of capital gains, subject to the financial constraint. Since capital gains are riskless, the solution to this problem is very simple: invest up to the financial constraint if capital gains are positive, and any amount up to the constraint if capital gains are zero. Formally,

$$
x_t = \frac{w_t}{2\left(\overline{\delta} - (\phi_t - \phi_{t+1})\right)} \qquad \text{if} \qquad \phi_t - \phi_{t+1} > 0 \qquad (4)
$$

$$
x_t \leq \frac{w_t}{2\left(\overline{\delta} - (\phi_t - \phi_{t+1})\right)} \qquad \text{if} \qquad \phi_t - \phi_{t+1} = 0. \qquad (5)
$$

The equilibrium is characterized by the market-clearing condition

$$y_t + \mu x_t = 0, \tag{6}$$

and Eqs. (2), (3), (4), and (5). This system of equations turns out to have a unique solution for ϕ_t and x_t, $t = 0, .., T-1$. While the solution depends on all parameters, it will prove useful to emphasize its dependence on the arbitrageurs' initial wealth w_0 and the supply shock u_0, and thus, to denote it as

$$\{(\phi(w_0, u_0, t), x(w_0, u_0, t))\}_{t=0,\ldots,T-1}.$$

Proposition 1 *There exists a unique symmetric competitive equilibrium. In this equilibrium, ϕ_t and x_t are given by the unique solution to the system of (2)-(6), i.e.,*

$$\phi_t = \phi(w_0, u_0, t) \quad \text{and} \quad x_t = x(w_0, u_0, t) \quad \text{for} \quad t = 0, .., T-1.$$

The equilibrium can take one of two forms:

- *If $w_0 \geq \overline{w}_0 \equiv \frac{2\overline{\delta}u_0}{\mu}$, the financial constraint never binds. The arbitrageurs fully absorb the supply shock, and close the price wedge in all periods, i.e., $\mu x_t = u_0$ and $\phi_t = 0$, for $t = 0, .., T-1$.*

- *If $w_0 < \overline{w}_0$, the financial constraint binds in all periods. The arbitrageurs do not fully absorb the supply shock, i.e., $\mu x_t < u_0$, for $t = 0, .., T-1$. The price wedge narrows over time and is closed only in period T, i.e., $\phi_t - \phi_{t+1} > 0$, for $t = 0, .., T-1$, and $\phi_T = 0$. The arbitrageurs' position in asset A is given by*

$$x_0 - x_0 \frac{f'(u_0 - \mu x_0)}{\overline{\delta}} = \frac{w_0}{2\overline{\delta}}, \tag{7}$$

$$x_t - x_t \frac{f'(u_0 - \mu x_t)}{\overline{\delta}} = x_{t-1}, \tag{8}$$

and it increases over time. The risk premium of asset A is given by $\phi_T = 0$ and

$$\phi_t - \phi_{t+1} = f'(u_0 - \mu x_t). \tag{9}$$

Proposition 1 provides a simple characterization of the equilibrium. The financial constraint either never binds, if the arbitrageurs' initial wealth is large enough, or binds in all periods. In the latter case, the price wedge is not closed, and the arbitrageurs realize capital gains. Due to these gains, the arbitrageurs' wealth increases over time, and so does their position.

A variable that will prove useful is the return on an agent's period t wealth, defined as the impact on the agent's wealth in period T of an increase in wealth in period t. For the i-investors, the return is equal to 1, since at the margin these investors invest in the riskless asset whose return is 1. For the arbitrageurs, we denote the return by R_t, and distinguish two cases. When the financial constraint does not bind, $R_t = 1$, since the price wedge is closed, and thus an arbitrage position is equivalent to a position in the riskless asset. To compute R_t when the financial constraint binds, we plug Eq. (4) into (3) and get

$$w_{t+1} = w_t \frac{1}{1 - \frac{\phi_t - \phi_{t+1}}{\bar{\delta}}}.$$

Therefore,

$$w_T = w_t \prod_{s=t}^{T-1} \frac{1}{1 - \frac{\phi_s - \phi_{s+1}}{\bar{\delta}}}$$

and

$$R_t = \prod_{s=t}^{T-1} \frac{1}{1 - \frac{\phi_s - \phi_{s+1}}{\bar{\delta}}}. \tag{10}$$

When the financial constraint binds, the price wedge narrows over time, and we have $\phi_s - \phi_{s+1} > 0$. Therefore, Eq. (10) implies that $R_t > 1$, which simply means that arbitrageurs have better investment opportunities than the i-investors. Eq. (10) also implies that R_t decreases over time. This is because arbitrageurs have fewer periods over which to exploit their better opportunities.

3.2. The uncertainty case

We first note that from period 1 on, we are in the certainty case, and can use Proposition 1. We thus have

$$\phi_t = \phi(w_1, u_1, t-1) \quad \text{and} \quad x_t = x(w_1, u_1, t-1) \quad \text{for} \quad t = 1, .., T-1. \tag{11}$$

To complete the derivation of the equilibrium, we need the agents' optimality conditions in period 0. We first derive the optimality condition of the A-investors (which is identical to that of the B-investors). The A-investors' expected utility is

$$\begin{aligned} & -E\exp\left[-\alpha\left(w_{A,0} + \sum_{t=0}^{T-1} y_t(\phi_t - \phi_{t+1}) + (y_0+u_0)\delta_1 + \sum_{t=1}^{T-1}(y_t+u_1)\delta_t\right)\right] \\ = & -E\exp\left[-\alpha\left(w_{A,0} + \sum_{t=0}^{T-1} y_t(\phi_t - \phi_{t+1}) - f(y_0+u_0) - \sum_{t=1}^{T-1} f(y_t+u_1)\right)\right], \end{aligned} \tag{12}$$

where the second expectation is taken only with respect to u_1. Maximizing this second expectation with respect to y_0, we get the optimality condition

$$E\left[(\phi_0 - \phi_1 - f'(y_0 + u_0))M_A\right] = 0, \tag{13}$$

where

$$M_A = \alpha \exp\left[-\alpha\left(w_{A,0} + \sum_{t=0}^{T-1} y_t(\phi_t - \phi_{t+1}) - f(y_0 + u_0) - \sum_{t=1}^{T-1} f(y_t + u_1)\right)\right].$$

To understand the intuition for Eq. (13), compare it to the optimality condition in the certainty case,

$$\phi_0 - \phi_1 - f'(y_0 + u_0) = 0.$$

The term $\phi_0 - \phi_1 - f'(y_0 + u_0)$ is the expected capital gain per unit of asset A, net of the marginal inventory cost. The optimality condition consists in setting this "net expected capital gain" to zero. In the uncertainty case, $\phi_0 - \phi_1 - f'(y_0 + u_0)$ is the net expected capital gain, conditional on u_1. Furthermore, M_A is the A-investors' expected marginal utility of wealth, conditional on u_1. The optimality condition consists in setting the expectation, with respect to u_1, of the product of these two terms to zero.

Next, we derive the arbitrageurs' optimality condition. Their dynamic budget constraint is

$$w_1 = w_0 + 2x_0(\phi_0 - \phi_1). \tag{14}$$

Using Eq. (14) and $w_T = w_1 R_1$, we can write expected utility as

$$EU(w_T) = EU\left[(w_0 + 2x_0(\phi_0 - \phi_1))R_1\right]. \tag{15}$$

Arbitrageurs maximize expected utility with respect to x_0, subject to the financial constraint

$$w_0 \geq 2 \max_{\delta_1, u_1}\{x_0(-\phi_0 + \phi_1 - \delta_1)\}. \tag{16}$$

The derivative of the arbitrageurs' expected utility is

$$\frac{dEU}{dx_0} = E\left[2(\phi_0 - \phi_1)R_1 M\right],$$

where $M = U'(w_T)$. The derivative is equal to the expectation, with respect to u_1, of the product of two terms: the capital gain per unit of the arbitrage opportunity, $2(\phi_0 - \phi_1)$, and the marginal utility derived from wealth received in period 1, $R_1 M$.

The derivative of the arbitrageurs' expected utility depends not only on the expected capital gain, but also on the covariance between the capital gain and $R_1 M$. Although the expected capital gain is nonnegative, the covariance between the capital gain and $R_1 M$ is generally negative. To see why, notice that the covariance between the capital gain and R_1 is negative. Indeed, the arbitrageurs' period 0 position pays off when the risk premium ϕ_1 is low, i.e., the price wedge narrows, which is exactly when R_1 is low. The covariance between the capital gain and R_1 can be interpreted as a covariance between internal funds (the arbitrageurs' wealth) and profitability of investment opportunities. This suggests a parallel to theories of corporate risk management based on financial constraints. For example, Froot, Scharfstein, and Stein (1993) posit that firms should manage risk to match their internal funds with the profitability of their investment opportunities, because external finance is costly, and the ability to invest depends on internal funds. In our setting, the arbitrageurs' ability to invest depends on their wealth (the "internal funds") because of the financial constraint. A negative covariance between the capital gain and R_1 implies a negative covariance between internal funds and the profitability of investment opportunities. This makes the arbitrage opportunity less desirable as a risk management instrument, relative to the riskless asset.

The form of the arbitrageurs' optimality condition depends on whether the financial constraint is binding upwards (preventing the arbitrageurs from increasing their position), slack, or binding downwards:

$$\frac{dEU}{dx_0} > 0, \quad w_0 = 2 \max_{\delta_1, u_1} \{x_0(-\phi_0 + \phi_1 - \delta_1)\}, \quad \text{and} \quad x_0 > 0, \tag{17}$$

$$\text{or} \qquad \frac{dEU}{dx_0} = 0 \quad \text{and} \quad w_0 > 2 \max_{\delta_1, u_1} \{x_0(-\phi_0 + \phi_1 - \delta_1)\}, \tag{18}$$

$$\text{or} \qquad \frac{dEU}{dx_0} < 0, \quad w_0 = 2 \max_{\delta_1, u_1} \{x_0(-\phi_0 + \phi_1 - \delta_1)\}, \quad \text{and} \quad x_0 < 0. \tag{19}$$

It is important to note that the financial constraint can be slack or binding downwards, even when the expected capital gain on the arbitrage opportunity is positive. This result is in contrast to the certainty case, and is due to the negative covariance between the capital gain and $R_1 M$. This result is obtained, in a different setting, in Shleifer and Vishny (1997) and Liu and Longstaff (2001).

The equilibrium is characterized by Eqs. (6), (11), (13), (14), and (16)-(19). In the uncertainty case, we have not shown existence or uniqueness of the equilibrium. In our numerical solutions, however, we have always been able to find an equilibrium, and this equilibrium seems to be unique.

4. Welfare

We now turn to the welfare analysis, which we view as this paper's main contribution. Understanding the welfare implications of investors' financial constraints is important, as they underlie many policy debates. An example is the debate on systemic risk, i.e., on whether a worsening of the financial condition of some market participants can propagate into the financial system with harmful effects. One important aspect of this debate concerns the ex ante incentives for risk taking. Do market participants take an appropriate level of risk, given that their potential losses can affect others? Our model provides a framework for studying this question.

In our model, arbitrageurs choose between investing in the riskless asset and the risky arbitrage position. Before examining whether they take an appropriate level of risk, we examine whether they should take any risk at all. More precisely, we compare the welfare of each type of agent under the equilibrium allocation, and under the *no-trade allocation* in which arbitrageurs are not allowed to invest in the risky assets, i.e., $x_t = 0$ for all t. For completeness, we also consider the *no-constraint allocation* in which arbitrageurs are not subject to the financial constraint, and therefore fully absorb the supply shocks, i.e., $\mu x_t = u_t$ for all t.

Lemma 2 *Compared to the equilibrium allocation,*

- *under the no-trade allocation, A- and B-investors and arbitrageurs are worse off,*
- *under the no-constraint allocation, A- and B-investors are better off, while arbitrageurs are worse off.*

Lemma 2 shows that allowing the arbitrageurs to invest in the risky assets is Pareto improving. The intuition is that under the equilibrium allocation, arbitrageurs make a profit by exploiting price discrepancies between assets A and B. At the same time, they provide liquidity to the A- and B-investors by absorbing, to some extent, these investors' supply shocks. Under the no-trade allocation, arbitrageurs cannot invest in assets A and B. Therefore, they make no profit, and provide no liquidity to the A- and B-investors. Under the no-constraint allocation, arbitrageurs close the price wedge, and thus make no profit. At the same time, they fully absorb the supply shocks of the A- and B-investors, thus providing perfect liquidity to these investors.

