

DISCUSSION PAPER SERIES

No. 10436

THE DYNAMICS OF FINANCIALLY CONSTRAINED ARBITRAGE

Denis Gromb and Dimitri Vayanos

FINANCIAL ECONOMICS



Centre for Economic Policy Research

THE DYNAMICS OF FINANCIALLY CONSTRAINED ARBITRAGE

Denis Gromb and Dimitri Vayanos

Discussion Paper No. 10436

February 2015

Submitted 11 February 2015

Centre for Economic Policy Research
77 Bastwick Street, London EC1V 3PZ, UK
Tel: (44 20) 7183 8801
www.cepr.org

This Discussion Paper is issued under the auspices of the Centre's research programme in **FINANCIAL ECONOMICS**. Any opinions expressed here are those of the author(s) and not those of the Centre for Economic Policy Research. Research disseminated by CEPR may include views on policy, but the Centre itself takes no institutional policy positions.

The Centre for Economic Policy Research was established in 1983 as an educational charity, to promote independent analysis and public discussion of open economies and the relations among them. It is pluralist and non-partisan, bringing economic research to bear on the analysis of medium- and long-run policy questions.

These Discussion Papers often represent preliminary or incomplete work, circulated to encourage discussion and comment. Citation and use of such a paper should take account of its provisional character.

Copyright: Denis Gromb and Dimitri Vayanos

THE DYNAMICS OF FINANCIALLY CONSTRAINED ARBITRAGE[†]

Abstract

We develop a model of financially constrained arbitrage, and use it to study the dynamics of arbitrage capital, liquidity, and asset prices. Arbitrageurs exploit price discrepancies between assets traded in segmented markets, and in doing so provide liquidity to investors. A collateral constraint limits their positions as a function of capital. We show that the dynamics of arbitrage activity are self-correcting: following a shock that depletes arbitrage capital, profitability increases, and this allows capital to be gradually replenished. Spreads increase more and recover faster for more volatile trades, although arbitrageurs cut their positions in these trades the least. When arbitrage capital is more mobile across markets, liquidity in each market generally becomes less volatile, but the reverse may hold for aggregate liquidity because of mobility-induced contagion.

JEL Classification: D52, D53, G01, G11, G12, G14 and G23

Keywords: arbitrage, financial constraints, financial crises and liquidity

Denis Gromb denis.gromb@insead.edu

INSEAD and CEPR

Dimitri Vayanos d.vayanos@lse.ac.uk

London School of Economics, NBER and CEPR

[†] We thank Philippe Bacchetta, Bruno Biais, Patrick Bolton, Darrell Duffie, Vito Gala, Jennifer Huang, Henri Pagès, Anna Pavlova, Matti Suominen, as well as seminar participants in Amsterdam, Bergen, Bordeaux, the Bank of Italy, la Banque de France, BI Oslo, Bocconi University, Boston University, CEMFI Madrid, Columbia, Copenhagen, Dartmouth College, Duke, Durham University, ESC Paris, ESC Toulouse, the HEC-INSEAD-PSE workshop, HEC Lausanne, Helsinki, the ICSTE-Nova seminar in Lisbon, INSEAD, Imperial College, Institut Henri Poincaré, LSE, McGill, MIT, Naples, NYU, Paris School of Economics, University of Piraeus, Porto, Queen's University, Stanford, Toulouse, Université Paris Dauphine, Science Po - Paris, the joint THEMA-ESSEC seminar, Vienna and Wharton for comments. Financial support from the Paul Woolley Centre at the LSE, and a grant from the Fondation Banque de France, are gratefully acknowledged. All errors are ours.

1 Introduction

The assumption of frictionless arbitrage is central to finance theory and all of its practical applications. It is hard to reconcile, however, with the large body of evidence on asset-market “anomalies,” especially those concerning the price discrepancies between assets with almost identical payoffs. One approach to address the anomalies has been to abandon the assumption of frictionless arbitrage and study the constraints faced by real-world arbitrageurs, e.g., hedge funds or trading desks in investment banks. Arbitrageurs often have limited external capital, and there is growing evidence that this constrains their activity and ultimately affects market liquidity and asset prices. These effects arise both during crises and more tranquil times, and in a variety of markets ranging from individual stocks all the way to currencies.¹

In this paper we develop a model of financially constrained arbitrage, and use it to study the dynamics of arbitrage capital, liquidity, and asset prices. These dynamics involve a two-way feedback. On one hand, arbitrageurs’ capital affects their investment capacity and hence asset prices. On the other hand, because arbitrageurs trade in asset markets, asset prices determine their trading profits and losses, which in turn drive the evolution of arbitrage capital.

We show that the dynamics of arbitrage activity are self-correcting. For example, following a shock that depletes arbitrage capital, arbitrage activity becomes more profitable, and this allows capital to be replenished and converge back towards its steady-state, pre-shock value. We also determine how the self-correcting pattern manifests itself in the cross-section of arbitrage trades. More volatile trades with higher margin requirements experience a larger increase in their spreads (i.e., price discrepancies relative to frictionless arbitrage) in response to the shock, and a faster recovery. Yet, arbitrageurs cut their positions in these trades the least. Trades with longer time to convergence also experience a larger increase in their spreads in response to the shock.

We finally use our model to examine how the degree of mobility of arbitrage capital affects market stability. When capital is more mobile across markets, the liquidity that arbitrageurs provide to investors in each market generally responds less to shocks. At the same time, mobility generates contagion, as arbitrageurs cut their positions across multiple markets in response to a shock in one market. Because of contagion, the aggregate liquidity, averaged across markets, may respond more to shocks when arbitrage capital is more mobile.

¹For example, Comerton-Forde, Hendershott, Jones, Moulton, and Seasholes (2010) find that bid-ask spreads quoted by specialists in the New York Stock Exchange widen when specialists experience losses. Hameed, Kang, and Viswanathan (2010) and Nagel (2012) show that the returns to a liquidity-providing strategy that exploits short-term reversals are higher following drops in the stock market or increases in volatility, times during which specialists are more likely to be constrained. Coval and Stafford (2007) show that stocks sold by distressed mutual funds, which experience extreme outflows, perform abnormally well after the outflows occur. Jylha and Suominen (2011) find that outflows from hedge funds that perform the carry trade predict poor performance of that trade, with low interest-rate currencies appreciating and high-interest rate ones depreciating.

The specifics of our model are as follows. We assume a discrete-time, infinite-horizon economy, with a riskless asset and a number of “arbitrage opportunities” consisting of pairs of risky assets with identical payoffs. Each risky asset is traded in a segmented market by investors who can trade only that asset and the riskless asset. Investors receive endowment shocks that affect their valuation for the risky asset in their market. Because of these shocks, the prices of the two assets in a pair can differ. The assumption that the two assets in each pair have identical payoffs is meant to capture situations where assets have closely related payoffs but can trade at significantly different prices. Market segmentation could arise because of informational asymmetries or regulation.²

An additional set of agents, arbitrageurs, seek to exploit the price discrepancies between the assets in each pair. In doing so, they intermediate trade across investors and provide liquidity to them. Arbitrageurs are “special” in that they can trade across segmented markets and thus have better opportunities than other investors. We term the price discrepancies that they seek to exploit “arbitrage spreads” and use them as an inverse measure of liquidity: arbitrageurs provide perfect liquidity if spreads are zero.

Arbitrageurs are constrained in their access to external capital. We derive their financial constraint following the logic of market segmentation and assuming that they can walk away from their liabilities unless these are backed by collateral. Consider an arbitrageur wishing to buy an asset and short the other asset in its pair. The arbitrageur could borrow the cash required to buy the asset, but the loan must be backed by collateral. Posting the asset as collateral would leave the lender exposed to a decline in its value. The arbitrageur could post as additional collateral the short position in the other asset, which offsets declines in the value of the long position. Market segmentation, however, prevents investors other than arbitrageurs from dealing in multiple risky assets. Hence, the additional collateral must come from the arbitrageur’s holdings of the riskless asset. We assume that collateral must be sufficient to fully protect the lender against default. The collateral requirement limits the positions that arbitrageurs can establish as a function of their capital. Note that positions in assets with more volatile payoffs require more collateral so that lenders are protected against larger fluctuations in asset value.

In the Appendix we derive the arbitrageurs’ financial constraint as an optimal contracting arrangement. We consider the full set of collateralized contracts that can be traded between arbitrageurs and investors in each segmented market. We show that when asset payoff distributions from one period to the next are binomial, contracts traded in collateral equilibrium (Geanakoplos

²Examples of assets with closely related payoffs that can trade at significantly different prices include “Siamese-twin” stocks, which are claims to identical dividend streams but trade in different countries (e.g., Rosenthal and Young (1990) and Dabora and Froot (1999)), “on-the-run” and “off-the-run” bonds, which have similar coupon rates and times to maturity but were issued at different times (e.g., Amihud and Mendelson (1991), Warga (1992) and Krishnamurthy (2002)), bonds and credit-default swaps, the two legs of covered interest arbitrage strategies in the currency market, etc. In the case of Siamese-twin stocks, for example, the investors in our model could be interpreted as domestic-equity mutual funds, which have a regulatory mandate to invest only in domestic stocks.

(1997, 2003)) take our assumed form.

In the absence of the financial constraint, arbitrageurs would drive all spreads down to zero and provide perfect liquidity to investors. Given the constraint, however, spreads may remain positive, and the arbitrageurs' optimal policy is to invest in the opportunities that offer maximum return per unit of collateral. Equilibrium is characterized by a cutoff return per unit of collateral: arbitrageurs invest in the opportunities above the cutoff, driving their return down to the cutoff, and do not invest in the opportunities below the cutoff. The cutoff return represents the profitability of arbitrage activity: it is a riskless per-period return that arbitrageurs earn above and beyond the riskless rate. Profitability is inversely related to arbitrage capital. When, for example, capital increases, arbitrageurs become less constrained and can hold larger positions. This drives down the returns of the opportunities they invest in, hence lowering profitability.

The self-correcting dynamics follow from the inverse relationship between profitability and capital. Following a (unanticipated) shock that depletes capital, arbitrageurs are forced to scale down their positions, and profitability increases. Because of the higher profits, the capital of arbitrageurs gradually increases. This, in turn, causes profitability to decrease, slowing down further capital accumulation. Capital converges towards a steady-state value, identical to that before the shock. In steady state, arbitrage remains profitable enough to offset the natural depletion of capital due to arbitrageurs' consumption.

Next, we examine how the self-correcting pattern manifests itself in the cross-section of arbitrage opportunities. Opportunities constituted by assets with more volatile payoffs offer higher returns to arbitrageurs because they require more collateral. For the same reason, their returns are more sensitive to movements in the aggregate return per unit of collateral. Therefore, they increase more following negative shocks to capital, and recover faster. The same applies to spreads, which are present values of future returns. At the same time, arbitrageurs cut their positions in more volatile opportunities the least because investors' demand functions for the corresponding assets are less price-elastic: high volatility makes investors more reluctant to give up the insurance they receive from arbitrageurs. Opportunities for which endowment shocks have longer duration, and hence price discrepancies take longer to disappear, have larger spreads as well. This is because spreads are present values of future returns and the summation includes more terms. For the same reason, the spreads of these opportunities increase more following negative shocks to capital.

Finally, we use our framework to examine whether markets are more stable when arbitrage capital can move more freely across them. We compare the case of integration of arbitrage markets, where all arbitrageurs can trade all assets as in our baseline model, to that of segmentation, where any given arbitrageur can trade only one given opportunity.

Integration attenuates the effect that shocks in one arbitrage market have on that market's liquidity. This is because by moving across markets, arbitrageurs bring in more investors to absorb the shocks. At the same time, liquidity becomes affected by shocks to other markets—a contagion effect. When arbitrage opportunities are symmetric in terms of their characteristics, e.g., the volatility of payoffs and the size and duration of endowment shocks, the contagion effect is dominated, and liquidity in each market is less volatile under integration. When opportunities are sufficiently asymmetric, however, the contagion effect can dominate for markets in which endowment shocks, and hence arbitrageur positions, are small. Integration exposes small markets to larger shocks from other markets.

The effect of integration on the aggregate liquidity, averaged across markets, is more complicated. Indeed, while liquidity can become less volatile in individual markets, it becomes more correlated across markets. When arbitrage opportunities are symmetric, the two effects exactly cancel out, and aggregate liquidity is equally volatile under integration and segmentation. Instead, when opportunities are asymmetric, aggregate liquidity can be more volatile under segmentation. This is because in markets where endowment shocks are large: (i) shocks to asset payoffs have larger effects on arbitrage capital holding prices constant and (ii) price changes amplify these effects with a larger multiplier. Integration eliminates this convexity because the multiplier is equalized across markets. The result can, however, reverse, if markets with large endowment shocks are also those where investor demand functions are less price-elastic.

Our paper belongs to a growing theoretical literature on the limits of arbitrage, and more precisely to its strand emphasizing arbitrageurs' financial constraints.³ Shleifer and Vishny (1997) are the first to derive the two-way relationship between arbitrage capital and asset prices. Gromb and Vayanos (2002) introduce some of our model's building blocks: arbitrageurs intermediate trade across segmented markets, and are subject to a collateral-based financial constraint. They assume, however, a finite horizon and no intermediate consumption, which rule out a steady state and the related analysis of self-correcting dynamics. They also assume a single arbitrage opportunity, which rules out cross-sectional effects.

Our result that arbitrage opportunities with higher collateral requirements offer higher returns is related to a number of papers. In Geanakoplos (2003), Brumm, Grill, Kubler, and Schmedders (2011), and Garleanu and Pedersen (2011), there are multiple risky assets differing in their collateral value, i.e., the amount that can be borrowed using the asset as collateral. Assets for which this amount is low are cheaper and offer higher expected returns. In these papers, however, there is no explicit intermediation by arbitrageurs because all agents can trade the same assets.⁴ The unique

³For a survey of this literature, see Gromb and Vayanos (2010).

⁴Detemple and Murthy (1997) and Basak and Croitoru (2000, 2006) derive related results for more general portfolio

ability of arbitrageurs to intermediate trade across investors is instead key to our model. In Brunnermeier and Pedersen (2009), collateral-constrained arbitrageurs engage in intermediation, and invest in opportunities with highest return per unit of collateral. Since more volatile opportunities require more collateral, they offer higher returns, and their returns are more sensitive to changes in arbitrage capital. The analysis is mostly static, however, and cannot address how arbitrage capital and liquidity recover after shocks, or how the mobility of arbitrage capital affects market stability.

Our results on self-correcting dynamics of arbitrage activity around a steady state are related to some recent papers. In Duffie and Strulovici (2012) arbitrageurs can supply insurance in one of two independent markets, and their movement across markets is hindered by a search friction. Losses in one market deplete arbitrage capital, causing insurance premia to rise and new capital to enter. The self-correcting dynamics in our model arise instead because the return on existing capital increases. This effect is not present in Duffie and Strulovici because arbitrageurs do not reinvest the premia they earn, and hence their capital does not grow faster following losses. In He and Krishnamurthy (2013), arbitrageurs can raise capital from other investors to invest in a risky asset, but this capital cannot exceed a fixed multiple of their internal capital. Capital recovers from negative shocks through increased profitability, as in our model. That paper focuses on the case of a single risky asset, while we focus instead on how the dynamics of liquidity and returns manifest themselves in the cross-section.⁵

Our analysis of optimal contracts in a binomial setting generalizes the no-default result of Fostel and Geanakoplos (2013a), shown under the assumption that contracts extend over one period.⁶ Besides allowing for dynamic contracts, we allow a contract to serve as collateral for other contracts, in a recursive manner. A similar recursive construction is in Gottardi and Kubler (2014).

Finally, our analysis of integration versus segmentation relates to Wagner (2011), who shows that investors choose not to hold the same diversified portfolio because this exposes them to the risk that they all liquidate at the same time, and to Guembel and Sussman (2015), who show that segmentation generally raises volatility and reduces investor welfare. In our model, arbitrageurs generally benefit from being diversified, but the resulting contagion effects can make aggregate liquidity more volatile. Contagion effects resulting from changes in arbitrageur capital or portfolio constraints are also derived in, e.g., Kyle and Xiong (2001) and Pavlova and Rigobon (2008).

The rest of the paper is organized as follows. Section 2 presents the model. Section 3 derives

constraints.