Note that while the result that arbitrage activity benefits all investors is intuitive, it is by no means general. Suppose, for example, that there were some A'-investors, who could invest only in asset A, but received no endowment shocks. In the absence of arbitrageurs, these investors would profit from providing liquidity to the A-investors. Introducing arbitrageurs would provide more liquidity to the A-investors, but "steal the business" of the A'-investors (see Zigrand, 2001). Our main result is that even when arbitrage activity is Pareto improving, arbitrageurs might fail to take a socially optimal level of risk.

We next examine whether arbitrageurs take an appropriate level of risk. Since the arbitrage opportunity is risky only in period 0, we focus on the arbitrageurs' position in that period, and consider the following thought experiment. Suppose that a social planner changes the arbitrageurs' period 0 position from its equilibrium value. The social planner affects only that position, and lets the market determine all other positions and prices. In addition, the change is subject to the financial constraint, i.e., the arbitrageurs' position can only be reduced (increased) when the constraint is binding upwards (downwards). Finally, for simplicity, the change is infinitesimal. (Considering noninfinitesimal changes would only strengthen our results on the nonoptimality of the arbitrageurs' position.) If the social planner can achieve a Pareto improvement by reducing (increasing) the arbitrageurs' position, then the position is said to involve too much (too little) risk. Otherwise, the position is said to be (locally) socially optimal.

To implement the experiment formally, we treat the arbitrageurs' period 0 position, x_0, as an exogenous parameter. For each value of x_0, we define an "x_0 equilibrium" by adding to Definition 1 the requirement that the arbitrageurs' period 0 position be x_0. We compute the agents' expected utilities in this x_0 equilibrium, and evaluate their derivatives at the value of x_0 that corresponds to the original equilibrium. Whether x_0 involves too much risk, too little risk, or is socially optimal depends on the sign of these derivatives.

Proposition 2 *The derivative of the i-investors' expected utility with respect to x_0 is*

$$E\left[\left(\sum_{t=0}^{T-1}\frac{d(\phi_t-\phi_{t+1})}{dx_0}y_t\right)M_i\right], \quad \textit{for} \quad i=A,B, \tag{20}$$

and the derivative of the arbitrageurs' expected utility is

$$E\left[\left(2(\phi_0-\phi_1)R_1+2\frac{d(\phi_0-\phi_1)}{dx_0}x_0R_1+2\sum_{t=1}^{T-1}\frac{d(\phi_t-\phi_{t+1})}{dx_0}x_tR_t\right)M\right]. \tag{21}$$

The intuition for Eqs. (20) and (21) is as follows. A change in x_0 has two effects on an agent's expected utility: first through the change in the agent's period 0 position (direct effect) and second through the change in the prices the agent is facing (indirect effect). For the A- and B-investors, the direct effect is zero since these investors' period 0 positions are unconstrained optima. To determine the indirect effect, we focus on the A-investors, and consider the capital gain on asset A between periods t and $t+1$. Changing x_0 changes this capital gain by $d(\phi_t - \phi_{t+1})/dx_0$. Therefore, it changes the A-investors' period $t+1$ wealth by that amount multiplied by the investors' period t position, y_t. Notice that the change in wealth depends only on u_1, and not on δ_t. Therefore, the change in expected utility can be computed by taking the expectation, with respect to u_1, of the change in wealth times M_A, the expected marginal utility of wealth conditional on u_1.

For the arbitrageurs, the direct effect can be nonzero, since the financial constraint could be binding in period 0. To determine the direct effect, we note that if prices are held constant, changing x_0 changes the arbitrageurs' period 1 wealth by $2(\phi_0 - \phi_1)$. The resulting change in expected utility is the expectation, with respect to u_1, of the change in wealth times $R_1 M$, the marginal utility of wealth received in period 1. The indirect effects are as for the A- and B-investors, with the difference that the change in period $t+1$ wealth is multiplied by $R_{t+1}M$, the marginal utility of wealth received in that period.[14]

We can now state the paper's main result.

Proposition 3 *The arbitrageurs' period 0 position may fail to be socially optimal. It sometimes involves too much and sometimes too little risk.*

Before the formal analysis, it is worth giving a broad intuition for the result. Consider first the case in which, in equilibrium, arbitrageurs are heavily invested in the arbitrage opportunity. Suppose that after they have chosen their positions, the other investors' relative demand increases. These investors are then eager for liquidity (in this and subsequent periods), which arbitrageurs are eager to provide because the price wedge is wide. However, the arbitrageurs' ability to do so is limited due to the capital losses on their positions. Reducing the positions would limit the losses, and enable arbitrageurs to provide more liquidity. Of course, arbitrageurs internalize that liquidity provision is profitable, which is why they may choose positions below the financial constraint. However, what competitive arbitrageurs fail to internalize is that reducing their positions affects prices, i.e., that with smaller losses, they can invest more aggressively, thus attenuating the widening of the price wedge. This

[14]The change in period $t+1$ wealth is equal to the change in the capital gain on the arbitrage opportunity between periods t and $t+1$, $2d(\phi_t - \phi_{t+1})/dx_0$, times the arbitrageurs' period t position, x_t. For $t \geq 1$, we also need to multiply by R_t/R_{t+1}. This is because a change in the capital gain affects the financial constraint in period t, and thus affects x_t.

would further reduce their losses, allowing them to exploit the wide price wedge. Hence, they can be better off. For the other investors, the benefit of increased liquidity when they need it most can dominate the cost of the initial reduction in liquidity. Hence, they too can be better off. Altogether, reducing the arbitrageurs' positions can be Pareto improving.

Consider next the case in which, in equilibrium, arbitrageurs are not invested in the arbitrage opportunity initially, and again, suppose that the other investors' relative demand increases. An increase in the arbitrageurs' initial positions would further widen the price wedge, an effect they fail to internalize. Unlike in the previous case, however, arbitrageurs benefit from a wider price wedge since they can exploit it without having realized capital losses. Hence they can be better off. For the other investors, a wider price wedge implies less liquidity. However, this can be more than offset by the fact that the arbitrageurs' increased wealth allows them to provide more liquidity in later periods. Hence, the other investors too can be better off. Altogether, increasing the arbitrageurs' positions can be Pareto improving.

We should emphasize that these Pareto improvements occur through price changes. This is consistent with Eqs. (20) and (21). Indeed, the direct effect in Eq. (21) always reduces the arbitrageurs' utility, since it is equal to the change in utility holding prices constant, and arbitrageurs maximize utility. Therefore, a Pareto improvement can occur only through the indirect effects, i.e., through a change in prices. The intuition is that a change in prices induces a redistribution of wealth, and this can be Pareto improving because agents' marginal rates of substitution (MRS) differ.

In the remainder of this section, we explain why agents' MRS differ. We complete our analysis in Section 5, where we derive the equilibrium in closed form in a special case, and determine the redistribution of wealth achieved by a change in prices. As we show in that section, both the redistribution across time and across states of nature are necessary for a Pareto improvement.

Recall the coefficients translating a change in period t wealth to a change in expected utility. For the i-investors, the coefficient is M_i, the expected marginal utility of wealth conditional on u_1. For the arbitrageurs, the coefficient is $R_t M$, the marginal utility of wealth received in period t.

Consider now the MRS across time periods, say t and $t' > t$. For the i-investors, M_i is independent of t, and thus the MRS is equal to 1. The intuition is that at the margin, i-investors invest in the riskless asset, whose return is 1. For the arbitrageurs, the MRS is $R_t M / R_{t'} M = R_t / R_{t'}$. When a price wedge remains, R_t decreases over time, and thus the MRS is greater than 1. The intuition is that at the margin, arbitrageurs invest in an arbitrage position, whose return exceeds 1. Since the arbitrageurs' MRS can exceed that of the i-investors, the arbitrageurs have a greater preference for receiving wealth in the early periods. Therefore, a redistribution of wealth that benefits the arbitrageurs in the early periods and the i-investors in the later periods has the potential to be Pareto improving.

Consider next the MRS across states, i.e., for different values of u_1, in the special case where $t = 1$, and where the arbitrageurs' financial constraint in period 0 is slack. The arbitrageurs' optimality condition is

$$E\left[(\phi_0 - \phi_1) R_1 M\right] = 0,$$

and that of the i-investors is

$$E\left[(\phi_0 - \phi_1 - f'(y_0 + u_0)) M_i\right] = 0.$$

For these equations to be consistent, $R_1 M$ must give more weight, relative to M_i, to the states where ϕ_1 is high. In other words, arbitrageurs must have a greater preference, relative to i-investors, for receiving wealth in states where the price wedge widens. The intuition is that at the margin, both the i-investors and the arbitrageurs invest in the riskless asset. Since the arbitrageurs do not face asset payoff (δ_t) risk, they must be more adversely affected (relative to the i-investors) by the risk that the price wedge widens. When this is the case, a redistribution of wealth that benefits the arbitrageurs in states where the price wedge widens, and the i-investors otherwise, has the potential to be Pareto improving.[15]

5. The continuous-time case

To gain further insight into the equilibrium, it is desirable to solve the model in closed form. One would expect this to be easiest for $T = 2$, i.e., when the number of periods is the smallest. However, the case $T = 2$ is not very tractable and, in addition, it fails to capture some interesting economic effects. In particular, the redistribution of wealth induced by a change in prices occurs not across time, but only across states of nature, and this turns out to be insufficient for a Pareto improvement.[16]

[15] Differences in MRS across states could be eliminated by trading claims contingent on u_1. These claims, however, would need to be contingent on the prices of both assets A and B, which would run counter to the market segmentation assumption.

[16] The analysis of the case $T = 2$ is available upon request. An additional limitation of that case is that a supply shock in period 1 increases, rather than decreases, the arbitrageurs' period 1 position. Intuitively,

In this section, we solve the model in the "continuous-time" case, where the number of periods goes to infinity, while the "calendar" time between the first and the last period remains constant. The continuous-time case captures the effects we want to consider, while remaining very tractable and allowing for closed-form solutions.

5.1. *The model*

We assume that calendar time, θ, belongs to an interval $[0, \Theta]$ that contains $T+1$ equally spaced periods. Period t, $t = 0, 1, .., T$, corresponds to calendar time $\theta = th$, where $h \equiv \Theta/T$ represents the calendar time between two consecutive periods. To obtain the continuous-time case, we assume that the number of periods, $T+1$, goes to infinity, and thus h goes to 0. In addition, we assume that the probability distribution of δ_t, the asset payoff information in period t, is given by

Outcome	$-\overline{\delta}$	$-\hat{\delta}\sqrt{h}$	$\hat{\delta}\sqrt{h}$	$\overline{\delta}$
Probability	h^a	$\frac{1}{2} - h^a$	$\frac{1}{2} - h^a$	h^a

and the probability distribution of u_1, the supply shock in period 1, is given by

Outcome	$u_0 - \hat{u}\sqrt{h}$	$u_0 + \hat{u}\sqrt{h}$
Probability	$\frac{1}{2}$	$\frac{1}{2}$

where $\overline{\delta}$, $\hat{\delta}$, $a > 1$, and $\hat{u}$ are constants that do not depend on h. The information δ_t on the asset payoff has two components: a "diffusion" component $\hat{\delta}\sqrt{h}$, which converges to a Brownian motion when h goes to zero, and a "jump" component $\overline{\delta}$. Without the latter, the maximum loss of a position within one period would converge to zero, and thus the financial constraint would vanish. We assume $a \neq 1$ only for simplicity. For simplicity, we also omit the jump component from the supply shock u_1. Note that the certainty case is obtained by setting $\hat{u} = 0$.

5.2. *Equilibrium*

We treat all variables as functions of the calendar time θ, and use the subscript θ rather than t. To study the equilibrium, we determine the asymptotic behavior of the risk premium

the supply shock has two effects. First, it decreases the arbitrageurs' period 1 wealth, through the widening of the price wedge. Second, it increases the profitability of the positions that arbitrageurs can establish in period 1 in assets A and B. This reduces the maximum loss of these positions, thus relaxing the financial constraint in period 1. The first effect dominates only when the number of periods is large enough (and, in particular, in the continuous-time case, as shown in Lemma 6). This is because the widening of the price wedge reflects the expectation of the supply shocks over all future periods, while the financial constraint concerns the maximum loss of a position only over the next period.

ϕ_θ and the arbitrageurs' position x_θ, when h goes to 0.

5.2.1. *The certainty case*

To focus on the case in which the financial constraint is binding, we assume $w_0 < \overline{w}_0 \equiv \frac{2\overline{\delta}u_0}{\mu}$. In this case, x_θ is given by Eq. (8), which we can write in terms of calendar time as

$$\begin{aligned} x_\theta - x_\theta \frac{f'(u_0 - \mu x_\theta)}{\overline{\delta}} &= x_{\theta-h} \\ \Rightarrow \quad \frac{x_\theta - x_{\theta-h}}{h} &= x_\theta \frac{f'(u_0 - \mu x_\theta)}{h\overline{\delta}}. \end{aligned} \tag{22}$$

To study Eq. (22), we use the following lemma.

Lemma 3 *We have*

$$\lim_{h\to 0} \frac{f'(y)}{h} = \alpha\hat{\delta}^2 y.$$

Lemma 3 implies that when h goes to 0, the inventory cost $f(y)$ is linear in h, the calendar time between two consecutive periods, and quadratic in y, the position in the risky asset. In addition, it is increasing in α, the risk-aversion coefficient, and in $\hat{\delta}$, which is a measure of asset payoff risk.

Taking the limit of Eq. (22) when h goes to 0, and using Lemma 3, we get

$$\frac{dx_\theta}{d\theta} = x_\theta \frac{\alpha\hat{\delta}^2(u_0 - \mu x_\theta)}{\overline{\delta}}. \tag{23}$$

Using Eq. (9), and following a similar procedure, we get

$$\frac{d\phi_\theta}{d\theta} = -\alpha\hat{\delta}^2(u_0 - \mu x_\theta). \tag{24}$$

The initial conditions for the differential Eqs. (23) and (24) are $x_0 = \frac{w_0}{2\overline{\delta}}$, which follows from Eq. (7), and $\phi_\Theta = 0$. We denote the solutions to the differential equations by x^*_θ and ϕ^*_θ. These functions represent the limits of x_θ and ϕ_θ when h goes to zero.

Lemma 4 *When h goes to 0, x_θ goes to*

$$x^*_\theta = \frac{w_0}{2\overline{\delta}} \frac{q}{1 + (q-1)e^{-r\theta}}, \tag{25}$$

and ϕ_θ goes to

$$\phi^*_\theta = \overline{\delta} \log \frac{1 + (q-1)e^{-r\theta}}{1 + (q-1)e^{-r\Theta}}, \tag{26}$$

where

$$q = \frac{\overline{w}_0}{w_0} > 1 \qquad \textit{and} \qquad r = \frac{u_0\alpha\hat{\delta}^2}{\overline{\delta}}.$$

Eq. (25) implies that x^*_θ increases with θ. This confirms the general result of Proposition 1 that the arbitrageurs' position increases over time. Eq. (25) also implies that when θ goes to infinity, x^*_θ converges to $(w_0/2\overline{\delta})q = u_0/\mu$. This means that when the calendar time interval is long enough, the arbitrageurs accumulate enough wealth from their arbitrage activity to be able to fully absorb the A- and B-investors' supply shock.

5.2.2. The uncertainty case

We first note that the limits of ϕ_θ and x_θ when h goes to zero are the same as in the certainty case, i.e., ϕ^*_θ and x^*_θ, except perhaps for the limit of x_0. This is because from period 1 on, the uncertainty case is identical to the certainty case. Moreover, when h goes to zero, period 1 converges to calendar time $\theta = 0$, and the "state variables" u_h and w_h, which correspond to period 1, converge to u_0 and w_0. (For w_h, this is because the risk between periods 0 and 1 converges to zero, and thus the difference between the risk premia in these periods – which determines the arbitrageurs' capital gain – also converges to zero.) Therefore, for all $\theta > 0$, the limits of ϕ_θ and x_θ are the same as in the certainty case. The limit of ϕ_0 is also the same because the difference between the risk premia in periods 0 and 1 converges to 0.

We next study the asymptotic behavior of ϕ_θ and x_θ around their limits. This will reveal some interesting properties of the equilibrium which, in addition, are relevant for the welfare analysis. We focus on the asymptotic behavior of $\phi_0 - \phi_h$, the difference between the risk premia in periods 0 and 1, and x_θ for $\theta \geq h$, the arbitrageurs' position in periods $t \geq 1$. To state our results, we define

$$s = \frac{2\tilde{x}^*_0\overline{\delta}}{w_0},$$

where $\tilde{x}^*_0$ denotes the limit of x_0 in the uncertainty case. The variable s measures the extent to which arbitrageurs invest in the arbitrage opportunity in period 0, when h goes to zero. Notice that when h goes to zero, the financial constraint in period 0, i.e., Eq. (16), becomes

$$w_0 \geq 2|\tilde{x}_0|\overline{\delta}.$$

Therefore, the financial constraint is binding upwards when $s = 1$, is slack when $|s| < 1$, and is binding downwards when $s = -1$.

Lemma 5 *We have*

$$\phi_0 - \phi_h = -\Phi(u_h - u_0) + \hat{\Phi}h + o(h), \tag{27}$$

where

$$\Phi = \frac{\overline{\delta}}{u_0} \frac{1 - e^{-r\Theta} + (q-1)r\Theta e^{-r\Theta}}{(1-s)(1-e^{-r\Theta}) + qe^{-r\Theta}}, \tag{28}$$

and

$$\hat{\Phi} = r\overline{\delta}\left(1 - \frac{s}{q}\right) + \frac{1}{2}\alpha\Phi\hat{u}^2 \left[\phi_0^* + \Phi u_0 \left(\left(1 - \frac{s}{q}\right) + e^{\frac{\phi_0^*}{\overline{\delta}}} \frac{1}{q}(1-s)\right)\right]. \tag{29}$$

The term $-\Phi(u_h - u_0)$ in the Taylor expansion (27) is stochastic, has zero expectation, and is of order $\sqrt{h}$. This term reflects the uncertainty that the period 1 supply shock introduces into the period 1 risk premium. The effect of the supply shock is measured by the coefficient Φ. This coefficient is positive, which simply means that the supply shock increases the risk premium. In addition, Φ increases with s, which means that the more heavily arbitrageurs are invested in the arbitrage opportunity in period 0, the more they amplify the effect of the period 1 supply shock on prices. This amplification effect is through the arbitrageurs' period 1 wealth, and is similar to effects derived in a number of recent papers.[17]

The term $\hat{\Phi}h$ is deterministic and of order h. This term is equal to the expected difference between the risk premia in periods 0 and 1, i.e., the expected capital gain per unit of asset A. This expected capital gain is determined by the optimality condition of the A-investors, and reflects the compensation that these investors require for bearing risk.

Lemma 6 *For $\theta \geq h$, we have*

$$x_\theta - x_\theta^* = X_\theta(u_h - u_0) + o\left(\sqrt{h}\right), \tag{30}$$

where

$$X_\theta = \frac{1}{\mu(1 + (q-1)e^{-r\theta})^2}\left[1 - e^{-r\theta} + (q-1)r\theta e^{-r\theta} - sqe^{-r\theta}\frac{1 - e^{-r\Theta} + (q-1)r\Theta e^{-r\Theta}}{(1-s)(1-e^{-r\Theta}) + qe^{-r\Theta}}\right]. \tag{31}$$

Moreover, $X_\theta < 0$ for $\theta \in [0, \theta(s))$ and $X_\theta > 0$ for $\theta \in (\theta(s), \Theta]$, where $\theta(s)$ is an increasing function such that $\theta(0) = 0$ and $\theta(1) = \Theta$.

The coefficient X_θ measures the effect of the period 1 supply shock on the arbitrageurs' position x_θ. When $s \in (0,1)$, i.e., arbitrageurs invest in period 0 but not up to the financial constraint, the sign of the effect changes over time. In a first phase, for $\theta \in [h, \theta(s))$, the supply shock reduces the arbitrageurs' position. The intuition is that the supply shock

[17]For example, papers on the limits of arbitrage (Shleifer and Vishny, 1997; Xiong, 2001), margin requirements (Aiyagari and Gertler, 1999; Sodini, 2001), macroeconomic credit cycles (Kiyotaki and Moore, 1998; Krishnamurthy, 2000; Mian, 2002), and corporate asset fire sales (Shleifer and Vishny, 1992).

increases the risk premium in period 1, and thus reduces the arbitrageurs' wealth in that period. In a second phase, for $\theta \in (\theta(s), \Theta]$, however, the supply shock increases the arbitrageurs' position. The intuition is that the supply shock increases the profitability of an arbitrage position established after period 1. This increases the rate at which arbitrageurs accumulate wealth and, eventually, increases their wealth. When $s = 1$, i.e., arbitrageurs invest up to the financial constraint, the reduction in period 1 wealth dominates the increase in the accumulation rate, and the second phase disappears. By contrast, when $s \leq 0$, i.e., arbitrageurs either do not invest or short the arbitrage opportunity, the period 1 wealth does not decrease, and the first phase disappears.

We finally determine the limit of the arbitrageurs' period 0 position, x_0.

Lemma 7 *Set*

$$g(s) = \hat{\Phi} - \frac{\Phi^2}{\overline{\overline{\delta}}}\hat{u}^2(1 - \gamma(1 - s)), \tag{32}$$

where γ is the coefficient of relative risk aversion of the arbitrageurs' utility function U at the point $w_0 e^{\frac{\phi_0^*}{\overline{\delta}}}$. Then

$$\begin{aligned} s = 1 \quad & \textit{if} \quad g(1) \geq 0, \\ g(s) = 0 \quad \textit{and} \quad s \in (-1, 1) \quad & \textit{if} \quad g(-1) > 0 > g(1), \\ s = -1 \quad & \textit{if} \quad g(-1) \leq 0. \end{aligned}$$

Arbitrageurs do not invest up to the financial constraint if $g(1) < 0$. Using the definition of $\hat{\Phi}$, we can write this equation as

$$r\overline{\delta}\left(1 - \frac{1}{q}\right) + \Phi\hat{u}^2\left[\frac{1}{2}\alpha\phi_0^* + \frac{1}{2}\alpha\Phi u_0\left(1 - \frac{1}{q}\right) - \frac{\Phi}{\overline{\overline{\delta}}}\right] < 0. \tag{33}$$

Eq. (33) is satisfied if $\hat{u}$ is large, i.e., there is enough uncertainty about the period 1 supply shock, and α is small, i.e., A- and B-investors are not very risk-averse. There must be enough uncertainty so that the arbitrageurs' risk-management motive to invest conservatively is important. Furthermore, A- and B-investors must not very risk-averse, otherwise the expected capital gain on the arbitrage opportunity would become large, making the opportunity very attractive to arbitrageurs.

5.3. Welfare

We next turn to the welfare analysis, and consider the derivatives of agents' expected utilities with respect to the arbitrageurs' period 0 position, x_0. In the limit when h goes to zero, the effect of changing the arbitrageurs' position in any given period becomes negligible. Therefore, the derivatives of agents' expected utilities converge to zero, and their sign for h close to zero depends on their asymptotic behavior around their limit.

Lemma 8 *The derivative of the i-investors' expected utility with respect to x_0 has the form*

$$\Psi_i h + o(h),$$

and the derivative of the arbitrageurs' expected utility has the form

$$\Psi h + o(h),$$

where $i = A, B$, and Ψ_i and Ψ are given by Eqs. (46) and (52) in the Appendix.

Lemma 8 implies that for small h, the sign of the derivatives is the same as that of Ψ_i and Ψ.

5.3.1. The certainty case

In the certainty case, the financial constraint in period 0 is binding upwards, and thus the social planner can only reduce the arbitrageurs' position.

Proposition 4 *In the certainty case, $\Psi_i > 0$, while Ψ can have either sign.*

Proposition 4 implies that reducing the arbitrageurs' period 0 position makes the *i*-investors worse off, and can make the arbitrageurs better or worse off. This means that in the certainty case, the social planner cannot achieve a Pareto improvement, and the arbitrageurs' position is socially optimal.

To explain the intuition, we use Eqs. (20) and (21), which give the derivatives of agents' expected utilities in the general case. The derivative of the arbitrageurs' expected utility is

$$E\left[\left(2(\phi_0 - \phi_1)R_1 + 2\frac{d(\phi_0 - \phi_1)}{dx_0}x_0 R_1 + 2\sum_{t=1}^{T-1}\frac{d(\phi_t - \phi_{t+1})}{dx_0}x_t R_t\right) M\right].$$

The first term in this equation represents the direct effect of x_0. Reducing x_0 means that arbitrageurs invest less aggressively in period 0. Holding prices constant, this reduces the arbitrageurs' period 1 wealth. The second term represents the indirect effect of x_0 through the capital gain on the arbitrage opportunity between periods 0 and 1. Since arbitrageurs invest less aggressively in period 0, the capital gain increases, and so does the arbitrageurs' period 1 wealth. Finally, the third term represents the indirect effect of x_0 through the capital gain on the arbitrage opportunity from period 1 on. This effect is through the arbitrageurs' period 1 wealth: if, for example, wealth increases, then the arbitrageurs can invest more aggressively, and the capital gain decreases. Reducing x_0 can make the arbitrageurs better or worse off, simply because it can increase or decrease their period 1 wealth. It is worth pointing out that the effects of x_0 on period 1 wealth are identical to those of the quantity chosen by a monopoly on the monopoly's profits.