⁵Capital recovery from negative shocks through increased profitability also arises in Xiong (2001), where arbitrageurs can share risk with long-term traders and noise traders, in Brunnermeier and Sannikov (2014), where arbitrageurs are more efficient holders of productive capital, and in Kondor and Vayanos (2014), where arbitrageurs can trade with hedgers. The first two papers focus on the case of a single risky asset. Moreover, in all three papers there is no explicit intermediation by arbitrageurs because all agents can trade the same assets.

⁶Simsek (2013) derives a general characterization of default rates in collateral equilibrium, in a static setting. For more references on leverage and collateral equilibrium, see the survey by Fostel and Geanakoplos (2013b).

the equilibrium and its key properties. Section 4 specializes the model to a stationary version, and derives the steady state and the convergence dynamics. Section 5 examines how markets recover from shocks to arbitrage capital and whether they are more stable when capital is more mobile. Section 6 concludes, and proofs are in the Appendix.

2 The Model

2.1 Assets

There is an infinite number of discrete periods indexed by $t \in \mathbb{N}$. There is one riskless asset with exogenous return $r > 0$. There is also a continuum \mathcal{I} of infinitely lived risky assets, all in zero net supply. Risky assets come in pairs, with assets in each pair having identical payoffs. We denote by $-i$ the other asset in i 's pair. Assets i and $-i$ pay off

$$d_{i,t} \equiv \bar{d}_i + \epsilon_{i,t} \tag{1}$$

in period t , where \bar{d}_i is a positive constant, and $\epsilon_{i,t}$ is a random variable distributed symmetrically around zero in an interval $[-\bar{\epsilon}_i, \bar{\epsilon}_i]$. We assume $\bar{d}_i \geq \bar{\epsilon}_i$ so that asset payoffs are non-negative. The variables $\frac{\epsilon_{i,t}}{\bar{\epsilon}_i}$ are *i.i.d.* across time and asset pairs. We denote by $p_{i,t}$ the ex-dividend price of asset i in period t , and define the asset's risk premium by

$$\phi_{i,t} \equiv \frac{\bar{d}_i}{r} - p_{i,t}, \tag{2}$$

i.e., the present value of expected future payoffs discounted at the riskless rate r , minus the price.

The assumption that the two assets in each pair have identical payoffs is for simplicity. Our intention is to capture situations where two assets or portfolios have closely related payoffs but can trade at significantly different prices. Examples include Siamese-twin stocks, traded in different countries but with identical dividend streams, bonds with similar coupon rates and times to maturity, e.g., on- and off-the-run, bonds and credit-default swaps, the two legs of covered interest arbitrage strategies in the currency market, etc. The assumption that risky assets are in zero net supply is also for simplicity: it ensures that arbitrageurs hold opposite positions in the assets in each pair and hence are not affected by shocks to asset payoffs. The assumption that payoff distributions have bounded support facilitates the derivation of the arbitrageurs' financial constraint (Section 2.3.2).

2.2 Outside Investors

2.2.1 Market Segmentation

For the outside investors, the markets for the risky assets are segmented. Each outside investor can invest in only two assets: the riskless asset and one specific risky asset. We refer to the outside investors who can invest in risky asset i as i -investors. We assume that i -investors are competitive and infinitely lived, form a continuum with measure μ_i , consume in each period, and have negative exponential utility. In period t , an i -investor chooses positions $\{y_{i,s}\}_{s \geq t}$ in asset i and consumption $\{c_{i,s}\}_{s \geq t+1}$ to maximize

$$-\mathbf{E}_t \left[\sum_{s=t+1}^{\infty} \gamma^{s-t} \exp(-\alpha c_{i,s}) \right], \quad (3)$$

where α is the coefficient of absolute risk aversion and γ is the subjective discount factor. The investor is subject to a budget constraint, derived in Section 3.2. We denote the investor's wealth in period t by $w_{i,t}$. We study optimization in period t after consumption $c_{i,t}$ has been chosen, which is why we optimize over $c_{i,s}$ for $s \geq t + 1$. Accordingly, we define $w_{i,t}$ as the wealth net of $c_{i,t}$. We assume that investors i and $-i$ are identical in terms of their measure, i.e., $\mu_i = \mu_{-i}$.

We take market segmentation as given, i.e., assume that i -investors face prohibitively large transaction costs for investing in any risky asset other than asset i . These costs could be due, for example, to informational asymmetries or regulation. For example, each risky asset could be traded in a different country, and lack of information or regulatory constraints could be preventing i -investors from investing in a country other than their own.

The assumption that i -investors cannot invest in any risky asset other than asset i can be relaxed: key for our analysis is only that i -investors cannot invest in asset $-i$. For example, we could assume that investors and risky assets are divided into two groups, with assets in each pair split across groups, and investors able to invest only in assets in their group. The assumption that outside investors have negative exponential utility eliminates wealth effects for these investors. The only wealth effects in our model concern the arbitrageurs.

2.2.2 Endowment Shocks

We assume that outside investors receive random endowments, which affect their willingness to hold risky assets. In period t each i -investor receives an endowment equal to

$$u_{i,t-1}\epsilon_{i,t}, \tag{4}$$

where $u_{i,t-1}$ is known in period $t - 1$. We assume that $u_{i,t}$ is equal to zero, except maybe over a sequence of M_i periods $t \in \{h_i - M_i, \dots, h_i - 1\}$ when it can become equal to a constant u_i . The latter outcome occurs with arbitrarily small probability, and when it occurs we say that i -investors experience an endowment shock of intensity u_i and duration M_i . We assume that the probability of an endowment shock is arbitrarily small so that the possibility of a shock does not affect asset i 's price before period $h_i - M_i$.

An endowment shock in market i is accompanied by one in market $-i$. To ensure that the prices of assets i and $-i$ can differ, we assume that the shocks differ. We further restrict the shocks to be opposites, i.e., $u_i = -u_{-i}$. This assumption, together with that of zero net supply, simplifies our analysis by ensuring that arbitrageurs' positions in assets i and $-i$ are opposites.

When $u_{i,t} = 0$, investors i and $-i$ are identical and so are the prices of assets i and $-i$. Suppose instead that i -investors experience a shock $u_i > 0$. Their endowment then becomes positively correlated with the shock $\epsilon_{i,t}$ and hence with asset i 's payoff. As a consequence, asset i becomes riskier and less attractive for i -investors. Conversely, asset $-i$ becomes more attractive for $-i$ -investors, who experience a shock $u_{-i} < 0$. If investors i and $-i$ could trade with each other, then they would offset the effect of the shock and the prices of assets i and $-i$ would remain identical: each i -investor would sell u_i shares of asset i or $-i$ to $-i$ -investors. Because, however, trade between investors i and $-i$ is ruled out by market segmentation, the price of asset i decreases and of asset $-i$ increases. This creates a role for arbitrageurs, who can invest in all risky assets and in the riskless asset. Arbitrageurs buy asset i from i -investors and sell asset $-i$ to $-i$ -investors, thus exploiting the price discrepancy between the two assets. In doing so, they provide liquidity to investors i and $-i$ because they allow them to trade.

We refer to an asset pair $(i, -i)$ as an arbitrage opportunity (employing that terminology even when the two assets are trading at the same price). When investors i and $-i$ experience endowment shocks, we say that opportunity $(i, -i)$ is active. We denote by

$$\mathcal{A}_t \equiv \{i \in \mathcal{I} : u_{i,t} > 0\}$$

the set of active opportunities in period t and assume that this set is finite. A finite set of active

opportunities is consistent with a continuum of opportunities each of which becomes active with an arbitrarily small probability. For simplicity, we eliminate stochastic variation in the characteristics of active opportunities by assuming that the set

$$\mathcal{C}_t \equiv \{(\bar{\epsilon}_i, \mu_i, u_i, h_i - t) : i \in \mathcal{A}_t\}$$

is deterministic. Thus, while arbitrageurs are uncertain as to which specific opportunities will arise, they know what the profitability of their overall portfolio will be. One setting that yields a deterministic \mathcal{C}_t is as follows. The universe \mathcal{I} of risky assets is divided into N disjoint families \mathcal{I}_n for $n = 1, \dots, N$, with the assets in each family forming a continuum and having the same characteristics $(\bar{\epsilon}_i, \mu_i, u_i, M_i)$. Moreover, a deterministic number of assets from each family are randomly drawn in each period to form an active opportunity (together with the other assets in their pairs). In Section 4, we specialize this setting to the case where one asset from each family is drawn in each period. This yields a stationary version of our model, whereby the set \mathcal{C}_t is not only deterministic but also constant over time.

Under the Siamese-twin interpretation of the asset pairs, outside investors can be interpreted as those who can trade only in one country. Such investors can be, for example, domestic-equity mutual funds, which have a mandate to invest only in domestic stocks. The demand of these funds is affected by investor inflows and outflows, which correspond to our endowment shocks. Dabora and Froot (1999) find empirically that there exists a non-trivial price wedge between a stock and its Siamese twin. Moreover, the wedge increases when the aggregate stock market in the country where that stock is traded goes up. They argue that the best explanation for their findings is that some investors have mandates preventing them from investing in a country other than their own, possibly because of agency problems. This is consistent with our interpretations of market segmentation.⁷

Under the bond interpretation of asset pairs, outside investors can be interpreted as those who must hold bonds with specific coupon rates and times to maturity. Such investors can be, for example, pension funds, and their preferences could be driven by asset-liability management or tax considerations.

⁷Because i -investors cannot invest in any risky asset other than asset i , our model cannot generate the finding of Dabora and Froot (1999) that the price wedge between a stock and its Siamese twin increases when the aggregate stock market in the country where that stock is traded goes up. But this finding can be generated in the extension of our model where outside investors can invest in groups of risky assets.

2.3 Arbitrageurs

2.3.1 Specialness

Arbitrageurs can invest in all risky assets and in the riskless asset, and so have better investment opportunities than outside investors. This assumption captures in the context of our model the idea that arbitrageurs are more sophisticated than other investors. Arbitrageurs can be interpreted, for example, as hedge funds since these are less subject to the informational or regulatory frictions that cause segmentation. Because arbitrageurs can invest in all assets, they are the only agents who can exploit price discrepancies across asset pairs and provide liquidity to outside investors. In this sense, arbitrageurs in our model are “special.”

We assume that arbitrageurs are competitive and infinitely lived, form a continuum with measure one, consume in each period, and have logarithmic utility.⁸ In period t , an arbitrageur chooses positions $\{x_{i,s}\}_{i \in \mathcal{I}, s \geq t}$ in all risky assets and consumption $\{c_s\}_{s \geq t+1}$ to maximize

$$\mathbb{E}_t \left[\sum_{s=t+1}^{\infty} \beta^{s-t} \log(c_s) \right], \quad (5)$$

where β is the subjective discount factor. The arbitrageur is subject to a budget constraint and a financial constraint, derived in Sections 3.3 and 2.3.2, respectively. We denote the arbitrageur’s wealth in period t by W_t and assume that $W_0 > 0$. Since arbitrageurs have measure one, W_t is also their aggregate wealth. Logarithmic utility of arbitrageurs simplifies our analysis because it ensures that their consumption is a constant fraction of their wealth. We use the terms “wealth” and “capital” interchangeably for arbitrageurs.

2.3.2 Financial Constraints

Financial constraints arise in our model because agents need to collateralize their asset positions. Consider an agent who wants to establish a long position in a risky asset. If the agent needs to borrow cash to buy the asset, then he must post collateral to ensure that the cash loan will be repaid. Consider next an agent who wants to establish a short position in a risky asset. The agent needs to borrow the asset so that he can sell it subsequently, and must post collateral to ensure that the asset loan will be repaid. We assume that i -investors have enough wealth to collateralize any

⁸By fixing the measure of arbitrageurs, we are ruling out entry and are focusing on changes in the wealth of existing arbitrageurs as the driver of price dynamics. Duffie and Strulovici (2012) study how entry impeded by search frictions affects price dynamics. Their analysis provides a complementary perspective to ours. Note that the duration M_i of endowment shocks can be interpreted as the time it takes for enough new arbitrageurs to enter the market for arbitrage opportunity $(i, -i)$ and eliminate that opportunity.

position they may want to establish, i.e., up to $\mu_i u_i$. Arbitrageurs, however, may be constrained by their wealth.⁹

Standard asset pricing models assume that agents can establish any combination of asset positions as long as they can honor any liabilities that their positions generate. One interpretation of this constraint is that a central clearinghouse registers all positions and prevents agents from undertaking liabilities that they cannot honor.

If arbitrageurs in our model were subject to the standard constraint only, they would be able to enforce the law of one price, i.e., the prices of the two assets in each pair would be identical. Indeed, if the prices were different, arbitrageurs could sell short the more expensive asset and use part of the proceeds to buy an equal number of shares of the cheaper asset. Because asset payoffs are identical, the liabilities from the short position would be offset by the long position. Therefore, arbitrageurs would be able to honor all their liabilities, and could earn an unlimited profit by scaling up their positions.

We assume that arbitrageurs are subject to a stronger constraint than in standard models. We require them to honor any liabilities that their positions generate, and do so market-by-market. The positions of arbitrageurs in market i consist of a position in asset i and a position in the riskless asset held within that market. We require that this combined position does not generate any liability. Thus, liability is calculated market by-market rather than by aggregating across all markets as in standard models. This is in the spirit of market segmentation: the same informational or regulatory frictions that prevent i -investors for investing in any risky asset other than asset i could also be preventing arbitrageurs' lenders in market i from accepting risky assets other than asset i as collateral.¹⁰

To derive the financial constraint of an arbitrageur, we denote by $x_{i,t}$ his position in asset i and by $z_{i,t}$ the value of his combined position in market i , both in period t . The quantity $z_{i,t}$ is the sum of the value $x_{i,t} p_{i,t}$ of the investment in asset i plus the value of an investment in the riskless asset held in market i . The value of the arbitrageur's combined position in market i in period $t + 1$ is

$$z_{i,t}(1+r) + x_{i,t}[d_{i,t+1} + p_{i,t+1} - (1+r)p_{i,t}] \tag{6}$$

⁹Our assumption that outside investors are unconstrained does not necessarily imply that they are wealthier than arbitrageurs because their positions could be smaller. This could be for two distinct reasons. First, the position that arbitrageurs as a group establish in asset i is the opposite to that of i -investors. Therefore, if arbitrageurs are in smaller measure than i -investors, then they hold a larger position per capita in asset i . Second, each arbitrageur can trade more risky assets than each outside investor, leading to a larger aggregate position.

¹⁰Using one asset as collateral for a position in the other is known as cross-netting. One situation where cross-netting is generally not possible is when one asset is traded over-the-counter and the other in an exchange, e.g., US bonds are traded over the counter and US bond futures in the Chicago Mercantile Exchange. For a more detailed description of the frictions that hamper cross-netting see, for example, Gromb and Vayanos (2002) and Shen, Yan, and Zhang (2014). While our analysis rules out cross-netting, it can be generalized to allow for partial cross-netting.

and must be positive. Indeed, if it were negative, the arbitrageur would have a liability in market i , from which he could walk away. Requiring (6) to be positive for all possible realizations of uncertainty in period $t + 1$ yields

$$z_{i,t} \geq \max_{\{\epsilon_{i,t+1}\}_{i \in \mathcal{I}}} \left\{ x_{i,t} \left(p_{i,t} - \frac{d_{i,t+1} + p_{i,t+1}}{1+r} \right) \right\}. \quad (7)$$

The right-hand side of (7) represents the maximum possible loss, in present-value terms, that the arbitrageur can realize in market i between periods t and $t + 1$. This loss has to be smaller than the value of the arbitrageur's combined position in market i in period t . Thus, the arbitrageur can finance a long position in asset i by borrowing cash with the asset as collateral, but must contribute enough cash of his own to cover against the most extreme price decline. Conversely, the arbitrageur can borrow and short-sell asset i using the cash proceeds as collateral for the loan, but must contribute enough cash of his own to cover against the most extreme price increase.

Aggregating (7) across markets yields the financial constraint

$$W_t \geq \sum_{i \in \mathcal{I}} \max_{\{\epsilon_{i,t+1}\}_{i \in \mathcal{I}}} \left\{ x_{i,t} \left(p_{i,t} - \frac{d_{i,t+1} + p_{i,t+1}}{1+r} \right) \right\} \quad (8)$$

since the value of the arbitrageur's positions summed across markets is his wealth W_t . If (8) is satisfied, then the arbitrageur can allocate his total investment in the riskless asset in period t across markets so that (7) is satisfied for each market. The constraint (8) requires the arbitrageur to have enough wealth to cover his maximum possible loss in each market separately.