The derivative of the i-investors' expected utility is

$$E\left[\left(\sum_{t=0}^{T-1}\frac{d(\phi_t-\phi_{t+1})}{dx_0}y_t\right)M_i\right].$$

The effects of x_0 for the i-investors are indirect, and of opposite signs to the indirect effects for the arbitrageurs. Intuitively, an increase in the capital gain on the arbitrage opportunity makes the arbitrageurs better off. It makes, however, the i-investors worse off, since the price wedge is wider, and thus these investors obtain less liquidity. Reducing x_0 increases the capital gain on the arbitrage opportunity between periods 0 and 1, and can either decrease or increase the capital gain from period 1 on. The i-investors are worse off because the decrease in liquidity in period 0 dominates any increase in liquidity from period 1 on.

5.3.2. The uncertainty case

For simplicity, we focus on the case in which the financial constraint in period 0 is slack. From Eq. (33), this occurs when there is enough uncertainty, and the i-investors are not very risk-averse.

Proposition 5 *In the uncertainty case:*

- *If Θ is small, then $\Psi_i > 0$ and $\Psi < 0$.*
- *If Θ is large and s is close to 1, then $\Psi_i < 0$ and $\Psi < 0$.*
- *If Θ is large and s is close to 0, then $\Psi > 0$, while Ψ_i can have either sign.*

Proposition 5 implies that the arbitrageurs' period 0 position is socially optimal when the calendar time interval is short (i.e., Θ is small). When the calendar time interval is long, however, the arbitrageurs' position may involve too much or too little risk. It involves too much risk when arbitrageurs are almost fully invested in the arbitrage opportunity (i.e., s

is close to 1), which in turn occurs when there is enough, but not too much, uncertainty. It may involve too little risk when arbitrageurs are not invested in the arbitrage opportunity (i.e., s is close to zero).

To explain the intuition, we consider the indirect effects of x_0 for the arbitrageurs. Consider first the indirect effect through the capital gain on the arbitrage opportunity between periods 0 and 1. Reducing x_0, and holding prices constant, increases the arbitrageurs' period 1 wealth more in the "bad" state, where the price wedge widens, than in the "good" state, where it narrows. This means that the arbitrageurs' ability to invest in period 1, and to close the price wedge, increases more in the bad state. Therefore, the capital gain on the arbitrage opportunity between periods 0 and 1 increases more in the bad state, and so does the arbitrageurs' period 1 wealth through the indirect effect. The indirect effect is, in fact, the "flip-side" of the amplification effect of Lemma 5, i.e., if arbitrageurs invest more aggressively, they amplify the effect of the supply shock on prices.

Consider next the indirect effect through the capital gain on the arbitrage opportunity from period 1 on. As explained in the previous paragraph, reducing x_0 increases the arbitrageurs' period 1 wealth, and ability to invest, more in the bad state. Therefore, the capital gain on the arbitrage opportunity from period 1 on decreases more in the bad state, and so does the rate at which arbitrageurs accumulate wealth from period 1 on.

Suppose now that the calendar time interval is long, and that in period 0 arbitrageurs are almost fully invested in the arbitrage opportunity. Reducing x_0 transfers wealth to the arbitrageurs in period 1 and in the bad state. (In what follows, we consider transfers relative to the certainty case.) This transfer, which occurs through the change in the capital gain between periods 0 and 1, is important, since in period 0 arbitrageurs are almost fully invested in the arbitrage opportunity. Reducing x_0 also transfers wealth away from the arbitrageurs in periods $t > 1$ and in the bad state. This transfer, which occurs through the change in the capital gain from period 1 on, is important when the arbitrageurs' investment is higher in the bad than in the good state. This occurs in the later periods, during the second phase described in Lemma 6.

Overall, wealth is transferred to the arbitrageurs in the early periods and in the bad state, and away from them in the later periods or in the good state. This redistribution of wealth has the potential to be (and, in fact, is) Pareto improving, because it is in the early periods and in the bad state when arbitrageurs value wealth the most.

Notice that a redistribution both across time and across states is required for achieving the Pareto improvement. Indeed, when the calendar time interval is short, and the redistribution across time becomes negligible, the arbitrageurs' position is socially optimal. Social optimality also holds in the certainty case, where there is no redistribution across states. A redistribution both across time and across states matters for the i-investors. These investors are made better off because they obtain liquidity in the later periods, and more so in the bad state where they value liquidity the most.

Suppose finally that the calendar time interval is long, and that in period 0 arbitrageurs are not invested in the arbitrage opportunity. Reducing x_0 does not imply any transfers in period 1, since the arbitrageurs' period 0 investment is zero. Therefore, the only transfers are in periods $t > 1$, and are away from the arbitrageurs in the bad state. This redistribution of wealth has the potential to make all agents worse off, since it is in the bad state where arbitrageurs value wealth the most. Therefore, the reverse redistribution of wealth, achieved by increasing x_0, has the potential to be Pareto improving. In fact, as shown in the Appendix, it is indeed Pareto improving when the arbitrageurs and the i-investors are not very risk-averse.

6. Concluding remarks

We propose a multiperiod model in which competitive arbitrageurs exploit discrepancies between the prices of two identical risky assets traded in segmented markets. Arbitrageurs need to collateralize separately their positions in each asset, and this implies a financial constraint limiting positions as a function of wealth. In our model, arbitrage activity benefits all investors because arbitrageurs supply liquidity to the market. However, arbitrageurs might fail to take a socially optimal level of risk, in the sense that a change in their positions can make all investors better off. We characterize conditions under which arbitrageurs take too much or too little risk.

Besides showing that the market outcome can fail to be allocatively efficient, we clarify the source of the inefficiency. This is that (i) competitive arbitrageurs fail to internalize that changing their positions affects prices, and (ii) due to market segmentation and financial constraints, agents' marginal rates of substitution differ, and so a redistribution of wealth induced by a change in prices can be Pareto improving. This source of inefficiency has already been identified in the literature on general equilibrium with incomplete markets. Our contribution is to explore the direction of the inefficiency when incompleteness is created by market segmentation and financial constraints. We also argue that this particular form of incompleteness could be fruitful for understanding features of financial markets and policy debates regarding these markets.

An example is the debate on systemic risk, i.e., on whether a worsening of the financial condition of some market participants can propagate into the financial system with harmful effects. Our model captures features of systemic risk, since a reduction in arbitrageurs' wealth can be detrimental not only to other arbitrageurs but also to other investors. In addition, our model lends theoretical support to the notion that market participants could fail to take an appropriate level of risk ex ante, an important argument in the systemic risk debate.

A second debate concerns the regulation of financial institutions. Much of this regulation (e.g., capital requirements) aims at controlling problems arising from the possibility of default. Our model rules out default since margin accounts need to be fully collateralized. That inefficiencies can still arise suggests that there might be a motive for regulation even without reference to default. Regulations could be aimed at alleviating the distortions on arbitrageurs' investments induced by frictions in their access to capital.

While our model suggests a motive for regulations, it is too stylized for discussing their implementation or effectiveness. For example, it is not explicit about what regulators can do or know, relative to the agents in the economy in particular. Therefore, our discussion of policy implications should be interpreted as speculative and preliminary. Nevertheless, we find these questions interesting and intend to address them more thoroughly in future research.

One channel of regulatory intervention is to affect arbitrageurs' financial constraints. Regulators might have some control over these constraints by altering, for example, arbitrageurs' capital or margin requirements, or by influencing their financiers (e.g., regulating investment banks' lending to hedge funds). Such issues were raised in the wake of 1998 crisis (Edwards, 1999). Our results suggest that a regulator with only limited control over financial constraints might prefer to tighten the constraints in some cases, since this may reduce overinvestment.

In our model, inefficiency also arises from the lack of entry into the arbitrage industry. While a full welfare discussion of entry is delicate, some insight can be gained by examining A- and B-investors' welfare. Sufficient entry would remove segmentation, and maximize A- and B-investors' welfare. Short of this level, however, entry might exacerbate arbitrageurs' overinvestment and be detrimental to A- and B-investors. Therefore, policies limiting entry can have benefits. Incidentally, while limiting entry might also reduce competition among arbitrageurs, this could come with benefits. Indeed, in our model, the arbitrageurs' overinvestment is linked to their price-taking behavior, a problem that would be reduced by monopolization (or cooperation among arbitrageurs). One can therefore conceive of A- and B-investors being better off with a less than perfectly competitive arbitrage industry.

Contagion phenomena can arise naturally in our framework. Suppose that arbitrageurs can take several unrelated arbitrage positions. An adverse shock to one position would trigger the liquidation of some of the other positions, creating a price linkage between otherwise unrelated assets (see, for example, Krishnamurthy, 2000; Kyle and Xiong, 2001). Our framework could be used to analyze the implications of contagion for welfare and policy. For example, the equilibrium level of liquidation itself could be inefficient (see Gromb and Vayanos, 2000).

APPENDIX

Proof to Lemma 1: To show that f is positive, we use Jensen's inequality:

$$\exp(\alpha f(y)) = E\exp(-\alpha y\delta) > \exp\left[E(-\alpha y\delta)\right] = \exp\left[-\alpha y E(\delta)\right] = 1.$$

(Jensen's inequality is strict since δ_t is stochastic.)

To show that f is strictly convex, we compute its second derivative. We have

$$f(y) = \frac{1}{\alpha}\log\left[E\exp(-\alpha y\delta)\right].$$

Therefore,

$$f'(y) = -\frac{E\left(\delta\exp(-\alpha y\delta)\right)}{E\exp(-\alpha y\delta)}, \tag{34}$$

and

$$f''(y) = \alpha\frac{E\left(\delta^2\exp(-\alpha y\delta)\right)E\exp(-\alpha y\delta) - \left[E\left(\delta\exp(-\alpha y\delta)\right)\right]^2}{\left[E\exp(-\alpha y\delta)\right]^2}.$$

That $f''(y) > 0$ follows from the Cauchy-Schwarz inequality

$$E(gh)^2 \leq E(g^2)E(h^2),$$

for the functions $g = \delta\exp(-\alpha y\delta/2)$ and $h = \exp(-\alpha y\delta/2)$. (The Cauchy-Schwarz inequality is strict since δ_t is stochastic, and thus g and h are not proportional.)

To show that $f(y) = f(-y)$, we use the symmetry of the probability distribution of δ around zero:

$$\exp(\alpha f(y)) = E\exp(-\alpha y\delta) = E\exp(\alpha y\delta) = \exp(\alpha f(-y)).$$

Finally, to show that $\lim_{y\to\infty} f'(y) = \overline{\delta}$, we note that

$$|f'(y) - \overline{\delta}| = \left|-\frac{E\left[\delta\exp(-\alpha y\delta)\right]}{E\exp(-\alpha y\delta)} - \overline{\delta}\right| = \left|\frac{E\left[(\delta+\overline{\delta})\exp(-\alpha y\delta)\right]}{E\exp(-\alpha y\delta)}\right|.$$

To show that the last term goes to zero when y goes to ∞, we fix $\epsilon > 0$. We have

$$\left|\frac{E\left[(\delta+\overline{\delta})\exp(-\alpha y\delta)\mathbf{1}_{\delta\leq-\overline{\delta}+\epsilon}\right]}{E\exp(-\alpha y\delta)}\right| \leq \epsilon\left|\frac{E\left[\exp(-\alpha y\delta)\mathbf{1}_{\delta\leq-\overline{\delta}+\epsilon}\right]}{E\exp(-\alpha y\delta)}\right| \leq \epsilon.$$

Moreover, for y large enough,

$$\left|\frac{E\left[(\delta+\overline{\delta})\exp(-\alpha y\delta)\mathbf{1}_{\delta>-\overline{\delta}+\epsilon}\right]}{E\exp(-\alpha y\delta)}\right|$$

becomes smaller than ϵ. Q.E.D.

Proof to Proposition 1: Suppose first that $w_0 \geq \overline{w}_0$. We will show that the only equilibrium can be $\phi_t = 0$ and $\mu x_t = u_0$. We proceed by induction. The induction hypothesis, $\mathcal{H}_t$, is that for all $s \leq t$, we have $\phi_s - \phi_{s+1} = 0$ and $\mu x_s = u_0$. We assume $\mathcal{H}_{t-1}$, and will show $\mathcal{H}_t$. (To start the induction, we note that $\mathcal{H}_{-1}$ obviously holds.) If $\phi_t - \phi_{t+1} > 0$, then the financial constraint is binding upwards, i.e.,

$$x_t = \frac{w_t}{2\left(\overline{\delta} - (\phi_t - \phi_{t+1})\right)}. \tag{35}$$

In that case, however,

$$u_0 - \mu x_t < u_0 - \mu \frac{w_t}{2\overline{\delta}} = u_0 - \mu \frac{w_0}{2\overline{\delta}} < 0,$$

and thus

$$\phi_t - \phi_{t+1} = f'(u_0 - \mu x_t) < 0,$$

a contradiction. (Since $f(y) = f(-y)$, we have $f'(0) = 0$. Moreover, since f is strictly convex, we have $f'(y) > 0$ for $y > 0$, and $f'(y) < 0$ for $y < 0$.) If $\phi_t - \phi_{t+1} < 0$, then the financial constraint is binding downwards. In that case, however,

$$u_0 - \mu x_t > u_0 > 0,$$

and thus

$$\phi_t - \phi_{t+1} = f'(u_0 - \mu x_t) > 0,$$

a contradiction. Therefore, we must have $\phi_t - \phi_{t+1} = 0$. Since

$$f'(u_0 - \mu x_t) = \phi_t - \phi_{t+1} = 0,$$

we must also have $\mu x_t = u_0$. Therefore, $\mathcal{H}_t$ holds, and thus the only equilibrium can be $\phi_t = 0$ and $\mu x_t = u_0$. It is easy to check that this is indeed an equilibrium.

Suppose next that $w_0 < \overline{w}_0$. We first show that for $x_t \geq 0$, the difference equation given by (7) and (8) has a unique solution, and this solution satisfies $\mu x_t < u_0$. We proceed by induction. The induction hypothesis, $\mathcal{H}_t$, is that the above is true for all $s \leq t$. We assume $\mathcal{H}_{t-1}$, and will show $\mathcal{H}_t$. Suppose first that $t = 0$. The LHS of Eq. (7) is equal to zero for $x_0 = 0$, and to $u_0/\mu > w_0/2\overline{\delta}$ for $x_0 = u_0/\mu$. Moreover, the derivative of the LHS is

$$1 - \frac{f'(u_0 - \mu x_0)}{\overline{\delta}} + \mu x_0 \frac{f''(u_0 - \mu x_0)}{\overline{\delta}},$$

and is positive for $x_0 \geq 0$, since $f'(y) < \overline{\delta}$ and $f''(y) > 0$. (These properties follow from the strict convexity of f, and from $\lim_{y\to\infty} f'(y) = \overline{\delta}$.) Therefore, for $x_0 \geq 0$, Eq. (7) has a unique solution, and this solution satisfies $\mu x_0 < u_0$. Therefore, $\mathcal{H}_0$ holds. For $t > 0$, we use Eq. (8) and proceed as above.

We next show that the only equilibrium is the one given by Eqs. (7), (8), and (9). We proceed by induction. The induction hypothesis, $\mathcal{H}_t$, is that for all $s \leq t$, x_s and $\phi_s - \phi_{s+1}$ are given by Eqs. (7), (8), and (9), and moreover $w_{s+1} = 2x_s\overline{\delta}$. We assume $\mathcal{H}_{t-1}$, and will show $\mathcal{H}_t$. If $\phi_t - \phi_{t+1} = 0$, then the financial constraint is $x_t \leq w_t/2\overline{\delta}$. In that case, however,

$$u_0 - \mu x_t \geq u_0 - \mu \frac{w_t}{2\overline{\delta}} = u_0 - \mu x_{t-1} > 0,$$

and thus

$$\phi_t - \phi_{t+1} = f'(u_0 - \mu x_t) > 0,$$

a contradiction. If $\phi_t - \phi_{t+1} < 0$, we get a contradiction as in the case $w_0 \geq \overline{w}_0$. Therefore, we must have $\phi_t - \phi_{t+1} > 0$. The financial constraint is thus binding upwards, i.e., Eq. (35) holds. We can write this equation as

$$w_t = 2x_t\overline{\delta} - 2x_t(\phi_t - \phi_{t+1}). \tag{36}$$

Combining Eq. (36) with Eq. (9), we get Eq. (7) for $t = 0$, and Eq. (8) for $t > 0$. Moreover, Eq. (36) implies that

$$w_{t+1} = w_t + 2x_t(\phi_t - \phi_{t+1}) = 2x_t\overline{\delta}.$$

Therefore, $\mathcal{H}_t$ holds, and the only equilibrium is the one given by Eqs. (7), (8), and (9). It is easy to check that this is indeed an equilibrium. Q.E.D.

Proof to Lemma 2: Consider first the no-trade allocation. All investors are worse off than under the equilibrium allocation because, when faced with the equilibrium prices, they can always choose not to trade.

Consider next the no-constraint allocation. Since $\phi_t = 0$ for all t, the arbitrageurs do not realize any capital gains. Therefore, they are equally well off as under the no-trade allocation, and worse off than under the equilibrium allocation. Note that the arbitrageurs' period T wealth under the no-constraint allocation is w_0, and under the equilibrium allocation is

$$w_0 + 2\sum_{t=0}^{T-1} x_t(\phi_t - \phi_{t+1}) \equiv w_0 + C.$$

Since the arbitrageurs are better off under the equilibrium allocation, we must have $E(C) \geq 0$. The i-investors' period T wealth under the equilibrium allocation is

$$
\begin{aligned}
& w_{i,0} + \sum_{t=0}^{T-1} y_t(\phi_t - \phi_{t+1}) + (y_0 + u_0)\delta_1 + \sum_{t=1}^{T-1}(y_t + u_1)\delta_t \\
= \quad & w_{i,0} - \frac{\mu}{2}C + (y_0 + u_0)\delta_1 + \sum_{t=1}^{T-1}(y_t + u_1)\delta_t.
\end{aligned}
$$

The expected utility of receiving this wealth is smaller than that of receiving

$$w_{i,0} - \frac{\mu}{2}C,$$

since δ_t is a mean-preserving spread. Moreover, the expected utility of receiving the latter wealth is smaller than that of receiving $w_{i,0}$, the wealth under the no-constraint allocation, since $E(C) \geq 0$. Q.E.D.

Proof to Proposition 2: Eq. (20) follows by differentiating Eq. (12) with respect to x_0, and ignoring the terms in dy_t/dx_0 since y_t are unconstrained optima.

We next differentiate Eq. (15) with respect to x_0, and get

$$E\left[\left(2(\phi_0 - \phi_1)R_1 + 2\frac{d(\phi_0 - \phi_1)}{dx_0}x_0R_1 + (w_0 + 2x_0(\phi_0 - \phi_1))\frac{dR_1}{dx_0}\right)M\right].$$

Eq. (21) will follow if we show that

$$(w_0 + 2x_0(\phi_0 - \phi_1))\frac{dR_1}{dx_0} = 2\sum_{t=1}^{T-1}\frac{d(\phi_t - \phi_{t+1})}{dx_0}x_tR_t. \tag{37}$$

Eq. (10) implies that

$$\frac{dR_1}{dx_0} = \frac{d}{dx_0}\left(\prod_{t=1}^{T-1}\frac{1}{1 - \frac{\phi_t - \phi_{t+1}}{\bar{\delta}}}\right) = \sum_{t=1}^{T-1}\left(R_1\frac{1}{1 - \frac{\phi_t - \phi_{t+1}}{\bar{\delta}}}\frac{d(\phi_t - \phi_{t+1})}{dx_0}\frac{1}{\bar{\delta}}\right). \tag{38}$$

Since the solution of the certainty case always applies for $t \geq 1$, we can use Eq. (7) for $t = 1$. Combining this equation with Eq. (8), we get

$$x_t = \frac{w_1}{2\bar{\delta}}\prod_{s=1}^{t}\frac{1}{1 - \frac{\phi_s - \phi_{s+1}}{\bar{\delta}}}. \tag{39}$$

Eqs. (38) and (39) imply that

$$(w_0 + 2x_0(\phi_0 - \phi_1))\frac{dR_1}{dx_0} = w_1\frac{dR_1}{dx_0} = 2\sum_{t=1}^{T-1}\left(x_t\frac{d(\phi_t - \phi_{t+1})}{dx_0}\prod_{s=t}^{T-1}\frac{1}{1 - \frac{\phi_s - \phi_{s+1}}{\bar{\delta}}}\right).$$

Combining this equation with Eq. (10), we get Eq. (37). Q.E.D.

Proof to Proposition 3: The Proposition is a simple corollary of Proposition 5 (see proof below). Q.E.D.

Proof to Lemma 3: Using Eq. (34), and the probability distribution of δ_t, we have

$$f'(y) = \frac{h^a \overline{\delta} \sinh(\alpha y \overline{\delta}) + \left(\frac{1}{2} - h^a\right) \hat{\delta}\sqrt{h} \sinh(\alpha y \hat{\delta}\sqrt{h})}{h^a \cosh(\alpha y \overline{\delta}) + \left(\frac{1}{2} - h^a\right) \cosh(\alpha y \hat{\delta}\sqrt{h})},$$

where "sinh" and "cosh" denote hyperbolic sine and cosine, respectively. The Lemma follows by noting that

$$\lim_{h\to 0} \left[h^a \cosh(\alpha y \overline{\delta}) + \left(\frac{1}{2} - h^a\right) \cosh(\alpha y \hat{\delta}\sqrt{h})\right] = \frac{1}{2},$$

$$\lim_{h\to 0} \frac{h^a \overline{\delta} \sinh(\alpha y \overline{\delta})}{h} = 0,$$

(since $a > 1$), and

$$\lim_{h\to 0} \frac{\left(\frac{1}{2} - h^a\right) \hat{\delta}\sqrt{h} \sinh(\alpha y \hat{\delta}\sqrt{h})}{h} = \frac{1}{2}\alpha \hat{\delta}^2 y.$$

Q.E.D.

Proof to Lemma 4: To prove the Lemma, we need to show that the functions (25) and (26) solve the differential equations (23) and (24), with the initial conditions $x_0 = \frac{w_0}{2\overline{\delta}}$ and $\phi_\Theta = 0$. This follows from simple algebra. Q.E.D.

Proof to Lemma 5: We assume that the Taylor expansion of $\phi_0 - \phi_h$ has the form given by Eq. (27), and we determine Φ and $\hat{\Phi}$. To determine Φ, we consider the Taylor expansions of ϕ_0 and ϕ_h, separately, up to order $\sqrt{h}$. For ϕ_0, the term of order 0 is ϕ_0^*, and the term of order $\sqrt{h}$ is zero since it corresponds to the realization of the period 1 supply shock. For ϕ_h, the term of order 0 is also ϕ_0^*. The term of order $\sqrt{h}$ is the same as that obtained by evaluating

$$\phi_0^* = \overline{\delta} \log \frac{q}{1 + (q-1)e^{-r\Theta}}$$

at u_h instead of u_0 and

$$w_h = w_0 + 2x_0(\phi_0 - \phi_h) = w_0 - 2\tilde{x}_0^* \Phi(u_h - u_0) + o\left(\sqrt{h}\right)$$

instead of w_0. (The discrepancy between ϕ_h, and ϕ_0^* evaluated at u_h and w_h, is of order h.) It thus is

$$\frac{\partial \phi_0^*}{\partial u_0}(u_h - u_0) + \frac{\partial \phi_0^*}{\partial w_0}[-2\tilde{x}_0^* \Phi(u_h - u_0)].$$

The term of order $\sqrt{h}$ in the Taylor expansion of $\phi_0 - \phi_h$ has to equal the difference between the corresponding terms for ϕ_0 and ϕ_h. Therefore,

$$-\Phi = -\left[\frac{\partial \phi_0^*}{\partial u_0} + \frac{\partial \phi_0^*}{\partial w_0}(-2\tilde{x}_0^*\Phi)\right] \Rightarrow \Phi = \frac{\frac{\partial \phi_0^*}{\partial u_0}}{1 + 2\tilde{x}_0^*\frac{\partial \phi_0^*}{\partial w_0}}. \tag{40}$$

We have

$$\begin{aligned}
\frac{\partial \phi_0^*}{\partial u_0} &= \overline{\delta}\left[\frac{\partial q}{\partial u_0}\left(\frac{1}{q} - \frac{e^{-r\Theta}}{1+(q-1)e^{-r\Theta}}\right) + \frac{\partial r}{\partial u_0}\frac{(q-1)\Theta e^{-r\Theta}}{1+(q-1)e^{-r\Theta}}\right] \\
&= \overline{\delta}\left[\frac{q}{u_0}\left(\frac{1}{q} - \frac{e^{-r\Theta}}{1+(q-1)e^{-r\Theta}}\right) + \frac{r}{u_0}\frac{(q-1)\Theta e^{-r\Theta}}{1+(q-1)e^{-r\Theta}}\right] \\
&= \frac{\overline{\delta}}{u_0}\frac{1 - e^{-r\Theta} + (q-1)r\Theta e^{-r\Theta}}{1+(q-1)e^{-r\Theta}},
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial \phi_0^*}{\partial w_0} &= \overline{\delta}\frac{\partial q}{\partial w_0}\left(\frac{1}{q} - \frac{e^{-r\Theta}}{1+(q-1)e^{-r\Theta}}\right) \\
&= -\overline{\delta}\frac{q}{w_0}\left(\frac{1}{q} - \frac{e^{-r\Theta}}{1+(q-1)e^{-r\Theta}}\right) \\
&= -\frac{\overline{\delta}}{w_0}\frac{1 - e^{-r\Theta}}{1+(q-1)e^{-r\Theta}}.
\end{aligned}$$

Plugging in Eq. (40), and using the definition of s, we get Eq. (28).