Our formulation of the financial constraint assumes that the only assets that arbitrageurs can trade with i -investors, or can use as collateral in market i , are asset i and the riskless asset. Under a more general formulation, arbitrageurs could trade with i -investors any contracts that are contingent on future uncertainty. These contracts could be collateralized by the riskless asset, by asset i , or by any other contracts traded in market i . Moreover, contracts could extend over any number of periods. In Appendix B we formulate equilibrium in our model under general contracts. We show that without loss of generality, contracts can be assumed to be fully collateralized and hence default-free. Moreover, when the distribution of the variables $\epsilon_{i,t}$ that describe asset payoffs is binomial, contracts can be restricted to those studied in this section: only asset i and the riskless asset can be traded and used as collateral. This generalizes, within our setting, the no-default result of Fostel and Geanakoplos (2013a), shown under the assumption that contracts extend over one period. Thus, under the binomial distribution, the financial constraint (8) can be derived from optimal contracting.

The financial constraint (8) limits the arbitrageurs' positions and their ability to provide liquidity as a function of their wealth. While we derive this constraint based on collateral, we can interpret it more broadly as a friction that arbitrageurs face in raising external capital to undertake positive present-value investments. We can also interpret the arbitrageurs' wealth as their internal capital or more broadly as the external capital that they can access without frictions.

3 Equilibrium

3.1 Symmetric Equilibrium

We look for competitive equilibria that are symmetric, in the sense that risk premia and agents' positions are opposites for the two assets in each pair. Intuitively, risk premia are opposites because assets are in zero net supply and the endowment shocks of i - and $-i$ -investors are opposites. Because premia are opposites, arbitrageurs find it optimal to establish opposite positions. The positions of i - and $-i$ -investors are opposites (and hence markets can clear) because premia and endowment shocks are opposites.

Definition 1. *A competitive equilibrium consists of prices $p_{i,t}$ and positions in the risky assets $y_{i,t}$ for the i -investors and $x_{i,t}$ for the arbitrageurs, such that positions are optimal given prices and the markets for all risky assets clear:*

$$\mu_i y_{i,t} + x_{i,t} = 0. \tag{9}$$

Definition 2. *A competitive equilibrium is symmetric if for the two assets ($i, -i$) in each pair the risk premia are opposites ($\phi_{i,t} = -\phi_{-i,t}$), the positions of outside investors are opposites ($y_{i,t} = -y_{-i,t}$), and so are the positions of arbitrageurs ($x_{i,t} = -x_{-i,t}$).*

Symmetry implies that the average of the prices of the two assets in a pair is the present value of their expected future payoff discounted at the riskless rate r :

$$\frac{p_{i,t} + p_{-i,t}}{2} = \frac{\bar{d}_i}{r}.$$

Moreover, the risk premium of each asset is one-half of the difference between its price and the price of the other asset:

$$\phi_{i,t} = \frac{p_{i,t} - p_{-i,t}}{2}.$$

Since the risk premium measures the price difference between the two assets in a pair, we refer to its absolute value as “arbitrage spread.” The absolute value of the risk premium is also an inverse measure of liquidity. When the risk premium is equal to zero, the two assets trade at the same price and arbitrageurs provide perfect liquidity to outside investors. When instead the risk premium is non-zero, liquidity is imperfect.

We look for symmetric competitive equilibria in which risk premia are deterministic. Intuitively, since the arbitrageurs’ positions in assets i and $-i$ are opposites, their wealth W_t does not depend on the payoff $d_{i,t}$ of the two assets. Hence, risk premia and arbitrageurs’ positions are also independent of $d_{i,t}$. Since asset payoffs are the only source of uncertainty, risk premia are deterministic.¹¹

In the rest of this section we study optimization by outside investors and arbitrageurs in an equilibrium of the conjectured form, i.e., symmetric with deterministic risk premia. We then impose market clearing and show that such an equilibrium exists.

3.2 Outside Investors’ Optimization

The budget constraint of an i -investor is

$$w_{i,t+1} = y_{i,t}(d_{i,t+1} + p_{i,t+1}) + (1+r)(w_{i,t} - y_{i,t}p_{i,t}) + u_{i,t}\epsilon_{i,t+1} - c_{i,t+1}. \quad (10)$$

The investor holds $y_{i,t}$ shares of asset i in period t , and these shares are worth $y_{i,t}(d_{i,t+1} + p_{i,t+1})$ in period $t+1$. The investor also holds $w_{i,t} - y_{i,t}p_{i,t}$ units of the riskless asset in period t , i.e., wealth $w_{i,t}$ minus the investment $y_{i,t}p_{i,t}$ in asset i . This investment is worth $(1+r)(w_{i,t} - y_{i,t}p_{i,t})$ in period $t+1$. Finally, the random endowment $u_{i,t}\epsilon_{i,t+1}$ is added to the investor’s wealth in period $t+1$, while consumption $c_{i,t+1}$ lowers wealth.

We can simplify (10) by introducing the return per share of asset i in excess of the riskless asset. This excess return is

$$\begin{aligned} R_{i,t+1} &\equiv d_{i,t+1} + p_{i,t+1} - (1+r)p_{i,t} \\ &= (1+r)\phi_{i,t} - \phi_{i,t+1} + \epsilon_{i,t+1}, \end{aligned} \quad (11)$$

where the second step follows from (1) and (2). Since risk premia are deterministic, the quantity

$$\Phi_{i,t} \equiv (1+r)\phi_{i,t} - \phi_{i,t+1} \quad (12)$$

¹¹Asset prices are also deterministic but this is only due to the simplifying assumption that payoffs are independent across time.

is also deterministic and represents the expected excess return of asset i . Using (11) and (12), we can write (10) as

$$w_{i,t+1} = (1+r)w_{i,t} + y_{i,t}\Phi_{i,t} + (y_{i,t} + u_{i,t})\epsilon_{i,t+1} - c_{i,t+1}. \quad (13)$$

The investor's wealth in period $t+1$ is uncertain as of period t because of the payoff shock $\epsilon_{i,t+1}$. As the third term in the right-hand side of (13) indicates, the investor's exposure to $\epsilon_{i,t+1}$ is the sum of his asset position $y_{i,t}$ and endowment shock $u_{i,t}$.

We conjecture that the investor's value function in period t is

$$V_{i,t}(w_{i,t}) = -\exp(-Aw_{i,t} - F_{i,t}), \quad (14)$$

where $F_{i,t}$ is a deterministic function of i and t , and A is a constant. The value function is negative exponential in wealth because the utility function depends on consumption in the same manner.

Proposition 1. *The value function of an i -investor in period t is given by (14), where $A = r\alpha$. The investor's optimal position in asset i is given by the first-order condition*

$$\bar{\epsilon}_i f'[(y_{i,t} + u_{i,t})\bar{\epsilon}_i] = \Phi_{i,t}, \quad (15)$$

where the function $f(y)$ is defined by

$$\exp\left[\frac{\alpha A f(y)}{\alpha + A}\right] \equiv E\left[\exp\left(-\frac{\alpha A y \epsilon_{i,t}}{(\alpha + A)\bar{\epsilon}_i}\right)\right]. \quad (16)$$

The first-order condition (15) takes an intuitive form. The right-hand side, $\Phi_{i,t}$, is the expected excess return of asset i , and can be interpreted as the marginal benefit of risk-taking. The investor equates it to the marginal cost, which is the left-hand side. Since the function $f(y)$ is convex, as shown in Lemma 1, the marginal cost is increasing in the investor's risk exposure $y_{i,t} + u_{i,t}$. Equating marginal benefit to marginal cost yields a standard downward-sloping demand: the investor's position y_i in asset i is increasing in the asset's expected excess return $\Phi_{i,t}$ and is hence decreasing in the asset's price $p_{i,t}$.

Lemma 1. *The function $f(y)$ is non-negative, symmetric around the vertical axis, and strictly convex. It also satisfies $\lim_{y \rightarrow \infty} f'(y) = 1$.*

The function $\frac{\alpha A f(y)}{\alpha + A}$ is the cumulant-generating function of $-\frac{\alpha A y \epsilon_{i,t}}{(\alpha + A)\bar{\epsilon}_i}$. Cumulant-generating functions are convex. Symmetry follows because $\epsilon_{i,t}$ is distributed symmetrically around zero.

The first-order condition of $-i$ -investors yields an optimal position that is the opposite to that of i -investors. This follows from (15) and the observations that risk premia, expected excess returns, and endowment shocks are opposites for assets i and $-i$, and that $f'(y) = -f'(-y)$ as implied by Lemma 1.

3.3 Arbitrageurs' Optimization

The budget constraint of an arbitrageur is

$$W_{t+1} = \sum_{i \in \mathcal{I}} x_{i,t} (d_{i,t+1} + p_{i,t+1}) + (1+r) \left(W_t - \sum_{i \in \mathcal{I}} x_{i,t} p_{i,t} \right) - c_{t+1}. \quad (17)$$

The differences with the budget constraint (10) of an i -investor are that the arbitrageur can invest in all assets and receives no endowment. We next rewrite (17) and the financial constraint (8) using (11), (12), $\epsilon_{i,t} = \epsilon_{-i,t}$, $\Phi_{i,t} = -\Phi_{-i,t}$, and the property that $\Phi_{i,t} = 0$ for assets that are not part of active opportunities. The latter property holds in equilibrium, as we show in Section 3.4.2. The budget constraint (17) becomes

$$W_{t+1} = (1+r)W_t + \sum_{i \in \mathcal{A}_t} (x_{i,t} - x_{-i,t}) \Phi_{i,t} + \sum_{i \in \mathcal{I}} x_{i,t} \epsilon_{i,t+1} - c_{t+1}, \quad (18)$$

and the financial constraint (8) becomes

$$W_t \geq \sum_{i \in \mathcal{I}} \frac{|x_{i,t}| \bar{\epsilon}_i}{1+r} - \sum_{i \in \mathcal{A}_t} \frac{(x_{i,t} - x_{-i,t}) \Phi_{i,t}}{1+r}. \quad (19)$$

Investing in assets that are not part of active opportunities exposes arbitrageurs to risk that is not compensated in terms of expected return, i.e., adds a mean-preserving spread to the right-hand side of the budget constraint (18). Investing in those assets also tightens the financial constraint (19). Hence, the optimal investment is zero. Holding non-opposite positions in the two legs of an active opportunity, i.e., $x_{i,t} + x_{-i,t} \neq 0$ for $i \in \mathcal{A}_t$, is suboptimal for the same reasons: by setting $x_{i,t} + x_{-i,t} = 0$ and holding $x_{i,t} - x_{-i,t}$ constant, arbitrageurs can reduce their risk without affecting their expected return, and can possibly relax their financial constraint. Hence, we can simplify the budget constraint (18) to

$$W_{t+1} = (1+r)W_t + 2 \sum_{i \in \mathcal{A}_t} x_{i,t} \Phi_{i,t} - c_{t+1}, \quad (20)$$

and the financial constraint (19) to

$$W_t \geq 2 \sum_{i \in \mathcal{A}_t} \frac{|x_{i,t}| \bar{\epsilon}_i - x_{i,t} \Phi_{i,t}}{1+r}. \quad (21)$$

Eq. (20) confirms that the dynamics of arbitrageur wealth are deterministic. The per-share return of an active opportunity $(i, -i)$ is $2\Phi_{i,t}$, i.e., twice the expected excess return $\Phi_{i,t}$ of asset i . While i -investors earn $\Phi_{i,t}$ as compensation for bearing risk, arbitrageurs earn it riskfree because they can combine a position in asset i with one in asset $-i$. Thus, when $\Phi_{i,t} \neq 0$, arbitrageurs can earn a riskless return that exceeds the riskless rate r .

Using (12), we can write the per-share return of active opportunity $(i, -i)$ as

$$2[r\phi_{i,t} + (\phi_{i,t} - \phi_{i,t+1})].$$

The first term in the square bracket represents the “carry” component of the return. This is what arbitrageurs would earn if the risk premium $\phi_{i,t}$ remained constant over time. The second term represents the “convergence” component. This is the additional return that arbitrageurs earn because the risk premium converges to zero. (The risk premium becomes zero when opportunity $(i, -i)$ stops being active.)

Eq. (21) shows that the financial constraint is more severe when asset payoffs are more volatile, i.e., $\bar{\epsilon}_i$ is larger. This is because the maximum possible loss of a position is larger. Eq. (21) shows additionally that the financial constraint is less severe when expected excess returns are larger in absolute value. Suppose, for example, that $\Phi_{i,t} > 0$, in which case arbitrageurs are long asset i , as shown in Proposition 2. Then, the larger $\Phi_{i,t}$ is, i.e., the more profitable it is to invest in asset i , the smaller is the maximum possible loss of a long position in that asset, and the less severe is the constraint. Thus, outside finance is easier to raise when arbitrage opportunities are more profitable.

The arbitrageurs’ optimization problem reduces to choosing positions in assets $i \in \mathcal{A}_t$, i.e., those with positive endowment shocks. Positions in the corresponding assets $-i$ are opposites, and positions in assets that are not part of active opportunities are zero. We solve the simplified optimization problem under the assumption that assets $i \in \mathcal{A}_t$ offer non-negative expected excess returns, i.e., $\Phi_{i,t} \geq 0$ for all $i \in \mathcal{A}_t$. This property holds in equilibrium, as we show in Section 3.4.2. We conjecture that the value function of an arbitrageur in period t is

$$V_t(W_t) = B \log(W_t) + G_t, \quad (22)$$

where G_t is a deterministic function of t , and B is a constant.

Proposition 2. *The value function of an arbitrageur in period t is given by (22), where $B = \frac{\beta}{1-\beta}$.*

The arbitrageur's optimal consumption is

$$c_t = \frac{1-\beta}{\beta} W_t. \quad (23)$$

- *If all active opportunities offer a zero return, i.e., $\Phi_{i,t} = 0$ for all $i \in \mathcal{A}_t$, then the arbitrageur is indifferent between any combination of positions in these opportunities.*
- *Otherwise, he holds non-zero positions only in opportunities with the highest return per unit of volatility:*

$$i \in \operatorname{argmax}_{j \in \mathcal{A}_t} \frac{\Phi_{j,t}}{\bar{\epsilon}_j}. \quad (24)$$

For these opportunities, positions are long in assets $i \in \mathcal{A}_t$, i.e., those with positive endowment shocks. Moreover, the financial constraint (21) binds.

The arbitrageurs' optimal investment policy can be derived intuitively as follows. Substituting the optimal consumption (23) into the budget constraint (20), we can write the latter as

$$W_{t+1} = \beta \left[(1+r)W_t + 2 \sum_{i \in \mathcal{A}_t} x_{i,t} \Phi_{i,t} \right]. \quad (25)$$

Since assets $i \in \mathcal{A}_t$ offer non-negative expected excess returns, arbitrageurs do not benefit from selling them short. Therefore, we can write the financial constraint (21) as

$$W_t \geq 2 \sum_{i \in \mathcal{A}_t} \frac{x_{i,t} (\bar{\epsilon}_i - \Phi_{i,t})}{1+r}. \quad (26)$$

Maximizing W_{t+1} in (25) subject to (26) and $x_{i,t} \geq 0$ is a simple linear-programming problem. Arbitrageurs invest only in those opportunities that offer the highest return $\Phi_{i,t}$ per unit of collateral $\bar{\epsilon}_i - \Phi_{i,t}$. Moreover, when some opportunities offer a non-zero return, arbitrageurs “max out” their financial constraint (26) because they can earn a riskless return that exceeds the riskless rate r . Maximizing return per unit of collateral, $\frac{\Phi_{i,t}}{\bar{\epsilon}_i - \Phi_{i,t}}$, is equivalent to maximizing return per unit of volatility, $\frac{\Phi_{i,t}}{\bar{\epsilon}_i}$, and we focus on the latter from now on.

3.4 Equilibrium

3.4.1 Arbitraging Arbitrage

Combining the arbitrageurs' optimal investment policy with that of outside investors, and imposing market clearing, we can derive a sharp characterization of equilibrium returns and positions. We denote by

$$\mathcal{T}_t \equiv \{i \in \mathcal{A}_t : x_{i,t} > 0\},$$

the set of active opportunities that arbitrageurs trade in period t , i.e., those in which they hold non-zero positions.