To determine $\hat{\Phi}$, we use the optimality condition (13) of the i-investors. We first determine the Taylor expansion of the marginal utility of wealth, M_i, up to order $\sqrt{h}$. Recall that

$$M_i = \alpha \exp\left[-\alpha\left(w_{i,0} + k\right)\right],$$

where

$$k \equiv \sum_{t=0}^{T-1} y_t(\phi_t - \phi_{t+1}) - f(y_0 + u_0) - \sum_{t=1}^{T-1} f(y_t + u_1).$$

Therefore, if the Taylor expansion of k is

$$k = k^* + K(u_h - u_0) + o\left(\sqrt{h}\right),$$

then the Taylor expansion of M_i is

$$M_i = \alpha \exp\left[-\alpha\left(w_{i,0} + k^*\right)\right]\left[1 - \alpha K(u_h - u_0)\right] + o\left(\sqrt{h}\right). \tag{41}$$

To determine k^*, we note that it is equal to the limit of

$$k_1 \equiv \sum_{t=1}^{T-1} y_t(\phi_t - \phi_{t+1}) - \sum_{t=1}^{T-1} f(y_t + u_1)$$

when h goes to zero. Using Lemmas 3 and 4, and Eqs. (23) and (24), we get

$$\begin{aligned}
k^* &= \int_0^\Theta (-\mu x^*_\theta)\left(-\frac{d\phi^*_\theta}{d\theta}\right)d\theta - \int_0^\Theta \frac{1}{2}\alpha\hat{\delta}^2(u_0-\mu x^*_\theta)^2 d\theta \\
&= -\int_0^\Theta \frac{1}{2}\alpha\hat{\delta}^2(u_0-\mu x^*_\theta)(u_0+\mu x^*_\theta)d\theta \\
&= \int_0^\Theta \frac{1}{2}\left(u_0\frac{d\phi^*_\theta}{d\theta} - \mu\overline{\delta}\frac{dx^*_\theta}{d\theta}\right)d\theta \\
&= -\frac{1}{2}u_0\phi^*_0 - \frac{1}{2}\mu\overline{\delta}(x^*_\Theta - x^*_0) \\
&= -\frac{1}{2}u_0\phi^*_0 - \frac{\mu w_0}{4}\left(e^{\frac{\phi^*_0}{\overline{\delta}}} - 1\right).
\end{aligned} \tag{42}$$

To determine K, we note that the term of order $\sqrt{h}$ in

$$y_0(\phi_0 - \phi_1) - f(y_0 + u_0)$$

is

$$(-\mu\tilde{x}^*_0)(-\Phi)(u_h - u_0).$$

Moreover, the term of order $\sqrt{h}$ in k_1 is the same as that obtained by evaluating Eq. (42) at u_h instead of u_0, w_h instead of w_0, and

$$\phi_h = \phi^*_0 + \Phi(u_h - u_0) + o\left(\sqrt{h}\right)$$

instead of ϕ^*_0. It thus is

$$\frac{\partial k^*}{\partial u_0}(u_h - u_0) + \frac{\partial k^*}{\partial w_0}[-2\tilde{x}^*_0\Phi(u_h - u_0)] + \frac{\partial k^*}{\partial \phi^*_0}\Phi(u_h - u_0).$$

Therefore,

$$K = \mu\tilde{x}^*_0\Phi + \frac{\partial k^*}{\partial u_0} - 2\tilde{x}^*_0\Phi\frac{\partial k^*}{\partial w_0} + \Phi\frac{\partial k^*}{\partial \phi^*_0}.$$

Plugging in this equation the partial derivatives of k^*, and using the definition of s, we get

$$K = -\frac{1}{2}\left[\phi^*_0 + \Phi u_0\left(\left(1 - \frac{s}{q}\right) + e^{\frac{\phi^*_0}{\overline{\delta}}}\frac{1}{q}(1-s)\right)\right].$$

Consider now the optimality condition (13) of the A-investors. We can write this equation, in order h, as

$$E\left[\left(-\Phi(u_h - u_0) + \hat{\Phi}h - \alpha\hat{\delta}^2(u_0 - \mu\tilde{x}^*_0)h\right)(1 - \alpha K(u_h - u_0))\right] = 0.$$

Solving for $\hat{\Phi}$, we get

$$\hat{\Phi} = r\overline{\delta}\left(1 - \frac{s}{q}\right) - \alpha\Phi K\hat{u}^2, \tag{43}$$

i.e., Eq. (29). Q.E.D.

Proof to Lemma 6: Proceeding as in the proof of Lemma 5, the term of order $\sqrt{h}$ in x_θ is

$$\frac{\partial x^*_\theta}{\partial u_0}(u_h - u_0) + \frac{\partial x^*_\theta}{\partial w_0}[-2\tilde{x}^*_0 \Phi(u_h - u_0)].$$

Therefore,

$$X_\theta = \frac{\partial x^*_\theta}{\partial u_0} - 2\tilde{x}^*_0 \Phi \frac{\partial x^*_\theta}{\partial w_0}. \tag{44}$$

We have

$$\begin{aligned}
\frac{\partial x^*_\theta}{\partial u_0} &= \frac{\partial}{\partial u_0}\left[\frac{1}{\mu}\frac{u_0}{1+(q-1)e^{-r\theta}}\right] \\
&= \frac{1}{\mu}\left[\frac{1}{1+(q-1)e^{-r\theta}} - \frac{\partial q}{\partial u_0}\frac{u_0 e^{-r\theta}}{(1+(q-1)e^{-r\theta})^2} + \frac{\partial r}{\partial u_0}\frac{u_0(q-1)r\theta e^{-r\theta}}{(1+(q-1)e^{-r\theta})^2}\right] \\
&= \frac{1}{\mu(1+(q-1)e^{-r\theta})^2}\left[1 - e^{-r\theta} + (q-1)r\theta e^{-r\theta}\right]
\end{aligned}$$

and

$$\begin{aligned}
\frac{\partial x^*_\theta}{\partial w_0} &= \frac{\partial}{\partial w_0}\left[\frac{1}{\mu}\frac{u_0}{1+(q-1)e^{-r\theta}}\right] \\
&= -\frac{1}{\mu}\frac{\partial q}{\partial w_0}\frac{u_0 e^{-r\theta}}{(1+(q-1)e^{-r\theta})^2} \\
&= \frac{1}{\mu(1+(q-1)e^{-r\theta})^2}\frac{q e^{-r\theta} u_0}{w_0}.
\end{aligned} \tag{45}$$

Plugging in Eq. (44), and using the definitions of s and Φ, we get Eq. (31).

The sign of X_θ is the same as that of the function

$$F(\theta) \equiv e^{r\theta} - 1 + (q-1)r\theta - sq\frac{1 - e^{-r\Theta} + (q-1)r\Theta e^{-r\Theta}}{(1-s)(1-e^{-r\Theta}) + qe^{-r\Theta}}.$$

This function is strictly increasing, and its value for $\theta = \Theta$ is

$$F(\Theta) = (1 - e^{-r\Theta} + (q-1)r\Theta e^{-r\Theta})e^{r\Theta}(1-s)(1+(q-1)e^{-r\Theta}).$$

Suppose now $s \leq 0$. Then $F(0) \geq 0$, and thus $F(\theta) > 0$ for $\theta \in (0, \Theta]$. Suppose next $s \in (0,1)$. Then $F(0) < 0$ and $F(1) > 0$. Therefore, $F(\theta) < 0$ for $\theta \in [0, \theta(s))$ and $F(\theta) > 0$ for $\theta \in (\theta(s), \Theta]$. Moreover, since $F(\theta)$ decreases in s, $\theta(s)$ increases in s. Suppose finally $s = 1$. Then $F(1) = 0$, and thus $F(\theta) < 0$ for $\theta \in [0, \Theta)$. Q.E.D.

Proof to Lemma 7: We first determine the Taylor expansion of the return, R_1, on period 1 wealth, up to order $\sqrt{h}$. Recall that

$$R_1 = \prod_{t=1}^{T-1}\frac{1}{1 - \frac{\phi_t - \phi_{t+1}}{\delta}} \Rightarrow \log(R_1) = -\sum_{t=1}^{T-1}\log\left(1 - \frac{\phi_t - \phi_{t+1}}{\delta}\right).$$

The limit of $\log(R_1)$ when h goes to zero is

$$-\int_0^T \left(\frac{d\phi_\theta^*}{d\theta}\right)\frac{1}{\bar{\bar{\delta}}}d\theta = \frac{\phi_0^*}{\bar{\bar{\delta}}}.$$

Therefore, the term of order 0 in the Taylor expansion of R_1 is $e^{\frac{\phi_0^*}{\bar{\bar{\delta}}}}$, and the term of order $\sqrt{h}$ is

$$\frac{\partial[e^{\frac{\phi_0^*}{\bar{\bar{\delta}}}}]}{\partial\phi_0^*}\Phi(u_h-u_0) = e^{\frac{\phi_0^*}{\bar{\bar{\delta}}}}\frac{\Phi}{\bar{\bar{\delta}}}(u_h-u_0).$$

We next determine the Taylor expansion of the marginal utility of wealth, M, up to order $\sqrt{h}$. Recall that

$$M = U'(w_T) = U'(w_h R_h),$$

(where we now denote R_1 by R_h, using calendar time). Therefore, the term of order 0 is $U'(w_0 e^{\frac{\phi_0^*}{\bar{\bar{\delta}}}})$, and the term of order $\sqrt{h}$ is

$$\begin{aligned} & \frac{\partial[U'(w_0 e^{\frac{\phi_0^*}{\bar{\bar{\delta}}}})]}{\partial w_0}[-2\tilde{x}_0^*\Phi(u_h-u_0)] + \frac{\partial[U'(w_0 e^{\frac{\phi_0^*}{\bar{\bar{\delta}}}})]}{\partial\phi_0^*}\Phi(u_h-u_0) \\ = \; & U''(w_0 e^{\frac{\phi_0^*}{\bar{\bar{\delta}}}})w_0 e^{\frac{\phi_0^*}{\bar{\bar{\delta}}}}\frac{\Phi}{\bar{\bar{\delta}}}(1-s)(u_h-u_0) \\ = \; & -U'(w_0 e^{\frac{\phi_0^*}{\bar{\bar{\delta}}}})\gamma\frac{\Phi}{\bar{\bar{\delta}}}(1-s)(u_h-u_0). \end{aligned}$$

Consider now the derivative of the arbitrageurs' expected utility, $E[2(\phi_0-\phi_h)R_h M]$. The sign of this derivative, in order h, is the same as that of

$$\begin{aligned} & E\left[\left(-\Phi(u_h-u_0)+\hat{\Phi}h\right)\left(1+\frac{\Phi}{\bar{\bar{\delta}}}(u_h-u_0)\right)\left(1-\gamma\frac{\Phi}{\bar{\bar{\delta}}}(1-s)(u_h-u_0)\right)\right] \\ = \; & \left[\hat{\Phi}-\frac{\Phi^2}{\bar{\bar{\delta}}}\hat{u}^2(1-\gamma(1-s))\right]h, \end{aligned}$$

i.e., that of $g(s)$. The Lemma then follows from the arbitrageurs' optimality conditions (17)-(19). Q.E.D.

Proof to Lemma 8: Consider first the i-investors, and suppose that the Taylor expansion of

$$\ell_0 \equiv \sum_{t=0}^{T-1}\frac{d(\phi_t-\phi_{t+1})}{dx_0}y_t,$$

up to order h, is

$$L_0(u_h-u_0)+\hat{L}_0 h + o(h).$$

Then, combining this Taylor expansion with that of M_i, i.e., Eq. (41), we conclude that the derivative of the i-investors' expected utility has the form given in this Lemma, with

$$\Psi_i = \alpha\exp\left[-\alpha\left(w_{i,0}+k^*\right)\right]\left[\hat{L}_0 - L_0\alpha K\hat{u}^2\right]. \tag{46}$$

To determine Ψ_i, we thus need to determine L_0 and $\hat{L}_0$. The Taylor expansion of

$$\frac{d(\phi_0 - \phi_1)}{dx_0} y_0$$

follows from Lemma 5, and is

$$\begin{aligned} & \frac{ds}{d\tilde{x}_0^*} \left[-\frac{d\Phi}{ds}(u_h - u_0) + \frac{d\hat{\Phi}}{ds} h + o(h) \right] (-\mu \tilde{x}_0^*) \\ = \; & \mu s \frac{d\Phi}{ds}(u_h - u_0) - \mu s \frac{d\hat{\Phi}}{ds} h + o(h). \end{aligned} \tag{47}$$

To determine the Taylor expansion of

$$\ell_1 \equiv \sum_{t=1}^{T-1} \frac{d(\phi_t - \phi_{t+1})}{dx_0} y_t,$$

we note that the effect of x_0 on the equilibrium from period 1 on is through the arbitrageurs' period 1 wealth. Therefore,

$$\ell_1 = \left[\sum_{t=1}^{T-1} \frac{d(\phi_t - \phi_{t+1})}{dw_1} y_t \right] \frac{dw_1}{dx_0}.$$

The Taylor expansion of

$$\frac{dw_1}{dx_0} = \frac{d}{dx_0}(w_0 + 2x_0(\phi_0 - \phi_1)) = 2(\phi_0 - \phi_1) + 2x_0 \frac{d(\phi_0 - \phi_1)}{dx_0}$$

follows from Lemma 5, and is

$$-2\left(\Phi + s\frac{d\Phi}{ds}\right)(u_h - u_0) + 2\left(\hat{\Phi} + s\frac{d\hat{\Phi}}{ds}\right) h + o(h). \tag{48}$$

The Taylor expansion of ℓ_1 will follow once we determine that of

$$\ell_w \equiv \sum_{t=1}^{T-1} \frac{d(\phi_t - \phi_{t+1})}{dw_1} y_t,$$

up to order $\sqrt{h}$. We denote the latter Taylor expansion by

$$\ell_w^* + L_w(u_h - u_0) + o\left(\sqrt{h}\right). \tag{49}$$

To determine ℓ_w^*, the limit of ℓ_w when h goes to zero, we first use Eq. (24), and get

$$\begin{aligned} \ell_w^* &= \int_0^\Theta \frac{\partial}{\partial w_0} \left(-\frac{d\phi_\theta^*}{d\theta} \right) (-\mu x_\theta^*) d\theta \\ &= \int_0^\Theta \frac{\partial}{\partial w_0} \left(\alpha \hat{\delta}^2 (u_0 - \mu x_\theta^*) \right) (-\mu x_\theta^*) d\theta \\ &= \mu^2 \alpha \hat{\delta}^2 \int_0^\Theta \frac{\partial x_\theta^*}{\partial w_0} x_\theta^* d\theta. \end{aligned}$$

We next note that Eqs. (25) and (45) imply that

$$\frac{\partial x_\theta^*}{\partial w_0} = \frac{q}{w_0 r(q-1)} \frac{dx_\theta^*}{d\theta}.$$

Therefore,

$$\begin{aligned} \ell_w^* &= \mu^2 \alpha \hat{\delta}^2 \frac{q}{w_0 r(q-1)} \frac{1}{2} \left((x_\Theta^*)^2 - (x_0^*)^2 \right) \\ &= \frac{\mu}{4} \frac{(1-e^{-r\Theta})(q+1+(q-1)e^{-r\Theta})}{(1+(q-1)e^{-r\Theta})^2}. \end{aligned}$$

To determine L_w, we note that

$$L_w = \frac{\partial \ell_w^*}{\partial u_0} - 2\tilde{x}_0^* \Phi \frac{\partial \ell_w^*}{\partial w_0}.$$

Plugging into this equation the partial derivatives of ℓ_w^*, and using the definition of s, we get

$$\begin{aligned} L_w &= \frac{\mu}{4} \frac{1}{(1+(q-1)e^{-r\Theta})^3} \frac{q}{u_0} \\ &\times \left[2r\Theta q e^{-r\Theta} + \left(1 + s\Phi \frac{u_0}{\bar{\delta}}\right) \left(1 - e^{-r\Theta}\right) \left(1 - (q+2)e^{-r\Theta} - (q-1)e^{-2r\Theta}\right) \right]. \end{aligned}$$

Combining the Taylor expansions (47)-(49), we finally get

$$L_0 = \mu s \frac{d\Phi}{ds} - 2\left(\Phi + s\frac{d\Phi}{ds}\right) \ell_w^* \tag{50}$$

and

$$\hat{L}_0 = -\mu s \frac{d\hat{\Phi}}{ds} + 2\left(\hat{\Phi} + s\frac{d\hat{\Phi}}{ds}\right) \ell_w^* - 2\left(\Phi + s\frac{d\Phi}{ds}\right) L_w \hat{u}^2. \tag{51}$$

Consider next the arbitrageurs, and suppose that the Taylor expansion of

$$\begin{aligned} n_0 &\equiv 2(\phi_0 - \phi_1) + 2\frac{d(\phi_0 - \phi_1)}{dx_0} x_0 + 2\sum_{t=1}^{T-1} \frac{d(\phi_t - \phi_{t+1})}{dx_0} x_t \frac{R_t}{R_1} \\ &= \left[2(\phi_0 - \phi_1) + 2\frac{d(\phi_0 - \phi_1)}{dx_0} x_0 \right] \left[1 + 2\sum_{t=1}^{T-1} \frac{d(\phi_t - \phi_{t+1})}{dw_1} x_t \frac{R_t}{R_1} \right], \end{aligned}$$

up to order h, is

$$N_0(u_h - u_0) + \hat{N}_0 h + o(h).$$

Then, combining this Taylor expansion with those of R_h and M, given in the proof of Lemma 7, we conclude that the derivative of the arbitrageurs' expected utility has the form given in this Lemma, with

$$\Psi = U'(w_0 e^{\frac{\phi_0^*}{\bar{\delta}}}) e^{\frac{\phi_0^*}{\bar{\delta}}} \left[\hat{N}_0 + N_0 \frac{\Phi}{\bar{\delta}} \hat{u}^2 (1 - \gamma(1-s)) \right]. \tag{52}$$

To determine Ψ, we thus need to determine N_0 and $\hat{N}_0$. These will follow once we determine the Taylor expansion of

$$n_w \equiv 2 \sum_{t=1}^{T-1} \frac{d(\phi_t - \phi_{t+1})}{dw_1} x_t \frac{R_t}{R_1},$$

up to order $\sqrt{h}$. We denote this Taylor expansion by

$$n_w^* + N_w(u_h - u_0) + o\left(\sqrt{h}\right). \tag{53}$$

Since

$$\log\left(\frac{R_t}{R_1}\right) = \sum_{s=1}^{t-1} \log\left(1 - \frac{\phi_s - \phi_{s+1}}{\bar{\delta}}\right),$$

the limit of $\log(R_t/R_1)$ when h goes to 0 (fixing a calendar time θ) is

$$\int_0^\theta \left(\frac{d\phi_\theta^*}{d\theta}\right) = \frac{\phi_\theta^* - \phi_0^*}{\bar{\delta}}.$$

Therefore,

$$n_w^* = 2\int_0^\Theta \frac{\partial}{\partial w_0}\left(-\frac{d\phi_\theta^*}{d\theta}\right) x_\theta^* e^{\frac{\phi_\theta^* - \phi_0^*}{\bar{\delta}}} d\theta.$$

Proceeding as in the derivation of ℓ_w^*, we have

$$\frac{\partial}{\partial w_0}\left(-\frac{d\phi_\theta^*}{d\theta}\right) = -\mu\alpha\hat{\delta}^2 \frac{\partial x_\theta^*}{\partial w_0} = -\mu\alpha\hat{\delta}^2 \frac{q}{w_0 r(q-1)} \frac{dx_\theta^*}{d\theta}.$$

Moreover, Eqs. (25) and (26) imply that

$$x_\theta^* e^{\frac{\phi_\theta^* - \phi_0^*}{\bar{\delta}}} = \frac{w_0}{2\bar{\delta}}.$$

Therefore,

$$n_w^* = -\mu\alpha\hat{\delta}^2 \frac{q}{w_0 r(q-1)} \frac{w_0}{2\bar{\delta}}(x_\Theta^* - x_0^*) = -\frac{1 - e^{-r\Theta}}{1 + (q-1)e^{-r\Theta}}.$$

To determine N_w, we note that

$$N_w = \frac{\partial n_w^*}{\partial u_0} - 2\tilde{x}_0^* \Phi \frac{\partial n_w^*}{\partial w_0}.$$

Plugging into this equation the partial derivatives of n_w^*, and using the definition of s, we get

$$N_w = -\frac{e^{-r\Theta}}{(1 + (q-1)e^{-r\Theta})^2} \frac{q}{u_0}\left[r\Theta - \left(1 + s\Phi\frac{u_0}{\bar{\delta}}\right)\left(1 - e^{-r\Theta}\right)\right].$$

Combining the Taylor expansions (48) and (53), we finally get

$$N_0 = -2\left(\Phi + s\frac{d\Phi}{ds}\right)(1 + n_w^*) \tag{54}$$

and

$$\hat{N}_0 = 2\left(\hat{\Phi} + s\frac{d\hat{\Phi}}{ds}\right)(1 + n_w^*) - 2\left(\Phi + s\frac{d\Phi}{ds}\right) N_w \hat{u}^2. \tag{55}$$

Q.E.D.

Proof to Proposition 4: The sign of Ψ_i is the same as that of $\hat{L}_0 - L_0 \alpha K \hat{u}^2$. Eqs. (43) and (51) imply that

$$\begin{aligned}\hat{L}_0 &= r\overline{\delta}\left[\mu s \frac{1}{q} + 2\left(1 - \frac{2s}{q}\right)\ell_w^*\right] \\ &\quad + \alpha\left[\mu s \frac{d(\Phi K)}{ds} - 2\left(\Phi K + s\frac{d(\Phi K)}{ds}\right)\ell_w^*\right]\hat{u}^2 - 2\left(\Phi + s\frac{d\Phi}{ds}\right)L_w \hat{u}^2.\end{aligned}$$

Eq. (50) implies that

$$L_0 \alpha K \hat{u}^2 = \alpha\left[\mu s \frac{d\Phi}{ds} - 2\left(\Phi + s\frac{d\Phi}{ds}\right)\ell_w^*\right] K\hat{u}^2.$$

Therefore,

$$\hat{L}_0 - L_0 \alpha K \hat{u}^2 = r\overline{\delta}\left[\mu s \frac{1}{q} + 2\left(1 - \frac{2s}{q}\right)\ell_w^*\right] + \alpha s(\mu - 2\ell_w^*)\Phi \frac{dK}{ds}\hat{u}^2 - 2\left(\Phi + s\frac{d\Phi}{ds}\right)L_w \hat{u}^2. \quad (56)$$

In the certainty case, we have $\hat{u} = 0$ and $s = 1$. Therefore,

$$\hat{L}_0 - L_0 \alpha K \hat{u}^2 = r\overline{\delta}\left[\mu \frac{1}{q} + 2\left(1 - \frac{2}{q}\right)\ell_w^*\right].$$

Plugging in for ℓ_w^*, we find that this is equal to

$$\mu r \overline{\delta}\frac{(q-1) + 2e^{-r\Theta} + (q-1)e^{-2r\Theta}}{2(1 + (q-1)e^{-r\Theta})^2} > 0.$$

The sign of Ψ is the same as that of $\hat{N}_0 + N_0(\Phi/\overline{\delta})\hat{u}^2(1 - \gamma(1-s))$. Eqs. (43), (54), and (55), imply that

$$\begin{aligned}\hat{N}_0 + N_0 \frac{\Phi}{\overline{\delta}}\hat{u}^2(1 - \gamma(1-s)) &= 2r\overline{\delta}\left(1 - \frac{2s}{q}\right)(1 + n_w^*) \\ &\quad - 2\left(\Phi + s\frac{d\Phi}{ds}\right)(1 + n_w^*)\left[\alpha K + \frac{\Phi}{\overline{\delta}}(1 - \gamma(1-s))\right]\hat{u}^2 \\ &\quad - 2\alpha s \Phi \frac{dK}{ds}(1 + n_w^*)\hat{u}^2 - 2\left(\Phi + s\frac{d\Phi}{ds}\right)N_w \hat{u}^2. \quad (57)\end{aligned}$$

In the certainty case, we have

$$\hat{N}_0 + N_0 \frac{\Phi}{\overline{\delta}}\hat{u}^2(1 - \gamma(1-s)) = 2r\overline{\delta}\left(1 - \frac{2}{q}\right)(1 + n_w^*) = 2r\overline{\delta}\left(1 - \frac{2}{q}\right)\frac{qe^{-r\Theta}}{1 + (q-1)e^{-r\Theta}}.$$

This is positive if $q > 2$, and negative if $q < 2$. Q.E.D.