Proposition 3. *There exists $\Pi_t \in [0, 1)$ such that in period t :*

- *Arbitrageurs trade only active opportunities $(i, -i)$ such that $f'(u_i \bar{\epsilon}_i) > \Pi_t$. That is,*

$$\mathcal{T}_t = \{i \in \mathcal{A}_t : f'(u_i \bar{\epsilon}_i) > \Pi_t\}.$$

- *All active opportunities traded by arbitrageurs offer the same return Π_t per unit of volatility, while active opportunities $(i, -i)$ not traded by arbitrageurs offer return $f'(u_i \bar{\epsilon}_i) \in (0, \Pi_t]$ per unit of volatility. That is,*

$$i \in \mathcal{T}_t \Rightarrow \frac{\Phi_{i,t}}{\bar{\epsilon}_i} = \Pi_t,$$

$$i \in \mathcal{A}_t / \mathcal{T}_t \Rightarrow \frac{\Phi_{i,t}}{\bar{\epsilon}_i} = f'(u_i \bar{\epsilon}_i) \in (0, \Pi_t].$$

Proposition 3 implies that active opportunities can be ranked according to $f'(u_i \bar{\epsilon}_i)$. As can be seen by setting $y_{i,t} = 0$ in the outside investors' first-order condition (15), $f'(u_i \bar{\epsilon}_i)$ is the return per unit of volatility that opportunity $(i, -i)$ would offer in the absence of arbitrageurs. Arbitrageurs trade the opportunities for which $f'(u_i \bar{\epsilon}_i)$ is above a cutoff $\Pi_t \in (0, 1)$. Their activity causes the return per unit of volatility offered by these opportunities to decrease to the common cutoff. Opportunities for which $f'(u_i \bar{\epsilon}_i)$ is below that cutoff are not traded, and their return per unit of volatility remains equal to $f'(u_i \bar{\epsilon}_i)$.

Since the function $f(y)$ is convex, $f'(u_i \bar{\epsilon}_i)$ is increasing in the endowment shock u_i and in the parameter $\bar{\epsilon}_i$ that characterizes the volatility of asset payoffs. Thus, arbitrageurs are more likely to trade opportunities with higher volatility and higher endowment shocks: these are the opportunities offering higher return per unit of volatility in the arbitrageurs' absence.

The equalization of returns across traded opportunities can be interpreted as “arbitraging arbitrage.” If a traded opportunity offered lower return per unit of volatility than another opportunity, then arbitrageurs could raise their profit by redeploying their scarce capital to the latter. The arbitraging-arbitrage result is at the basis of the cross-sectional implications and contagion effects derived in Section 5. Suppose, for example, that arbitrageurs experience losses in opportunity $(i, -i)$. This forces them to scale back their position in that opportunity, causing its return to increase. Arbitraging arbitrage induces them, in turn, to redeploy capital to that opportunity and away from others, causing the return of others to increase as well.

3.4.2 Dynamics of Arbitrage Capital

Using Proposition 3, we can determine the dynamics of arbitrageur wealth and the relationship between wealth and Π_t .

Proposition 4. *Arbitrageur wealth evolves according to*

$$W_{t+1} = \beta \frac{1+r}{1-\Pi_t} W_t. \quad (27)$$

- If $W_t > W_{c,t} \equiv \frac{2}{1+r} \sum_{i \in \mathcal{A}_t} \mu_i u_i \bar{\epsilon}_i$, then the financial constraint is slack, arbitrageurs earn the riskless rate r , all active opportunities are traded, and their return Π_t per unit of volatility is equal to zero.
- If $W_t < W_{c,t}$, then the financial constraint binds, and arbitrageurs earn a riskless return that exceeds the riskless rate r . The return Π_t per unit of volatility offered by all traded opportunities is the unique positive solution of

$$2 \frac{1-\Pi_t}{1+r} \sum_{i \in \mathcal{T}_t} \mu_i [u_i \bar{\epsilon}_i - (f')^{-1}(\Pi_t)] = W_t, \quad (28)$$

and decreases in W_t .

When the variables $\epsilon_{i,t}$ have a binomial distribution, Π_t is a convex function of W_t .

The financial constraint is slack when all active opportunities offer a zero return, i.e., $\Phi_{i,t} = 0$ for all $i \in \mathcal{A}_t$. This happens when arbitrageurs fully absorb the endowment shocks of outside investors, i.e., $x_{i,t} = \mu_i u_i$ for all $i \in \mathcal{A}_t$. Setting $\Phi_{i,t} = 0$ and $x_{i,t} = \mu_i u_i$ in (26), we find that W_t must exceed the threshold $W_{c,t}$ defined in Proposition 4. Since all active opportunities offer a zero return, $\Pi_t = 0$ and arbitrageurs earn the riskless rate r .

When instead $W_t < W_{c,t}$, arbitrageurs cannot fully absorb the endowment shocks of outside investors. Therefore, all active opportunities offer a positive return, the return Π_t per unit of volatility offered by all traded opportunities is also positive, and arbitrageurs earn a riskless return that exceeds r . Moreover, when W_t decreases, Π_t increases because arbitrageurs are less able to absorb the endowment shocks.

Convexity of Π_t means that a given drop in W_t causes a larger increase in Π_t when it occurs in a region where W_t is smaller. Clearly, this comparison holds between the constrained and the unconstrained regions: a drop in W_t raises Π_t when $W_t < W_{c,t}$, but has no effect on Π_t when $W_t > W_{c,t}$. The intuition why the comparison can also hold within the constrained region is as follows. When W_t is smaller than but close to $W_{c,t}$, all active opportunities are traded, and hence a drop in W_t causes arbitrageurs to reduce their positions in all of them. Since the effect is spread out across many opportunities, the reduction in each position is small, and so is the increase in Π_t . When instead W_t is close to zero, arbitrageurs concentrate their investment on a small number of opportunities, and a drop in W_t triggers a large reduction in each position. Proposition 4 confirms the convexity of Π_t under the sufficient condition that the variables $\epsilon_{i,t}$ that describe asset payoffs have a binomial distribution.

Proposition 5. *A symmetric equilibrium exists in which risk premia $\phi_{i,t}$, outside investors' positions $y_{i,t}$, and arbitrageurs' positions $x_{i,t}$ and wealth W_t are deterministic. In this equilibrium, expected excess returns $\Phi_{i,t}$ are zero for assets that are not part of active opportunities, and are non-negative for assets with positive endowment shocks.*

4 Steady State and Convergence Dynamics

We next analyze a stationary version of our model. Stationarity allows us to characterize more fully the dynamics of risk premia and arbitrageur wealth. We build on this characterization to derive implications of our model in Section 5.

We assume that the universe \mathcal{I} of risky assets can be divided into $2N$ disjoint families, with the assets in each family forming a continuum and having the same characteristics $(\bar{\epsilon}_i, \mu_i, u_i, M_i)$. Moreover, one asset from each family is randomly drawn in each period to form an active opportunity (together with the other asset in its pair). Under these assumptions, the set \mathcal{C}_t describing the characteristics of active opportunities is constant over time, and our model becomes stationary. We show that the stationary version of our model has a deterministic steady state, and we derive the dynamics of convergence to the steady state.

We index families by $n \in \{-N, \dots, -1, 1, \dots, N\}$, with the convention that for an asset in fam-

ily n the other asset in its pair belongs to family $-n$, and that families $n = 1, \dots, N$ comprise the assets with the positive endowment shocks. We denote by $(\bar{\epsilon}_n, \mu_n, u_n, M_n)$ the characteristics $(\bar{\epsilon}_i, \mu_i, u_i, M_i)$ for all assets i in family n . The set of active opportunities in period t is

$$\mathcal{A} = \{(n, m) : n \in \{1, \dots, N\}, m \in \{1, \dots, M_n\}\}.$$

Opportunity (n, m) consists of one asset in family $n \in \{1, \dots, N\}$ and one asset in family $-n$, and remains active for $m - 1$ more periods. We denote the former asset by (n, m) and the latter by $(-n, m)$, and refer to m as the horizon of opportunity (n, m) . The expected excess returns of assets (n, m) and $(-n, m)$ do not depend on m (Proposition 3), and neither do the arbitrageurs' and outside investors' positions (Eqs. (9) and (15)). We hence index these quantities by the family subscript, n or $-n$, and the time subscript, t . The risk premia of the two assets depend on m , and we index them by the additional subscript m . Since arbitrageurs' positions do not depend on m , we can write the set of active opportunities traded in period t as

$$\mathcal{T}_t = \{(n, m) : n \in \mathcal{N}_t, m \in \{1, \dots, M_n\}\},$$

where we denote by \mathcal{N}_t the subset of families in $\{1, \dots, N\}$ whose assets are traded. We drop the time subscript for steady-state values.

Proposition 6. *The wealth W_t of arbitrageurs and the return Π_t per unit of volatility offered by all traded opportunities converge over time monotonically to steady-state values W and Π .*

- If $\beta(1+r) > 1$, then W_t increases toward $W = \infty$ and Π_t decreases toward $\Pi = 0$.
- If $\beta(1+r) < 1 - \bar{\Pi}$, where $\bar{\Pi} \equiv \max_{n=1, \dots, N} f'(u_n \bar{\epsilon}_n) < 1$, then W_t decreases toward $W = 0$ and Π_t increases toward $\Pi = \bar{\Pi}$.
- Otherwise, the steady-state values are given by

$$W = 2 \frac{1 - \Pi}{1 + r} \sum_{n \in \mathcal{N}} \mu_n M_n [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)] \in (0, W_c), \quad (29)$$

$$\Pi = 1 - \beta(1+r) \in (0, \bar{\Pi}), \quad (30)$$

where $W_c \equiv \frac{2}{1+r} \sum_{n=1}^N \mu_n M_n u_n \bar{\epsilon}_n$. If $W_t < W$, then W_t increases toward W and Π_t decreases toward Π . If $W_t > W$, then W_t decreases toward W and Π_t increases toward Π .

The dynamics in Proposition 6 can be derived by specializing Proposition 4 to the stationary

case. According to Proposition 4, the wealth of arbitrageurs increases between periods t and $t + 1$ if $\beta \frac{1+r}{1-\Pi_t} > 1$. Intuitively, wealth increases if the return earned by arbitrageurs is large relative to their consumption. Arbitrageurs earn the riskless (net) return $\frac{1+r}{1-\Pi_t} - 1$. The effect of consumption is captured by the subjective discount factor β , which is inversely related to consumption, as shown in (23).

Using Proposition 4, we can characterize how the return $\frac{1+r}{1-\Pi_t} - 1$ earned by arbitrageurs depends on their wealth W_t . When $W_t > W_c$, all active opportunities offer a zero return, $\Pi_t = 0$, and arbitrageurs earn the riskless rate r . When instead $W_t < W_c$, Π_t is positive, and arbitrageurs earn a return that exceeds r . Decreases in W_t within that region raise Π_t and hence raise arbitrageurs' return. Arbitrageurs' return reaches its maximum value, corresponding to the maximum value of Π_t , when W_t goes to zero. Setting $y_{i,t} = 0$ in the outside investors' first-order condition (15), we find that the return per unit of volatility from an active opportunity $(i, -i)$ in the absence of arbitrageurs is equal to $f'(u_i \bar{\epsilon}_i)$. Therefore, the maximum value of Π_t is $\bar{\Pi} \equiv \max_{n=1, \dots, N} f'(u_n \bar{\epsilon}_n)$. Specializing Proposition 4 to the stationary case ensures that the function linking Π_t to W_t , and in particular the parameters W_c and $\bar{\Pi}$, are constant over time.

The dynamics of wealth in the stationary case follow from the above observations. When $\beta(1+r) > 1$, the wealth of arbitrageurs increases over time even if $\Pi_t = 0$, i.e., if all active opportunities offer a zero return. Thus, wealth becomes arbitrarily large. When instead $\beta(1+r) < 1 - \bar{\Pi}$, the wealth of arbitrageurs decreases over time even if active opportunities offer their maximum return. Thus, wealth converges to zero.

In the intermediate case $1 - \bar{\Pi} < \beta(1+r) < 1$, the wealth of arbitrageurs converges to an interior steady-state value. Indeed, when wealth is large, all active opportunities offer a zero return, and wealth decreases because $\beta(1+r) < 1$. When instead wealth is close to zero, active opportunities offer their maximum return, and wealth increases because $1 - \bar{\Pi} < \beta(1+r)$. Dynamics are self-correcting: wealth decreases when it is large because arbitrageurs earn a low return, and wealth increases when it is small because arbitrageurs earn a high return. The steady-state value W implied by these dynamics is smaller than W_c because the steady-state return earned by arbitrageurs must exceed r to offset consumption. An increase in the subjective discount factor raises consumption, and hence raises the steady-state return and lowers the steady-state wealth.

We illustrate the dynamics in the stationary case with a numerical example. We assume that periods correspond to months. We set the subjective discount factor β to $0.9^{\frac{1}{12}}$ and the riskless rate r to $(1 + 2\%)^{\frac{1}{12}} - 1$. The annualized values of these variables are 0.9 and 2%, respectively. We assume that there are $2N = 6$ families of risky assets: families 1, 2 and 3 comprise the assets

with the positive endowment shocks, and families -1, -2 and -3 comprise the other assets in their pairs. We assume that the total measure of outside investors for all assets in family n that are part of active opportunities is equal to one. This measure is the product of the measure μ_n of outside investors for any given asset in family n , times the number M_n of assets in that family that are part of active opportunities. We impose no additional restrictions on μ_n and M_n . We assume that endowment and volatility parameters are $(u_1, \bar{\epsilon}_1) = (1, 1)$, $(u_2, \bar{\epsilon}_2) = (2, 1)$ and $(u_3, \bar{\epsilon}_3) = (2, 2)$, and that the random variables $\epsilon_{i,t}$ that describe asset payoffs have a binomial distribution. We set the coefficient of absolute risk aversion of outside investors to $\alpha = 5$.

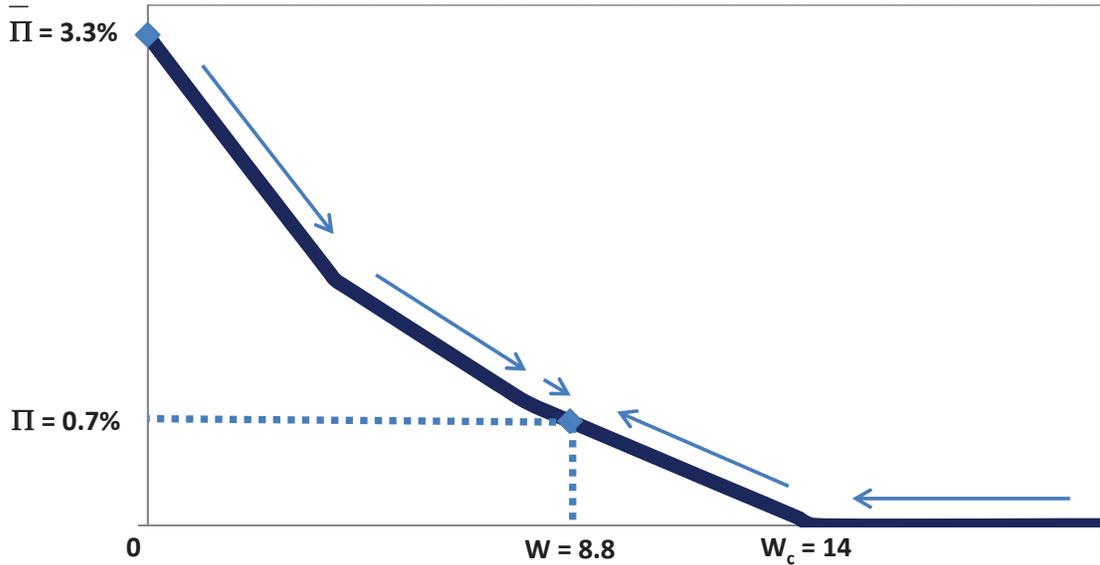


Figure 1: **Dynamics of arbitrageur wealth W_t and return Π_t per unit of volatility in the stationary case.** Wealth W_t is in the x -axis. Return Π_t per unit of volatility is in the vertical axis and is expressed in monthly terms. The point (W, Π) corresponds to the steady state. The figure is drawn for $\beta = 0.9^{\frac{1}{12}}$, $r = (1 + 2\%)^{\frac{1}{12}} - 1$, $N = 3$, $\mu_n M_n = 1$ for all n , $(u_1, \bar{\epsilon}_1) = (1, 1)$, $(u_2, \bar{\epsilon}_2) = (2, 1)$, $(u_3, \bar{\epsilon}_3) = (2, 2)$, binomial distributions for the random variables $\epsilon_{i,t}$, and $\alpha = 5$.