Proof to Proposition 5: The sign of Ψ_i is the same as that of

$$L \equiv \frac{\hat{L}_0 - L_0 \alpha K \hat{u}^2}{\Phi \hat{u}^2}.$$

Since the financial constraint is slack, we have $g(s) = 0$. Using Eqs. (32) and (43), we can write the equation $g(s) = 0$ as

$$r\overline{\delta}\left(1 - \frac{s}{q}\right) = \Phi\hat{u}^2\left[\alpha K + \frac{\Phi}{\overline{\delta}}(1 - \gamma(1 - s))\right]. \tag{58}$$

Eq. (28) implies that

$$\Phi + s\frac{d\Phi}{ds} = \Phi\frac{1 + (q - 1)e^{-r\Theta}}{(1 - s)(1 - e^{-r\Theta}) + qe^{-r\Theta}}. \tag{59}$$

Combining Eqs. (56), (58), and (59), we get

$$\begin{aligned} L &= \frac{\mu s\frac{1}{q} + 2\left(1 - \frac{2s}{q}\right)\ell_w^*}{1 - \frac{s}{q}}\left[\alpha K + \frac{\Phi}{\overline{\delta}}(1 - \gamma(1 - s))\right] \\ &\quad + \alpha s(\mu - 2\ell_w^*)\frac{dK}{ds} - 2\frac{1 + (q - 1)e^{-r\Theta}}{(1 - s)(1 - e^{-r\Theta}) + qe^{-r\Theta}}L_w. \end{aligned} \tag{60}$$

The sign of Ψ is the same as that of

$$N \equiv \frac{\hat{N}_0 + N_0(\Phi/\overline{\delta})\hat{u}^2(1 - \gamma(1 - s))}{\Phi\hat{u}^2}.$$

Combining Eqs. (57), (58), and (59), we get

$$\begin{aligned} N &= 2\left[\frac{1 - \frac{2s}{q}}{1 - \frac{s}{q}} - \frac{1 + (q - 1)e^{-r\Theta}}{(1 - s)(1 - e^{-r\Theta}) + qe^{-r\Theta}}\right](1 + n_w^*)\left[\alpha K + \frac{\Phi}{\overline{\delta}}(1 - \gamma(1 - s))\right] \\ &\quad - 2\alpha s\frac{dK}{ds}(1 + n_w^*) - 2\frac{1 + (q - 1)e^{-r\Theta}}{(1 - s)(1 - e^{-r\Theta}) + qe^{-r\Theta}}N_w. \end{aligned} \tag{61}$$

In the rest of the proof, we will focus on the sign of L and N.

Case 1: Θ small. When Θ is close to 0, we have the following asymptotic behavior:

$$\phi_0^* = \overline{\delta}\frac{q - 1}{q}r\Theta + o(\Theta), \qquad \Phi = \frac{\overline{\delta}}{u_0}r\Theta + o(\Theta), \qquad K = -\overline{\delta}\left(1 - \frac{s}{q}\right)r\Theta + o(\Theta),$$

$$\frac{dK}{ds} = \overline{\delta}\frac{1}{q}r\Theta + o(\Theta), \qquad \ell_w^* = \frac{\mu}{2}\frac{1}{q}r\Theta + o(\Theta), \qquad L_w = o(\Theta),$$

$$n_w^* = -\frac{1}{q}r\Theta + o(\Theta), \qquad \text{and} \qquad N_w = o(\Theta).$$

Plugging into Eqs. (60) and (61), we get

$$L = \frac{\mu s\frac{1}{q}}{1 - \frac{s}{q}}\frac{1}{u_0}(1 - \gamma(1 - s))r\Theta + o(\Theta),$$

and

$$N = -\frac{2s\frac{1}{q}}{1 - \frac{s}{q}}\frac{1}{u_0}(1 - \gamma(1 - s))r\Theta + o(\Theta).$$

Therefore, for small Θ, we have $L > 0$ and $N < 0$.

Case 2: Θ large and s close to 1. When Θ goes to ∞, we have the following limits:

$$\phi_0 = \overline{\delta}\log(q), \qquad \Phi = \frac{\overline{\delta}}{u_0}\frac{1}{1-s}, \qquad K = -\frac{1}{2}\overline{\delta}\left[\log(q) + 1 + \frac{1-\frac{s}{q}}{1-s}\right],$$

$$\frac{dK}{ds} = -\frac{1}{2}\overline{\delta}\frac{q-1}{q}\frac{1}{(1-s)^2}, \qquad \ell_w^* = \frac{\mu}{4}(q+1), \qquad \text{and} \qquad L_w = \frac{\mu}{4}\frac{q}{u_0}\frac{1}{1-s}.$$

Moreover, we have the following asymptotic behavior:

$$1 + n_w^* = qe^{-r\Theta}(1 + o(1)) \qquad \text{and} \qquad N_w = -\frac{qe^{-r\Theta}}{u_0}\left[r\Theta - \frac{1}{1-s} + o(1)\right].$$

Plugging into Eqs. (60) and (61), we find that for s close to 1,

$$L = \frac{1}{(1-s)^2}\frac{\mu}{2}\frac{q}{u_0}\left[\frac{1}{2}\alpha\overline{\delta}u_0\frac{(q-1)^2}{q^2} - 1\right] + o\left[\frac{1}{(1-s)^2}\right]$$

and

$$N = \frac{2}{(1-s)^2}\frac{qe^{-r\Theta}}{u_0}\left[\alpha\overline{\delta}u_0\frac{q-1}{q} - 2 + r\Theta(1-s) + o(1)\right] + o\left[\frac{1}{(1-s)^2}\right].$$

Since the financial constraint is slack, the term in brackets in Eq. (33) is negative. For s close to 1, this term is

$$\frac{1}{1-s}\frac{1}{u_0}\left[\frac{1}{2}\alpha\overline{\delta}u_0\frac{q-1}{q} - 1\right].$$

Therefore, for s close to 1, we have $L < 0$ and $N < 0$.

Case 3: Θ large and s close to 0. Plugging into Eq. (60), we find that for $s = 0$,

$$L = \frac{\mu}{2}\frac{1}{u_0}\left[1 - (q+1)\left(\gamma + \frac{1}{2}\alpha\overline{\delta}u_0(\log(q) + 2)\right)\right].$$

This can be positive (for example, when α and γ are close to 0) or negative (for example, when $\gamma = 1$). Plugging into Eq. (61), we find that for $s = 0$,

$$N = \frac{2qe^{-r\Theta}}{u_0}(r\Theta - 1 + o(1)).$$

Therefore, for s close to 0, we have $N > 0$, while L can have either sign. Q.E.D.

References

- Admati, A., Pfleiderer, P., 1990. Direct and indirect sale of information. Econometrica 58, 901-928.
- Aiyagari, R., Gertler, M., 1999. “Overreaction” of asset prices in general equilibrium. Review of Economic Dynamics 2, 3-35.
- Allen, F., Gale, D., 1994. Financial Innovation and Risk Sharing. MIT Press, Cambridge, MA.
- Allen, F., Gale, D., 1999. Bubbles and crises. Unpublished working paper. University of Pennsylvania.
- Attari, M., Mello, A., 2001. Financially constrained arbitrage in illiquid markets. Unpublished working paper. University of Wisconsin.
- Basak, S., Croitoru, B., 2000. Equilibrium mispricing in a capital market with portfolio constraints. Review of Financial Studies 13, 715-748.
- Basak, S., Cuoco, D., 1998. An equilibrium model with restricted stock market participation. Review of Financial Studies 11, 309-341.
- Cuoco, D., 1997. Optimal consumption and equilibrium prices with portfolio constraints and stochastic income. Journal of Economic Theory 72, 33-73.
- Cvitanic, J., Karatzas, I., 1992. Convex duality in constrained portfolio optimization. Annals of Applied Probability 2, 767-818.
- Dow, J., Gorton, G., 1994. Arbitrage chains. Journal of Finance 49, 819-849.
- Duffie, D., Rahi, R., 1995. Financial market innovation and security design: an introduction. Journal of Economic Theory 65, 1-42.
- Edwards, F., 1999. Hedge funds and the collapse of Long-Term Capital Management. Journal of Economic Perspectives 13, 189-210.
- Froot, K., Scharfstein, D., Stein, J., 1993. Risk management: coordinating corporate investment and financing policies. Journal of Finance 48, 1629-1658.
- Geanakoplos, J., 2001. Liquidity, default and crashes: endogenous contracts in general equilibrium. Unpublished working paper. Yale University.

- Geanakoplos, J., Polemarchakis, H., 1986. Existence, regularity, and constrained sub-optimality of competitive allocations when the asset market is incomplete. In: Heller, W., Starr, R., Starrett, D. (Eds.), Uncertainty, Information and Communication, Essays in Honor of Kenneth J. Arrow, Vol. 3. Cambridge University Press, Cambridge, UK.

- Gromb, D., Vayanos, D., 2000. Equilibrium and welfare in markets with financially constrained arbitrageurs. Unpublished working paper. Massachusetts Institute of Technology.

- Grossman, S., Vila, J-L., 1992. Optimal dynamic trading with leverage constraints. Journal of Financial and Quantitative Analysis 27, 151-168.

- Holmström, B., Tirole, J., 2001. LAPM: a liquidity-based asset pricing model. Journal of Finance 56, 1837-1867.

- Kiyotaki, N., Moore, J., 1998. Credit cycles. Journal of Political Economy 105, 211-248.

- Ko, J., 2000. Breakup of relative-value trades. Unpublished working paper. Massachusetts Institute of Technology.

- Kogan, L., Uppal, R., 2001. Risk aversion and optimal portfolio policies in partial and general equilibrium economies. Unpublished working paper. Massachusetts Institute of Technology.

- Krishnamurthy, A., 2000. Collateral constraints and the amplification mechanism. Unpublished working paper. Northwestern University.

- Kyle, A., Xiong, W., 2001. Contagion as a wealth effect of financial intermediaries. Journal of Finance 56, 1401-1440.

- Liu, J., Longstaff, F., 2001. Losing money on arbitrages: optimal dynamic portfolio choice in markets with arbitrage opportunities. Unpublished working paper. University of California at Los Angeles.

- Loewenstein, M., Willard, G., 2001. Liquidity and arbitrage. Unpublished working paper. University of Maryland.

- Mian, A., 2002. Capital structure, financial fragility, and the oversupply of debt. Unpublished working paper. University of Chicago.
- O'Hara, M., 1995. Market Microstructure Theory. Blackwell Publishers, Cambridge, MA.
- Shleifer, A., Vishny, R., 1992. Liquidation values and debt capacity: a market equilibrium approach. Journal of Finance 47, 1367-1400.
- Shleifer, A., Vishny, R., 1997. The limits of arbitrage. Journal of Finance 52, 35-55.
- Sodini, P., 2001. Margin rules, welfare and asset prices. Unpublished working paper. University of Stockholm.
- Tuckman, B., Vila, J.-L., 1992. Arbitrage with holding costs: A utility-based approach. Journal of Finance 47, 1283-1302.
- Tuckman, B., Vila, J.-L., 1993. Holding costs and equilibrium arbitrage. Unpublished working paper. MIT.
- Xiong, W., 2001. Convergence trading with wealth effects. Journal of Financial Economics 62, 247-292.
- Yuan, K., 1999. Asymmetric price movements and borrowing constraints: a REE model of crisis, contagion, and confusion. Unpublished working paper. University of Michigan.
- Yuan, K., 2001, A nonlinear rational expectations equilibrium model for single or multi-asset securities markets. Unpublished working paper. University of Michigan.
- Zigrand, J-P., 2001. Arbitrage and endogenous market integration. Unpublished working paper. London School of Economics.