Figure 1 plots the return Π_t per unit of volatility as a function of arbitrageur wealth W_t . The wealth threshold W_c above which Π_t is equal to zero is $\frac{2}{1+r}(u_1 \bar{\epsilon}_1 + u_2 \bar{\epsilon}_2 + u_3 \bar{\epsilon}_3) \approx 14$. The maximum value of Π_t , which corresponds to zero arbitrageur wealth, is 3.3% in monthly terms. The steady-state value of Π_t is 0.7% in monthly terms, and the corresponding steady-state value of W_t is 8.8. The implied returns per unit of collateral are $\frac{3.3\%}{1-3.3\%} = 3.4\%$ when arbitrageur wealth is zero, and $\frac{0.5\%}{1-0.5\%} = 0.7\%$ in steady state.

Figure 1 illustrates the result of Proposition 4 that Π_t is a decreasing function of W_t . When W_t exceeds the steady-state value W , arbitrageurs earn a low return Π_t and their wealth decreases

to W , as shown by the arrows to the right of W . When instead $W_t < W$, arbitrageurs earn a high return Π_t and their wealth increases to W , as shown by the arrows to the left of W .

An additional result of Proposition 4 shown in Figure 1 is that Π_t is a convex function of W_t . A drop in W_t has no effect on Π_t in the unconstrained region $W_t > W_c$, but raises Π_t in the constrained region $W_t < W_c$. Moreover, the effect strengthens as W_t decreases within the constrained region. Within that region Π_t is approximately a piecewise linear function of W_t . The rightmost segment ($W_t \in [7.9, 14]$) corresponds to the case where all active opportunities are traded, the middle segment ($W_t \in [3.9, 7.9]$) to the case where assets in families (1,-1) are not traded, and the leftmost segment ($W_t \in [0, 3.9]$) to the case where assets in families (2,-2) are also not traded. Arbitrageurs stop trading assets in families (1,-1) the first as their wealth decreases because the product $u_n \bar{e}_n$ is the smallest for those assets and hence the corresponding arbitrage opportunities are the least profitable. As arbitrageurs concentrate their investment on a smaller number of opportunities, a given drop in their wealth W_t has a larger effect on their positions and on Π_t .

5 Implications

In this section we explore the implications of our model for two related issues. First, how do markets adjust over time following shocks to arbitrage capital? Second, are markets more stable when arbitrage capital is more mobile across opportunities? We consider shocks relative to the steady state of the stationary version of our model (Section 4). We focus on parameters for which the steady state is interior, i.e., arbitrageur wealth does not converge to zero or infinity.

5.1 Recovery From Shocks

To study how markets recover from shocks, we consider the following thought experiment. Suppose that in period t arbitrageur wealth drops below its steady-state value. This could be due, for instance, to an unanticipated shock, e.g., assets in one or several pairs failing to pay the exact same dividend. We study both the immediate and longer-term effects that the shock has on returns, spreads, liquidity (of which spreads are an inverse measure), and positions. We focus on opportunities that are traded in steady state; those that are not traded are not affected by drops in wealth.

Corollary 1. *Suppose that arbitrageur wealth drops in period t below its steady-state value.*

- *The immediate effect is that returns and spreads increase, liquidity decreases, and arbitrageurs*

scale down their positions, possibly to zero:

$$\forall (n, m) \in \mathcal{T} : \quad \Phi_{n,t} > \Phi_n, \quad \phi_{n,m,t} > \phi_{n,m}, \quad 0 \leq x_{n,t} < x_n.$$

- Following this immediate reaction, returns, spreads, liquidity and positions revert gradually toward their steady state values:

$$\begin{aligned} \forall (n, m) \in \mathcal{T} : \quad & \Phi_{n,t} \geq \Phi_{n,t+1} \geq \dots \geq \Phi_{n,t+m-1} > \Phi_n, \\ & \frac{\phi_{n,m,t}}{\phi_{n,m}} \geq \frac{\phi_{n,m-1,t+1}}{\phi_{n,m-1}} \geq \dots \geq \frac{\phi_{n,1,t+m-1}}{\phi_{n,1}} > 1, \\ & x_{n,t} \leq x_{n,t+1} \leq \dots \leq x_{n,t+m-1} < x_n. \end{aligned}$$

Proposition 6 implies that the profitability of arbitrage, as measured by the return Π_t per unit of volatility, increases immediately following the shock, and then decreases gradually over time toward its steady-state value. Corollary 1 shows that the dynamics for individual arbitrage opportunities are similar to those of Π_t : an immediate movement away from steady state, followed by gradual reversion. The reversion pattern can, however, be different from that of Π_t . Following its initial rise, Π_t decreases over time. Returns of individual arbitrage opportunities decrease over time only for those opportunities in which arbitrageurs remain invested after their initial drop in wealth. For an opportunity that arbitrageurs exit, the return remains constant until Π_t decreases to the level at which the opportunity becomes attractive again. From that time onward, the return decreases.

Changes in returns between one period and the next are not always a monotone function of time. Clearly, returns change slowly when they approach their steady-state values. But they can also change slowly when a large drop in wealth drives them far above their steady-state values. This applies not only to opportunities that arbitrageurs exit (returns are constant), but also to opportunities in which they remain invested. The intuition is that changes in returns are driven by absolute rather than relative changes in wealth, and the former are small when wealth is small. Changes in returns can hence be the most rapid in an intermediate period, i.e., can be a hump-shaped function of time. Changes in spreads can also be hump-shaped, for the same reason. In our numerical example, the hump shape arises following large reductions in wealth: wealth must drop to 3 or below from its steady-state value of 8.8.

Corollary 1 traces the evolution of the return, spread and position for opportunity (n, m) during the time when the opportunity is active, i.e., its horizon m . The steady-state value of the return and position are constant during that time, but the steady-state value of the spread decreases towards zero as horizon shortens. Corollary 1 adjusts for horizon by dividing the spread by its

time-varying steady-state value. The spread decreases over time both because it approaches its steady-state value, as shown in Corollary 1, and because that value decreases.

We next examine how the dynamics of returns, spreads, liquidity and positions depend on the characteristics of arbitrage opportunities. Specifically, we compare opportunities that differ in their volatility parameter $\bar{\epsilon}_n$ (Corollary 2) and in their horizon m (Corollary 3).

Corollary 2. *Consider two arbitrage opportunities (n, m) and (n', m) , with $\bar{\epsilon}_n > \bar{\epsilon}_{n'}$ and $(\mu_n, u_n) = (\mu_{n'}, u_{n'})$, which are among the opportunities traded in steady state.*

- *Return and spread are larger for the more volatile opportunity, and liquidity is smaller. Immediately following a drop in arbitrageur wealth in period t below the steady-state value, return and spread increase more for the more volatile opportunity, and liquidity decreases more. During the recovery phase, return and spread decrease more for that opportunity, and liquidity increases more. That is,*

$$\Phi_n > \Phi_{n'}, \quad \phi_{n,m} > \phi_{n',m},$$

$$\Phi_{n,t} - \Phi_n > \Phi_{n',t} - \Phi_{n'}, \quad \phi_{n,m,t} - \phi_{n,m} > \phi_{n',m,t} - \phi_{n',m},$$

$$\Phi_{n,t+s} - \Phi_{n,t+s+1} \geq \Phi_{n',t+s} - \Phi_{n',t+s+1},$$

$$\phi_{n,m-s,t+s} - \phi_{n,m-s-1,t+s+1} \geq \phi_{n',m-s,t+s} - \phi_{n',m-s-1,t+s+1} \quad \forall s = 0, \dots, m-2.$$

- *Arbitrageurs hold a larger position in the more volatile opportunity. If arbitrageur wealth drops in period t below the steady-state value, and the drop is not large enough for arbitrageurs to exit any of the opportunities, their position in the less volatile opportunity is scaled down by a larger amount. During the recovery phase, it is scaled up by a larger amount. That is,*

$$x_n > x_{n'}, \quad x_n - x_{n,t} < x_{n'} - x_{n',t},$$

$$x_{n,t+s+1} - x_{n,t+s} < x_{n',t+s+1} - x_{n',t+s} \quad \forall s = 0, \dots, m-2.$$

For larger drops in wealth, arbitrageurs exit the less volatile opportunity, and possibly the more volatile one as well.

The first part of Corollary 2 shows that the return and spread of the more volatile opportunity are larger and more sensitive to changes in arbitrageur wealth. If the drop in wealth is small enough so that both opportunities remain traded, the result follows from Proposition 3, which shows that arbitrageurs equalize return per unit of volatility for all traded opportunities. Indeed, if return is

proportional to volatility, then it is larger for the more volatile opportunity. Since, in addition, wealth affects the proportionality coefficient Π_t , the return of the more volatile opportunity is more sensitive to changes in wealth.

Suppose next that the drop in wealth is large enough for arbitrageurs to exit one of the opportunities. According to the second part of Corollary 2, this has to be the less volatile opportunity. Hence, its return and spread remain less sensitive to changes in wealth.

The second part of Corollary 2 shows additionally that arbitrageurs hold a larger position in the more volatile opportunity and that position is less sensitive to changes in wealth. Arbitrageurs' position in the more volatile opportunity is larger because outside investors in that opportunity are more eager to share risk. The position is less sensitive to changes in wealth because the stronger risk-sharing motive by outside investors renders their demand less price-elastic. Indeed, suppose that arbitrageurs reduce their positions equally in both opportunities following a drop in wealth. Because outside investors in the more volatile opportunity would suffer more from the reduced risk-sharing, they would value risk-sharing more in the margin. This would cause the more volatile opportunity to offer higher return per unit of volatility, and would induce arbitrageurs to re-balance towards that opportunity.

In summary, Corollary 2 shows that following a drop in wealth, the returns and spreads of the more volatile opportunities increase the most, and yet arbitrageurs may cut their positions in those opportunities the least.

Corollary 3. *Consider two arbitrage opportunities (n, m) and (n, m') , with $m > m'$, which are among the opportunities traded in steady state. The spread is larger for the opportunity with the longer horizon, and liquidity is smaller. Immediately following a drop in arbitrageur wealth in period t below the steady-state value, the spread increases more for the opportunity with the longer horizon, and liquidity decreases more. During the recovery phase, the spread decreases more slowly for that opportunity, and liquidity increases more slowly. That is,*

$$\begin{aligned} \phi_{n,m} &> \phi_{n,m'}, & \phi_{n,m,t} - \phi_{n,m} &> \phi_{n,m',t} - \phi_{n,m'}, \\ \phi_{n,m-s,t+s} - \phi_{n,m-s-1,t+s+1} &\leq \phi_{n,m'-s,t+s} - \phi_{n,m'-s,t+s+1} & \forall s = 0, \dots, m' - 2. \end{aligned}$$

Corollary 3 shows that the spread of an opportunity with longer horizon is larger and more sensitive to changes in arbitrageur wealth. The intuition can be seen from the following relationship

between spread and returns:

$$\phi_{n,m,t} = \sum_{s=0}^{m-1} \frac{\Phi_{n,t+s}}{(1+r)^{s+1}}, \quad (31)$$

which can be derived by solving (12) backwards with the terminal condition $\phi_{n,0,t} = 0$. The spread associated to an arbitrage opportunity is the present value of the opportunity's per-period returns discounted at the riskless rate. An opportunity with longer horizon has larger spread because the present value includes more terms. This spread is more sensitive to changes in wealth because all the terms depend on wealth. It decreases more slowly during the recovery phase (and, in fact, starting from any level of wealth) because it must reach zero at the end of a longer horizon.

5.2 Mobility of Arbitrage Capital

We next examine how the degree of mobility of arbitrage capital affects market stability. Are markets more stable when arbitrageurs are better diversified across opportunities? Or does diversification cause large contagion effects as arbitrageurs react to negative shocks in one market by cutting their positions across all markets? To study this issue, we compare two polar cases. Under full integration, our maintained assumption so far, all arbitrageurs can trade all assets. Instead, under full segmentation, the assets in each family pair $(n, -n)$ for $n = 1, \dots, N$ are traded by a separate set of arbitrageurs, and constitute segmented arbitrage market n . Integration could be triggered, for example, from a deregulation of international capital flows.

To study the effects of integration, we start from the steady state under segmentation, and lift the segmentation restriction. Proposition 6 applied to each segmented arbitrage market implies that arbitrageurs in market n have non-zero wealth in steady state if $f'(u_n \bar{\epsilon}_n) > 1 - \beta(1+r)$. Moreover, the return per unit of volatility is $\Pi = 1 - \beta(1+r)$ in the markets where arbitrageur wealth is non-zero, and $f'(u_n \bar{\epsilon}_n) \leq \Pi$ in the markets where it is zero. Since this return is the same across the non-zero-wealth markets, and is lower in the zero-wealth markets, lifting the segmentation restriction has no effect: arbitrageurs are indifferent between staying in their market or diversifying into other non-zero-wealth markets, and the return per unit of volatility in all markets does not change.

Corollary 4. *In steady state, integration of arbitrage markets has no effect on spreads and returns.*

While segmentation and integration are equivalent in steady state, they yield different outcomes following stochastic shocks. As in Section 5.1, we assume that shocks are unanticipated. Section 5.1 identifies shocks with a decrease in arbitrageur wealth. In this section we are more explicit

about the mechanism through which shocks affect wealth, because it differs across segmentation and integration.

We assume that shocks concern the relative payoff of the assets in each pair. In particular, in period t the payoff of asset (n, m) is not identical to that of asset $(-n, m)$, but instead exceeds it by $\bar{\epsilon}_n \eta_{nt}$, where η_{nt} is an *i.i.d.* shock across $n = 1, \dots, N$. We assume that η_{nt} is distributed in a small interval around zero, and hence shocks are small, so that we can linearize around the steady state.

The effect of shocks on arbitrageur wealth can be derived from the arbitrageurs' budget constraint. Since markets are in steady state before the shocks occur in period t , we can use (12) to write the budget constraint (25) as

$$W_t = \beta \left[(1+r)W_{t-1} + 2 \sum_{(n,m) \in \mathcal{T}} x_n (\bar{\epsilon}_n \eta_{nt} + (1+r)\phi_{n,m} - \phi_{n,m-1,t}) \right]. \quad (32)$$

If the shock η_{nt} is negative, then it reduces the wealth W_t of arbitrageurs through the term $x_n \bar{\epsilon}_n \eta_{nt}$. This is a direct effect, holding arbitrage spreads $\phi_{n,m-1,t}$ constant. The shock has also an indirect effect, through a change in spreads: because W_t decreases, spreads increase, and this amplifies the reduction in W_t . Amplification effects have been derived in the literature on limited arbitrage (see, e.g., Gromb and Vayanos (2010) for a survey). We examine how they differ across segmentation and integration.

We compare segmentation and integration based on the variance of spreads and of arbitrageur wealth in period t . We also consider the aggregate spread, defined as the average of individual spreads weighted by arbitrageurs' positions:

$$\phi_t \equiv \sum_{n \in \mathcal{N}} x_n \sum_{m=1}^{M-1} \phi_{n,m,t}.$$

The aggregate spread is the liquidity provided to a hypothetical outside investor who is either on the long or on the short side of all of the arbitrageurs' positions. Because shocks are independent and small, the variance of a random variable X_t , e.g., spreads or wealth, is

$$\text{Var}(X_t) = \sum_{n=1}^N \left(\frac{\partial X_t}{\partial \eta_{n,t}} \right)^2 \text{Var}(\eta), \quad (33)$$

where the partial derivatives are evaluated at the steady state and $\text{Var}(\eta)$ is the common vari-

ance of $\eta_{n,t}$ for $n = 1, \dots, N$. The variance of X_t is the sum over $n = 1, \dots, N$ of a variance induced by each shock. We first consider the case where arbitrage opportunities are symmetric, i.e., $(\bar{\epsilon}_n, \mu_n, u_n, M_n) \equiv (\bar{\epsilon}, \mu, u, M)$ for all $n = 1, \dots, N$.

Proposition 7. *Suppose that arbitrage opportunities are symmetric. Integration of arbitrage markets:*

- *Lowers the variance of each spread and of each arbitrageur's wealth.*
- *Does not affect the variance of the arbitrageurs' total wealth and of the aggregate spread.*

Integration of arbitrage markets makes them more stable in the sense that the spread associated to each opportunity, and hence the liquidity of the underlying assets, become less volatile. Risk-sharing among arbitrageurs also improves because the variance of each arbitrageur's wealth is reduced. At the same time, integration does not affect the variance of the arbitrageurs' total wealth and of aggregate liquidity.

The intuition why integration does not affect the variance of the arbitrageurs' total wealth is as follows. Since the aggregate position of all arbitrageurs in an opportunity (n, m) remains the same under integration, a negative shock η_{nt} causes the same drop in total wealth through the direct effect. The additional drop through the indirect effect is also the same. Indeed, under segmentation, the shock affects only market n , and arbitrageur positions are cut only for the corresponding active opportunities. Under integration, positions are cut for all active opportunities, as arbitrageurs move across markets to equate returns per unit of volatility (Proposition 3), but each position is cut by less. Aggregating across positions and arbitrageurs, the cut is the same in both cases because it is triggered by the same drop in total wealth (i.e., the same direct effect). The resulting rise in the aggregate spread is also the same because the aggregate cut is the same and because symmetry ensures that the spreads for all opportunities are equally sensitive to changes in positions. Therefore, the indirect effect is the same.

Since integration does not affect the variance of the aggregate spread but causes individual spreads to become more correlated, it also causes them to become less volatile. Intuitively, arbitrageurs smooth the effect that a shock in any given market has on spreads because by moving across markets they bring in more outside investors to absorb the shock. Because individual spreads become less volatile, each arbitrageur bears less risk: this can be seen more simply in the case where he does not diversify across markets prior to the shock. In the proof of Proposition 7 we show that integration causes the variance of individual spreads and of each arbitrageur's wealth to be divided by N , the number of arbitrage markets.

We next consider the case where arbitrage opportunities are asymmetric. For simplicity we

keep horizon symmetric, i.e., $M_n \equiv M$ for all $n = 1, \dots, N$.

Proposition 8. *Suppose that arbitrage opportunities have the same horizon parameter $M_n = M$, but differ in terms of the volatility parameter $\bar{\tau}_n$ and the endowment shock u_n . Integration of arbitrage markets:*

- *Lowers the variance of some but not necessarily all spreads.*
- *Lowers the variance of the arbitrageurs' total wealth and of the aggregate spread, if opportunities have the same measure $\mu_n = \mu$ of outside investors. Otherwise, the variance can rise.*

When arbitrage opportunities differ only in terms of the volatility parameter $\bar{\tau}_n$ and the endowment shock u_n , integration lowers the variance of the arbitrageurs' total wealth. This is because amplification effects exhibit a form of convexity: the indirect effect multiplies the direct effect, and one is larger in those segmented markets where the other is larger. Indeed, markets in which the direct effect is large are those where u_n is large, because arbitrageurs hold a large position, or where $\bar{\tau}_n$ is large, because the shock η_{nt} has a large impact on the payoff difference $\bar{\tau}_n \eta_{nt}$. In the former case the indirect effect is large because the change in spreads multiplies a large position, and in the latter case because spreads are more sensitive to changes in positions. Integration mitigates the convexity because the indirect effect becomes the same across markets, and approximately equal to the average of indirect effects under segmentation. This reduces the average effect that shocks have on total wealth.

When arbitrage opportunities differ also in terms of the measure μ_n of outside investors, integration can raise the variance of the arbitrageurs' total wealth. This happens when markets with large ϵ_n or u_n have small μ_n . Indirect effects under segmentation remain large in these markets because they are independent of μ_n : with few outside investors, arbitrageurs hold small positions (implying small indirect effects) but spreads are highly sensitive to changes in their positions (implying large indirect effects). On the other hand, direct effects are small because positions are small. Therefore, convexity is reversed.

Integration can raise the variance of aggregate liquidity. Indeed, the gains or losses of an outside investor who is either on the long or on the short side of all of the arbitrageurs' positions are opposite to those of arbitrageurs. Therefore, they reflect the movements of the arbitrageurs' total wealth.

Finally, integration can raise the variance of some individual spreads. This happens for the opportunities with low ϵ_n or u_n . Intuitively, direct and indirect effects are small for these opportunities under segmentation, but the indirect effect becomes larger under integration. Integration cannot raise the variance of all spreads, however, consistent with Proposition 7.

6 Conclusion

We develop a model in which arbitrageurs' limited access to capital affects the functioning of financial markets. Arbitrageurs in our model are uniquely able to exploit mispricings, but face financial constraints limiting their ability to do so. Two properties structure the equilibrium: the equalization of return per unit of collateral across opportunities in which arbitrageurs invest, and the self-correcting dynamics of arbitrage capital and profitability. Based on these properties, we derive—analytically—a rich set of implications for the dynamics of liquidity and returns in the cross-section, and for the effects of the mobility of arbitrage capital on market stability.

Our model focuses on riskless arbitrage: while each of the legs of an arbitrage trade is risky, the risk cancels out. Considering unanticipated shocks, as we do in this paper, gives a preview of some of the results that can be derived under risky arbitrage. Yet, a fuller analysis of risky arbitrage can yield important new economic insights. Our model can be extended to accommodate arbitrage risk: both “fundamental risk,” whereby the two assets in a pair fail to have the same payoff, and “demand risk,” whereby endowment shocks of outside investors fluctuate in each period. When both types of risk are small, an equilibrium in continuous time can be derived analytically using perturbation methods around the solution under riskless arbitrage (Gromb and Vayanos (2015)). Within this framework, it is possible to address questions such as how arbitrageurs manage the risk of their portfolio, how the risk of each trade in the portfolio is affected by their aggregate capital, and how that risk is related to expected returns in the cross-section.

APPENDIX

A Proofs

Proof of Proposition 1: The investor's Bellman equation is

$$V_{i,t}(w_{i,t}) = \max_{c_{i,t+1}, y_{i,t}} \mathbb{E}_t \{-\gamma \exp(-\alpha c_{i,t+1}) + \gamma V_{i,t+1}(w_{i,t+1})\}. \quad (\text{A.1})$$

Substituting (13) and (14) into (A.1), we find

$$\begin{aligned} -\exp(-Aw_{i,t} - F_{i,t}) &= \max_{c_{i,t+1}, y_{i,t}} \mathbb{E}_t \{-\gamma \exp(-\alpha c_{i,t+1}) \\ &\quad -\gamma \exp(-A[(1+r)w_{i,t} + y_{i,t}\Phi_{i,t} + (y_{i,t} + u_{i,t})\epsilon_{i,t+1} - c_{i,t+1}] - F_{i,t+1})\}. \end{aligned} \quad (\text{A.2})$$

The first-order condition with respect to consumption is

$$\alpha \exp(-\alpha c_{i,t+1}) = A \exp(-A[(1+r)w_{i,t} + y_{i,t}\Phi_{i,t} + (y_{i,t} + u_{i,t})\epsilon_{i,t+1} - c_{i,t+1}] - F_{i,t+1}) \quad (\text{A.3})$$

$$\Rightarrow c_{i,t+1} = \frac{A[(1+r)w_{i,t} + y_{i,t}\Phi_{i,t} + (y_{i,t} + u_{i,t})\epsilon_{i,t+1}] + F_{i,t+1} + \log\left(\frac{\alpha}{A}\right)}{\alpha + A}. \quad (\text{A.4})$$

Hence, we can write the right-hand side of (A.2) as

$$\begin{aligned} &\max_{y_{i,t}} \mathbb{E}_t \left\{ -\frac{\gamma(\alpha + A)}{\alpha} \exp(-A[(1+r)w_{i,t} + y_{i,t}\Phi_{i,t} + (y_{i,t} + u_{i,t})\epsilon_{i,t+1} - c_{i,t+1}] - F_{i,t+1}) \right\} \\ &= \max_{y_{i,t}} \mathbb{E}_t \left\{ -\frac{\gamma(\alpha + A)}{\alpha} \exp\left(\frac{-\alpha \{A[(1+r)w_{i,t} + y_{i,t}\Phi_{i,t} + (y_{i,t} + u_{i,t})\epsilon_{i,t+1}] + F_{i,t+1}\} + A \log\left(\frac{\alpha}{A}\right)}{\alpha + A}\right) \right\} \end{aligned} \quad (\text{A.5})$$

$$= \max_{y_{i,t}} \left\{ -\frac{\gamma(\alpha + A)}{\alpha} \exp\left(\frac{-\alpha \{A[(1+r)w_{i,t} + y_{i,t}\Phi_{i,t} - f[(y_{i,t} + u_{i,t})\bar{\epsilon}_i]] + F_{i,t+1}\} + A \log\left(\frac{\alpha}{A}\right)}{\alpha + A}\right) \right\}, \quad (\text{A.6})$$

where the first step follows from (A.3), the second from (A.4), and the third by from (16) by setting $y \equiv (y_{i,t} + u_{i,t})\bar{\epsilon}_i$. Eq. (A.6) implies that the first-order condition with respect to $y_{i,t}$ is (15).

Equating (A.6) to the right-hand side of (A.2), we find that the Bellman equation holds for all

values of the single state variable $w_{i,t}$ if

$$A = \frac{\alpha A(1+r)}{\alpha + A}, \quad (\text{A.7})$$

$$F_{i,t} = -\log\left(\frac{\gamma(\alpha + A)}{\alpha}\right) + \frac{\alpha \{A [y_{i,t}\Phi_{i,t} - f[(y_{i,t} + u_{i,t})\bar{\epsilon}_i]] + F_{i,t+1}\} - A \log\left(\frac{\alpha}{A}\right)}{\alpha + A}. \quad (\text{A.8})$$

Eq. (A.7) implies that $A = r\alpha$. Substituting into (A.8), we find

$$F_{i,t} = -\log[\gamma(1+r)] + \frac{r\alpha [y_{i,t}\Phi_{i,t} - f[(y_{i,t} + u_{i,t})\bar{\epsilon}_i]] + F_{i,t+1} + r \log(r)}{1+r}. \quad (\text{A.9})$$

Eq. (A.9) and the transversality condition $\lim_{s \rightarrow \infty} \frac{F_{i,t+s}}{(1+r)^s} = 0$ determine $F_{i,t}$. ■

Proof of Lemma 1: To prove the properties in the lemma, we set $\hat{\alpha} \equiv \frac{\alpha A}{\alpha + A}$ and $\hat{\epsilon}_{i,t} \equiv \frac{\epsilon_{i,t}}{\bar{\epsilon}_i}$. Since the distribution of $\hat{\epsilon}_{i,t}$ is independent of i and t , so is the function $f(y)$. Since $\hat{\epsilon}_{i,t}$ has mean zero, Jensen's inequality implies that

$$E[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})] \geq \exp(0) = 1,$$

and hence $f(y) \geq 0$. Since $\hat{\epsilon}_{i,t}$ is distributed symmetrically around zero,

$$E[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})] = E[\exp(\hat{\alpha}y\hat{\epsilon}_{i,t})],$$

and hence $f(y)$ is symmetric around the vertical axis. To show that $f(y)$ is strictly convex, we show that $f''(y) > 0$. Since

$$f(y) = \frac{\log\{E[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})]\}}{\hat{\alpha}},$$

differentiating once we find

$$f'(y) = -\frac{E[\hat{\epsilon}_{i,t} \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})]}{E[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})]}, \quad (\text{A.10})$$

and differentiating twice we find

$$f''(y) = \hat{\alpha} \frac{E[\hat{\epsilon}_{i,t}^2 \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})] E[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})] - \{E[\hat{\epsilon}_{i,t} \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})]\}^2}{\{E[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})]\}^2}. \quad (\text{A.11})$$

The numerator in (A.11) is positive because of the Cauchy-Schwarz inequality $[E(XY)]^2 \leq E(X^2)E(Y^2)$, which is strict when the random variables X and Y are not proportional. We can use the Cauchy-Schwarz inequality by setting

$$X \equiv \hat{\epsilon}_{i,t} \exp\left(-\frac{\hat{\alpha}y\hat{\epsilon}_{i,t}}{2}\right),$$

$$Y \equiv \exp\left(-\frac{\hat{\alpha}y\hat{\epsilon}_{i,t}}{2}\right),$$

and noting that X and Y are not proportional because $\hat{\epsilon}_{i,t}$ is stochastic. Therefore, $f''(y) > 0$. To show that $\lim_{y \rightarrow \infty} f'(y) = 1$, we show that $|f'(y) - 1|$ can be made smaller than 2η for any arbitrary $\eta > 0$ when y goes to infinity. Using (A.10) and the fact that $\hat{\epsilon}_{i,t}$ is distributed symmetrically around zero with the supremum of its support being one, we find

$$\begin{aligned} |f'(y) - 1| &= \frac{E[(1 + \hat{\epsilon}_{i,t}) \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})]}{E[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})]} \\ &= \frac{E[(1 + \hat{\epsilon}_{i,t}) \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \mathbf{1}_{\{\hat{\epsilon}_{i,t} \in [-1, -1+\eta]\}}]}{E[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})]} + \frac{E[(1 + \hat{\epsilon}_{i,t}) \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \mathbf{1}_{\{\hat{\epsilon}_{i,t} \in (-1+\eta, 1]\}}]}{E[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})]}. \end{aligned} \tag{A.12}$$

Since

$$(1 + \hat{\epsilon}_{i,t}) \mathbf{1}_{\{\hat{\epsilon}_{i,t} \in [-1, -1+\eta]\}} \leq \eta,$$

the first term in the right-hand side of (A.12) is smaller than η . The second term can also be made smaller than η for large y . Indeed, multiplying numerator and denominator by $\exp(-\hat{\alpha}y(1 - \eta))$, we can write this term as

$$\frac{E[(1 + \hat{\epsilon}_{i,t}) \exp(-\hat{\alpha}y(\hat{\epsilon}_{i,t} + 1 - \eta)) \mathbf{1}_{\{\hat{\epsilon}_{i,t} \in (-1+\eta, 1]\}}]}{E[\exp(-\hat{\alpha}y(\hat{\epsilon}_{i,t} + 1 - \eta))]} \tag{A.13}$$

Since $\hat{\epsilon}_{i,t}$ in the numerator of (A.13) exceeds $-1 + \eta$, the numerator remains bounded when y goes to infinity. The denominator of (A.13) converges to infinity, however, because $\epsilon_{i,t}$ takes values in $[-1, -1 + \eta)$ with positive probability. ■

Proof of Proposition 2: The arbitrageur's Bellman equation is

$$V_t(W_t) = \max_{c_{t+1}, \{x_{i,t}\}_{i \in \mathcal{A}_t}} \{\beta \log(c_{i,t+1}) + \beta V_{t+1}(W_{t+1})\}. \quad (\text{A.14})$$

Substituting (20) and (22) into (A.14), we find

$$B \log(W_t) + G_t = \max_{c_{t+1}, \{x_{i,t}\}_{i \in \mathcal{A}_t}} \left\{ \beta \log(c_{i,t+1}) + \beta B \log \left((1+r)W_t + 2 \sum_{i \in \mathcal{A}_t} x_{i,t} \Phi_{i,t} - c_{t+1} \right) + \beta G_{t+1} \right\}. \quad (\text{A.15})$$

The first-order condition with respect to consumption is

$$\begin{aligned} \frac{1}{c_{i,t+1}} - \frac{B}{(1+r)W_t + 2 \sum_{i \in \mathcal{A}_t} x_{i,t} \Phi_{i,t} - c_{t+1}} &= 0 \\ \Rightarrow c_{i,t+1} &= \frac{(1+r)W_t + 2 \sum_{i \in \mathcal{A}_t} x_{i,t} \Phi_{i,t} - c_{t+1}}{B} = \frac{W_{t+1}}{B} \end{aligned} \quad (\text{A.16})$$

$$\Rightarrow c_{i,t+1} = \frac{(1+r)W_t + 2 \sum_{i \in \mathcal{A}_t} x_{i,t} \Phi_{i,t}}{B+1}, \quad (\text{A.17})$$

where the second equality in (A.16) follows from (20). Using (A.17), we can write the right-hand side of (A.15) as

$$\max_{\{x_{i,t}\}_{i \in \mathcal{A}_t}} \left\{ \beta(B+1) \log \left((1+r)W_t + 2 \sum_{i \in \mathcal{A}_t} x_{i,t} \Phi_{i,t} \right) + \beta B \log(B) - \beta(B+1) \log(B+1) + \beta G_{t+1} \right\}. \quad (\text{A.18})$$

The maximization in (A.18) is subject to the financial constraint (21). Expected excess returns are assumed to satisfy $\Phi_{i,t} \geq 0$ for all $i \in \mathcal{A}_t$. Moreover, (15) and Lemma 1 imply that $\Phi_{i,t} < \bar{\epsilon}_i$ for all $i \in \mathcal{A}_t$.

When $\Phi_{i,t} = 0$ for all $i \in \mathcal{A}_t$, the arbitrageur is indifferent between any combination of positions in the active opportunities. When $\Phi_{i,t} > 0$ for some $i \in \mathcal{A}_t$, (21) binds, and this implies that $x_{i,t} \geq 0$ for all $i \in \mathcal{A}_t$. Indeed, if $x_{i,t} < 0$ for some $i \in \mathcal{A}_t$, then setting those $x_{i,t}$ to zero would relax (21). Hence, (A.18) could be raised by increasing those $x_{i,t}$ for which $\Phi_{i,t} > 0$. Since $x_{i,t} \geq 0$ for all $i \in \mathcal{A}_t$, (21) becomes (26). Maximizing (A.18) subject to (26) and $x_{i,t} \geq 0$, we find that the arbitrageur holds non-zero positions only in opportunities $i \in \mathcal{M}_t \equiv \operatorname{argmax}_{j \in \mathcal{A}_t} \frac{\Phi_{j,t}}{\bar{\epsilon}_j}$.

Setting

$$\Pi_t \equiv \max_{j \in \mathcal{A}_t} \frac{\Phi_{j,t}}{\bar{\epsilon}_j} \in [0, 1)$$

and using the characterization of the arbitrageur's optimal positions, we find

$$(1+r)W_t + 2 \sum_{i \in \mathcal{A}_t} x_{i,t} \Phi_{i,t} = \frac{1+r}{1-\Pi_t} W_t. \quad (\text{A.19})$$

When $\Pi_t = 0$, (A.19) follows because $\Phi_{i,t} = 0$ for all $i \in \mathcal{A}_t$. When $\Pi_t > 0$, (A.19) follows by writing the left-hand side as

$$(1+r)W_t + 2 \sum_{i \in \mathcal{M}_t} x_{i,t} \Phi_{i,t} = (1+r)W_t + 2\Pi_t \sum_{i \in \mathcal{M}_t} x_{i,t} \bar{\epsilon}_i,$$

and combining with (26), which binds and can be written as

$$W_t = 2 \sum_{i \in \mathcal{M}_t} \frac{x_{i,t} (\bar{\epsilon}_i - \Phi_{i,t})}{1+r} = 2(1-\Pi_t) \sum_{i \in \mathcal{M}_t} \frac{x_{i,t} \bar{\epsilon}_i}{1+r}.$$

Substituting (A.19) into (A.18), and equating (A.18) to the right-hand side of (A.15), we find that the Bellman equation holds for all values of the single state variable W_t if

$$B = \beta(B+1), \quad (\text{A.20})$$

$$G_t = \beta(B+1) \log \left(\frac{(1+r)\Pi_t}{1-\Pi_t} \right) + \beta B \log(B) - \beta(B+1) \log(B+1) + \beta G_{t+1}. \quad (\text{A.21})$$

Eq. (A.7) implies that $B = \frac{\beta}{1-\beta}$. Substituting into (A.16), we find (23). Eq. (A.21) and the transversality condition $\lim_{s \rightarrow \infty} \beta^s G_{t+s} = 0$ determine G_t . ■

Proof of Proposition 3: We define Π_t as in the proof of Proposition 2. Proposition 2 implies that if arbitrageurs trade opportunity $(i, -i)$ then $\frac{\Phi_{i,t}}{\bar{\epsilon}_i} = \Pi_t$, and if they do not trade it then $\frac{\Phi_{i,t}}{\bar{\epsilon}_i} \leq \Pi_t$. In the former case, $x_{i,t} > 0$ and (9) imply that $y_{i,t} < 0$. Substituting into (15) and using the convexity of $f(y)$, we find $f'(u_i \bar{\epsilon}_i) > \frac{\Phi_{i,t}}{\bar{\epsilon}_i} = \Pi_t$. In the latter case, $x_{i,t} = 0$ and (9) imply that $y_{i,t} = 0$. Substituting into (15), we find $f'(u_i \bar{\epsilon}_i) = \frac{\Phi_{i,t}}{\bar{\epsilon}_i} \leq \Pi_t$. ■

Proof of Proposition 4: Substituting c_{t+1} from (23) into (20), and solving for W_{t+1} , we find (25). Combining with (A.19) yields (27).

If $\Pi_t = 0$, then Proposition 3 implies that all active opportunities are traded, and (A.19) implies that arbitrageurs earn the riskless rate r . To determine a lower bound on W_t , we use market clearing and the financial constraint. Eq. (15) implies that $f'[(y_{i,t} + u_i)\bar{\epsilon}_i] = \frac{\Phi_{i,t}}{\bar{\epsilon}_i} = 0$ for all $i \in \mathcal{A}_t$. Since $f(y)$ is symmetric around the vertical axis (Lemma 1), $f'(0) = 0$. Strict convexity of $f(y)$ implies that $f'(y)$ is invertible and hence $y_{i,t} + u_i = 0$. Combining with (9), we find $x_{i,t} = \mu_i u_i$. Substituting $x_{i,t} = \mu_i u_i$ into (26) and using $\Phi_{i,t} = 0$ for all $i \in \mathcal{A}_t$, we find $W_t \geq W_{c,t}$.

If $\Pi_t > 0$, then Proposition 2 implies that the financial constraint binds, and (A.19) implies that arbitrageurs earn the riskless return $\frac{1+r}{1+\Pi_t} - 1 > r$. To determine how Π_t relates to W_t , we use market clearing and the financial constraint. Eq. (15) and Proposition 3 imply that $f'[(y_{i,t} + u_i)\bar{\epsilon}_i] = \frac{\Phi_{i,t}}{\bar{\epsilon}_i} = \Pi_t$ for all $i \in \mathcal{T}_t$. Inverting this equation yields

$$\begin{aligned} (y_{i,t} + u_i)\bar{\epsilon}_i &= (f')^{-1}(\Pi_t) \\ \Rightarrow x_{i,t} &= \mu_i u_i - \mu_i \frac{(f')^{-1}(\Pi_t)}{\bar{\epsilon}_i}, \end{aligned} \tag{A.22}$$

where the second step follows from (9). Since $x_{i,t} = 0$ for all $i \in \mathcal{A}_t/\mathcal{T}_t$, we can write (26) (which binds) as

$$W_t = 2 \sum_{i \in \mathcal{T}_t} \frac{x_{i,t}(\bar{\epsilon}_i - \Phi_{i,t})}{1+r} \tag{A.23}$$

$$= 2(1 - \Pi_t) \sum_{i \in \mathcal{T}_t} \frac{x_{i,t}\bar{\epsilon}_i}{1+r}, \tag{A.24}$$

where the second step follows from Proposition 3. Since $0 < x_{i,t} < \mu_i u_{i,t}$ for all $i \in \mathcal{T}_t$ (because of (A.22)) and $\Pi_t \in (0, 1)$, (A.23) implies that $W_t < W_{c,t}$. Substituting $x_{i,t}$ from (A.22) into (A.24), we find (28). The left-hand side of (28) decreases in Π_t because $f''(y) > 0$ implies that $f'(y)$ is increasing. Moreover, it is equal to zero for $\Pi_t = \max_{i \in \mathcal{A}_t} f'(u_i \bar{\epsilon}_i)$ and to $W_{c,t}$ for $\Pi_t = 0$. Therefore, (28) has a unique positive solution for $W_t \in (0, W_{c,t})$, which decreases in W_t .

To show convexity of Π_t , we differentiate (28) implicitly with respect to W_t . We find

$$\frac{d\Pi_t}{dW_t} = - \frac{1}{\frac{2}{1+r} \sum_{i \in \mathcal{T}_t} \mu_i \left[u_i \bar{\epsilon}_i - (f')^{-1}(\Pi_t) + \frac{1 - \Pi_t}{f''[(f')^{-1}(\Pi_t)]} \right]}. \tag{A.25}$$

The derivative $\frac{d\Pi_t}{dW_t}$ is continuous, except at $W_t = W_{c,t}$ and at the points where the set \mathcal{T}_t changes. For those values of W_t , the left derivative is smaller than the right derivative. For $W_t = W_{c,t}$, this is because the left derivative is negative and the right derivative is zero. For a point where \mathcal{T}_t changes, this is because the denominator for the right derivative minus that for the left derivative is

$$\frac{2}{1+r} \sum_{i \in \mathcal{DT}_t} \mu_i \left[u_i \bar{e}_i - (f')^{-1}(\Pi_t) + \frac{1 - \Pi_t}{f''[(f')^{-1}(\Pi_t)]} \right] = \frac{2}{1+r} \sum_{i \in \mathcal{DT}_t} \mu_i \left[\frac{1 - \Pi_t}{f''[(f')^{-1}(\Pi_t)]} \right] > 0,$$

where \mathcal{DT}_t denotes the additional opportunities that become traded to the right of that point. Therefore, Π_t is convex if the function

$$u_i \bar{e}_i - (f')^{-1}(\Pi_t) + \frac{1 - \Pi_t}{f''[(f')^{-1}(\Pi_t)]}$$

is increasing in W_t , or equivalently is decreasing in Π_t . This is also equivalent to the function

$$G(y) \equiv -y + \frac{1 - f'(y)}{f''(y)}$$

being decreasing in y because $f'(y)$ is increasing. The derivative of $G(y)$ with respect to y has the same sign as the function

$$G_1(y) \equiv -2f''(y)^2 - f'''(y)(1 - f'(y)).$$

Differentiating (A.11) we find

$$\begin{aligned} f'''(y) = \hat{\alpha}^2 & \left[2E[\hat{\epsilon}_{i,t} \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})] \frac{E[\hat{\epsilon}_{i,t}^2 \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})] E[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})] - \{E[\hat{\epsilon}_{i,t} \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})]\}^2}{\{E[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})]\}^3} \right. \\ & \left. + \frac{E[\hat{\epsilon}_{i,t}^2 \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})] E[\hat{\epsilon}_{i,t} \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})] - E[\hat{\epsilon}_{i,t}^3 \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})] E[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})]}{\{E[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t})]\}^2} \right]. \end{aligned} \tag{A.26}$$

Using (A.10), (A.11) and (A.26), we find

$$\begin{aligned}
G_1(y) = & -\hat{\alpha}^2 \left[2 \left(E \left[\hat{\epsilon}_{i,t}^2 \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right] E \left[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right] + E \left[\hat{\epsilon}_{i,t} \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right] \right) \right. \\
& \times \frac{E \left[\hat{\epsilon}_{i,t}^2 \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right] E \left[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right] - \{E \left[\hat{\epsilon}_{i,t} \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right]\}^2}{\{E \left[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right]\}^4} \\
& + E \left[(1 + \hat{\epsilon}_{i,t}) \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right] \\
& \left. \times \frac{E \left[\hat{\epsilon}_{i,t}^2 \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right] E \left[\hat{\epsilon}_{i,t} \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right] - E \left[\hat{\epsilon}_{i,t}^3 \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right] E \left[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right]}{\{E \left[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right]\}^3} \right].
\end{aligned}$$

When the distribution of $\epsilon_{i,t}$ is binomial, $\hat{\epsilon}_{i,t}$ has also a binomial distribution that takes the values 1 and -1 with equal probabilities. Therefore,

$$E \left[\exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right] = E \left[\hat{\epsilon}_{i,t}^2 \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right] = \cosh(\hat{\alpha}y),$$

$$E \left[\hat{\epsilon}_{i,t} \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right] = E \left[\hat{\epsilon}_{i,t}^3 \exp(-\hat{\alpha}y\hat{\epsilon}_{i,t}) \right] = -\sinh(\hat{\alpha}y),$$

and the function $G_1(y)$ becomes

$$G_1(y) = -\hat{\alpha}^2 \frac{2 (\cosh^2(\hat{\alpha}y) - \sinh(\hat{\alpha}y))}{\cosh^4(\hat{\alpha}y)}.$$

Since $\cosh(x) \geq 1$ and $\cosh(x) > \sinh(x)$, $G_1(y)$ is negative and hence $G(y)$ is decreasing. ■

Proof of Proposition 5: Suppose that in equilibrium (i) expected excess returns $\Phi_{i,t}$ for assets $i \in \mathcal{A}_t$ are given by Propositions 3 and 4, (ii) risk premia $\phi_{i,t}$ for assets $i \in \mathcal{A}_t$ are given by solving (12) backwards with the terminal condition $\phi_{i,h_i} = 0$:

$$\phi_{i,t} = \sum_{s=0}^{h_i-t-1} \frac{\Phi_{i,t+s}}{(1+r)^{s+1}}, \tag{A.27}$$

(iii) expected excess returns and risk premia for assets $-i$, $i \in \mathcal{A}_t$, are opposites to those for assets i , and (iv) expected excess returns and risk premia for assets that are not part of active opportunities are zero. The optimization problems of investors and arbitrageurs are then as in Sections 3.2 and 3.3. The analysis in these sections and in Section 3.4 ensures that the markets for all assets clear and that the quantities $(\phi_{i,t}, y_{i,t}, x_{i,t}, W_t)$ have the properties in the proposition. ■

Proof of Proposition 6: The dynamics of W_t in the three cases of the proposition are as follows:

- If $\beta(1+r) > 1$, then (27) implies that W_t increases to $W = \infty$.
- If $\beta(1+r) < 1 - \bar{\Pi}$, then (27) implies that W_t decreases to $W = 0$ because $\bar{\Pi}$ is the maximum value of Π_t .
- If $1 - \bar{\Pi} < \beta(1+r) < 1$, then (27) implies that W_t remains constant when Π_t is equal to the steady-state value Π given by (30). The steady-state value W of W_t is given by (29) because of (28). When $W_t < W$, (27) implies that $W_{t+1} > W_t$ because $\Pi_t > \Pi$. Conversely, when $W_t > W$, (27) implies that $W_{t+1} < W_t$ because $\Pi_t < \Pi$. To show that convergence of W_t to W is monotone, we need to show that $W_{t+1} < W$ in the former case and $W_{t+1} > W$ in the latter case. Since (27) implies that

$$W_{t+1} - W = \beta(1+r) \left(\frac{W_t}{1 - \Pi_t} - \frac{W}{1 - \Pi} \right),$$

convergence is monotone if the function $F(W_t) = \frac{W_t}{1 - \Pi_t}$ is increasing in W_t , where Π_t is defined implicitly as function of W_t from Proposition 4. When $W_t > W_c$, $F(W_t)$ is increasing in W_t because $\Pi_t = 0$. When $W_t < W_c$, (28) implies that

$$F(W_t) = \frac{2}{1+r} \sum_{i \in \mathcal{I}_t} \mu_i [u_i \bar{\epsilon}_i - (f')^{-1}(\Pi_t)].$$

Since $f(y)$ is strictly convex and Π_t decreases in W_t , $F(W_t)$ is increasing in W_t . Since $\Pi > 0$, Proposition 4 implies that $W \in (0, W_c)$.

The dynamics of Π_t in each of the three cases follow from the dynamics of W_t , and from the dependence of Π_t on W_t derived in Proposition 4. ■

Proof of Corollary 1: We first show the properties of returns. Since arbitrageurs' return per unit of volatility is decreasing in their wealth within the constrained region (Proposition 4), it increases to a value $\Pi_t > \Pi$ when wealth drops to $W_t < W$. It then decreases over time and converges back to Π (Proposition 6). Proposition 3 implies that an opportunity (n, m) traded in steady state satisfies $\Phi_n = \bar{\epsilon}_n \Pi$ and $f'(u_n \bar{\epsilon}_n) > \Pi$. If arbitrageurs remain invested in this opportunity after their initial drop in wealth, then $\Phi_{n,t} = \bar{\epsilon}_n \Pi_t > \bar{\epsilon}_n \Pi$. If they exit the opportunity, then $\Phi_{n,t} = \bar{\epsilon}_n f'(u_n \bar{\epsilon}_n) > \bar{\epsilon}_n \Pi$. In both cases, $\Phi_{n,t} > \Phi_n$. If the opportunity is traded in a period $t + s$ for $s = 0, \dots, m - 2$, then it is traded in all periods until $t + m - 1$ because Π_t decreases over

time. Moreover, $\Phi_{n,t+s} = \bar{\epsilon}_n \Pi_{t+s} > \bar{\epsilon}_n \Pi_{t+s+1} = \Phi_{n,t+s+1} > \bar{\epsilon}_n \Pi$. If the opportunity is not traded in period $t+s$ but is traded in period $t+s+1$, then $\Phi_{n,t+s} = \bar{\epsilon}_n f'(u_n \bar{\epsilon}_n) > \bar{\epsilon}_n \Pi_{t+s+1} = \Phi_{n,t+s+1} > \bar{\epsilon}_n \Pi$. If the opportunity is not traded in both periods, then $\Phi_{n,t+s} = \bar{\epsilon}_n f'(u_n \bar{\epsilon}_n) = \Phi_{n,t+s+1} > \bar{\epsilon}_n \Pi$. In all three cases, $\Phi_{n,t+s} \geq \Phi_{n,t+s+1} > \Phi_n$.

We next show the properties of spreads. Using (31), its steady-state version

$$\phi_{n,m} = \sum_{s=0}^{m-1} \frac{\Phi_n}{(1+r)^{s+1}}, \quad (\text{A.28})$$

and $\Phi_{n,t+s} > \Phi_n$ for $s = 0, \dots, m-1$, we find $\phi_{n,m,t} > \phi_{n,m}$. Using (31) written for $(m-s, t+s)$ and $(m-s-1, t+s+1)$ instead of (m, t) , (A.28), and $\Phi_{n,t+s} \geq \Phi_{n,t+s'}$ for $s' = s+1, \dots, m-1$, we find $\frac{\phi_{n,m-s,t+s}}{\phi_{n,m-s}} \geq \frac{\phi_{n,m-s-1,t+s+1}}{\phi_{n,m-s-1}}$ for $s = 0, \dots, m-2$.

The properties of positions follow from those of returns by using (9), (15) and the fact that $f'(y)$ is increasing. ■

Proof of Corollary 2: We first show the properties of returns. Since opportunities (n, m) and (n', m) are traded in steady state, $\Phi_n = \bar{\epsilon}_n \Pi > \bar{\epsilon}_{n'} \Pi = \Phi_{n'}$. If arbitrageurs remain invested in both opportunities after their initial drop in wealth, then

$$\Phi_{n,t} - \Phi_n = \bar{\epsilon}_n (\Pi_t - \Pi) > \bar{\epsilon}_{n'} (\Pi_t - \Pi) = \Phi_{n',t} - \Phi_{n'}.$$

If they remain invested only in one opportunity, it has to be (n, m) : if they remained invested only in (n', m) , then

$$\frac{\Phi_{n,t}}{\bar{\epsilon}_n} = f'(u_n \bar{\epsilon}_n) \leq \Pi_t = \frac{\Phi_{n',t}}{\bar{\epsilon}_{n'}} < f'(u_{n'} \bar{\epsilon}_{n'}),$$

which contradicts $\bar{\epsilon}_n > \bar{\epsilon}_{n'}$ and $u_n = u_{n'}$. Since they are invested in (n, m) and not in (n', m) ,

$$\Phi_{n,t} - \Phi_n = \bar{\epsilon}_n (\Pi_t - \Pi) > \bar{\epsilon}_{n'} (\Pi_t - \Pi) \geq \bar{\epsilon}_{n'} (f'(u_{n'} \bar{\epsilon}_{n'}) - \Pi) = \Phi_{n',t} - \Phi_{n'}.$$

If they exit both opportunities, then

$$\Phi_{n,t} - \Phi_n = \bar{\epsilon}_n (f'(u_n \bar{\epsilon}_n) - \Pi) > \bar{\epsilon}_{n'} (f'(u_{n'} \bar{\epsilon}_{n'}) - \Pi) = \Phi_{n',t} - \Phi_{n'}.$$

In all three cases, $\Phi_{n,t} - \Phi_n > \Phi_{n',t} - \Phi_{n'}$. If opportunities (n, m) and (n', m) are traded in a period

$t + s + 1$ for $s = 0, \dots, m - 2$, then the inequality

$$\Phi_{n,t+s} - \Phi_{n,t+s+1} \geq \Phi_{n',t+s} - \Phi_{n',t+s+1} \quad (\text{A.29})$$

follows from the previous arguments by replacing Π by Π_{t+s+1} . If only one opportunity is traded in period $t + s + 1$, it has to be (n, m) from the previous arguments. Eq. (A.29) then follows from $\Phi_{n,t+s} > \Phi_{n,t+s+1}$, shown in the proof of Corollary 1, and $\Phi_{n',t+s} = \Phi_{n',t+s+1} = f'(u_{n'}\bar{\epsilon}_{n'})$. If both opportunities are not traded in period $t + s + 1$, then (A.29) follows from $\Phi_{n,t+s} = \Phi_{n,t+s+1} = f'(u_n\bar{\epsilon}_n)$ and $\Phi_{n',t+s} = \Phi_{n',t+s+1} = f'(u_{n'}\bar{\epsilon}_{n'})$. In all three cases, (A.29) holds.

We next show the properties of spreads. Using (A.28) and $\Phi_n > \Phi_{n'}$, we find $\phi_{n,m} > \phi_{n',m}$. Subtracting (31) from (A.28), and using $\Phi_{n,t+s} - \Phi_n > \Phi_{n',t+s} - \Phi_{n'}$ for $s = 0, \dots, m - 1$ (which holds for $s > 0$ by the same argument as for $s = 0$), we find $\phi_{n,m,t} - \phi_{n,m} > \phi_{n',m,t} - \phi_{n',m}$. Subtracting (31) written for $(m - s - 1, t + s + 1)$ from the same equation written for $(m - s, t + s)$, and using $\Phi_{n,t+s'} - \Phi_{n,t+s'+1} \geq \Phi_{n',t+s'} - \Phi_{n',t+s'+1}$ for $s' = s, \dots, m - 1$, we find $\phi_{n,m-s,t+s} - \phi_{n,m-s-1,t+s+1} \geq \phi_{n',m-s,t+s} - \phi_{n',m-s-1,t+s+1}$ for $s = 0, \dots, m - 2$.

We finally show the properties of positions. Since opportunities (n, m) and (n', m) are traded in steady state, and $(\mu_n, u_n) = (\mu_{n'}, u_{n'})$, (A.22) implies that

$$x_n = \mu_n u_n - \mu_n \frac{(f')^{-1}(\Pi)}{\bar{\epsilon}_n} > \mu_{n'} u_{n'} - \mu_{n'} \frac{(f')^{-1}(\Pi)}{\bar{\epsilon}_{n'}} = x_{n'}.$$

If both opportunities are traded in period t , and hence in all periods until $t + m - 1$, then (A.22) implies that

$$x_n - x_{n,t} = \frac{\mu_n}{\bar{\epsilon}_n} [(f')^{-1}(\Pi_t) - (f')^{-1}(\Pi)] < \frac{\mu_{n'}}{\bar{\epsilon}_{n'}} [(f')^{-1}(\Pi_t) - (f')^{-1}(\Pi)] = x_{n'} - x_{n',t},$$

and

$$\begin{aligned} x_{n,t+s+1} - x_{n,t+s} &= \frac{\mu_n}{\bar{\epsilon}_n} [(f')^{-1}(\Pi_{t+s}) - (f')^{-1}(\Pi_{t+s+1})] \\ &< \frac{\mu_{n'}}{\bar{\epsilon}_{n'}} [(f')^{-1}(\Pi_{t+s}) - (f')^{-1}(\Pi_{t+s+1})] \\ &= x_{n',t+s+1} - x_{n',t+s} \end{aligned}$$

for $s = 0, \dots, m - 2$. If only one opportunity is traded in period t , it has to be (n, m) as shown previously. ■

Proof of Corollary 3: Combining (A.28) and $\Phi_n > 0$, we find $\phi_{n,m} > \phi_{n,m'}$. Combining (31), (A.28) and $\Phi_{n,t+s} > \Phi_n$ for $s = 0, \dots, m-1$, we find $\phi_{n,m,t} - \phi_{n,m} > \phi_{n,m',t} - \phi_{n,m'}$. Combining (31) written for $(m-s, t+s)$ and $(m-s-1, t+s+1)$, and $\Phi_{n,t+s'} > 0$ for $s' = m', \dots, m-1$, we find $\phi_{n,m-s,t+s} - \phi_{n,m-s-1,t+s+1} \leq \phi_{n,m'-s,t+s} - \phi_{n,m'-s-1,t+s+1}$ for $s = 0, \dots, m' - 2$. ■

Proof of Corollary 4: The proof follows from the argument in the paragraph just before the proposition. ■

Proof of Proposition 7: We first compute the partial derivatives of W_t under integration. Differentiating (32) with respect to η_{nt} , and using $M_n = M$ for all $n = 1, \dots, N$, we find

$$\begin{aligned} \frac{\partial W_t}{\partial \eta_{nt}} &= \frac{2\beta M x_n \bar{\epsilon}_n}{1 + 2\beta \sum_{n'' \in \mathcal{N}} x_{n''} \sum_{m=1}^{M-1} \frac{\partial \phi_{n'',m,t}}{\partial W_t}} \\ &= \frac{2\beta M \mu_n [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)]}{1 + 2\beta \sum_{n'' \in \mathcal{N}} \mu_{n''} \left[u_{n''} - \frac{(f')^{-1}(\Pi)}{\bar{\epsilon}_{n''}} \right] \sum_{m=1}^{M-1} \frac{\partial \phi_{n'',m,t}}{\partial W_t}}, \end{aligned} \quad (\text{A.30})$$

where the second step follows by assuming $n \in \mathcal{N}$ and using (A.22). (For $n \notin \mathcal{N}$, $\frac{\partial W_t}{\partial \eta_{nt}} = 0$ because $x_n = 0$.) Using (31), we find that for $n' \in \mathcal{N}$,

$$\begin{aligned} \frac{\partial \phi_{n',m,t}}{\partial W_t} &= \sum_{s=0}^{m-1} \frac{\partial \Phi_{n',t+s}}{\partial W_t} \frac{1}{(1+r)^{s+1}} \\ &= \bar{\epsilon}_{n'} \sum_{s=0}^{m-1} \frac{\partial \Pi_{t+s}}{\partial W_t} \frac{1}{(1+r)^{s+1}} \\ &= \bar{\epsilon}_{n'} \sum_{s=0}^{m-1} \frac{\partial \Pi_{t+s}}{\partial W_{t+s}} \frac{\partial W_{t+s}}{\partial W_t} \frac{1}{(1+r)^{s+1}} \\ &= \bar{\epsilon}_{n'} \sum_{s=0}^{m-1} \frac{\partial \Pi_{t+s}}{\partial W_{t+s}} \left(\prod_{s'=0}^{s-1} \frac{\partial W_{t+s'+1}}{\partial W_{t+s'}} \right) \frac{1}{(1+r)^{s+1}}, \end{aligned} \quad (\text{A.31})$$

where the second step follows from Proposition 3 and because opportunities that are traded in steady state remain traded close to steady state. The partial derivatives in (A.30) and (A.31) are evaluated at the steady state. Eq. (A.25), written for $t+s$ instead of t , implies that in steady state

$$\frac{\partial \Pi_{t+s}}{\partial W_{t+s}} = - \frac{1}{\frac{2M}{1+r} \sum_{n'' \in \mathcal{N}} \mu_{n''} \left[u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]} \quad (\text{A.32})$$

for all $s \geq 0$. Writing (27) for $t + s$ instead of t , and differentiating with respect to W_{t+s} , we find

$$\frac{\partial W_{t+s+1}}{\partial W_{t+s}} = \beta \frac{1+r}{1-\Pi_t} + \beta \frac{1+r}{(1-\Pi_t)^2} W_{t+s} \frac{\partial \Pi_{t+s}}{\partial W_{t+s}}. \quad (\text{A.33})$$

Eq. (A.33) implies that in steady state

$$\begin{aligned} \frac{\partial W_{t+s+1}}{\partial W_{t+s}} &= \beta \frac{1+r}{1-\Pi} + \beta \frac{1+r}{(1-\Pi)^2} W \frac{\partial \Pi_{t+s}}{\partial W_{t+s}} \\ &= 1 + \frac{2M}{1+r} \sum_{n'' \in \mathcal{N}} \mu_{n''} [u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi)] \frac{\partial \Pi_{t+s}}{\partial W_{t+s}} = Z, \end{aligned} \quad (\text{A.34})$$

where the second step follows from (29) and (30), and the third from (A.32) and by setting

$$Z \equiv \frac{\sum_{n'' \in \mathcal{N}} \mu_{n''} \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}{\sum_{n'' \in \mathcal{N}} \mu_{n''} \left[u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]}.$$

Using (A.32) and (A.34), we can write (A.31) as

$$\frac{\partial \phi_{n',m,t}}{\partial W_t} = - \frac{\bar{\epsilon}_{n'} \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}}{2M \sum_{n'' \in \mathcal{N}} \mu_{n''} \left[u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]}, \quad (\text{A.35})$$

and (A.30) as

$$\frac{\partial W_t}{\partial \eta_{nt}} = \frac{2\beta M \mu_n [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)]}{1 - \frac{\beta \sum_{n'' \in \mathcal{N}} \mu_{n''} [u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}}{M \sum_{n'' \in \mathcal{N}} \mu_{n''} \left[u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]}. \quad (\text{A.36})$$

We next compute the partial derivatives of W_t under segmentation. Denoting by W_{nt} the wealth of arbitrageurs in segmented market $n \in \mathcal{N}$, and differentiating their budget constraint with respect to η_{nt} , we find the following counterpart of (A.30):

$$\frac{\partial W_{nt}}{\partial \eta_{nt}} = \frac{2\beta M \mu_n [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)]}{1 + 2\beta \mu_n \left[u_n - \frac{(f')^{-1}(\Pi)}{\bar{\epsilon}_n} \right] \sum_{m=1}^{M-1} \frac{\partial \phi_{n,m,t}}{\partial W_t}}. \quad (\text{A.37})$$

The counterparts of (A.35) and (A.36) are

$$\frac{\partial \phi_{n,m,t}}{\partial W_{nt}} = - \frac{\bar{\epsilon}_n \sum_{s=0}^{m-1} \frac{Z_n^s}{(1+r)^s}}{2M\mu_n \left[u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]}, \quad (\text{A.38})$$

$$\frac{\partial W_{nt}}{\partial \eta_{nt}} = \frac{2\beta M \mu_n [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)]}{1 - \frac{\beta [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z_n^s}{(1+r)^s}}{M \left[u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]}, \quad (\text{A.39})$$

respectively, where

$$Z_n \equiv \frac{\frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}{u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}.$$

Since $W_t = \sum_{n' \in \mathcal{N}} W_{n't}$ and the shock η_{nt} does not affect $W_{n't}$ for $n' \neq n$, $\frac{\partial W_t}{\partial \eta_{nt}}$ is also given by (A.39).

When opportunities are symmetric, $\mathcal{N} = \{1, \dots, N\}$ and

$$Z_n = Z = \frac{\frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}{u\bar{\epsilon} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}$$

for all $n = 1, \dots, N$. In the case of segmentation, (A.39) implies that

$$\frac{\partial W_{nt}}{\partial \eta_{nt}} = \frac{\partial W_t}{\partial \eta_{nt}} = \frac{2\beta M \mu [u\bar{\epsilon} - (f')^{-1}(\Pi)]}{1 - \frac{\beta [u\bar{\epsilon} - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}}{M \left[u\bar{\epsilon} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]} \equiv D_W. \quad (\text{A.40})$$

Therefore, (33) implies that the variance of the arbitrageurs' total wealth is $ND_W^2 \text{Var}(\eta)$, and of the wealth of arbitrageurs in market n is $D_W^2 \text{Var}(\eta)$. Eqs. (A.38) and (A.39) imply that

$$\begin{aligned} \frac{\partial \phi_{n,m,t}}{\partial \eta_{nt}} &= \frac{\partial \phi_{n,m,t}}{\partial W_{nt}} \frac{\partial W_{nt}}{\partial \eta_{nt}} \\ &= - \frac{\beta \bar{\epsilon} [u\bar{\epsilon} - (f')^{-1}(\Pi)] \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}}{u\bar{\epsilon} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} - \frac{\beta [u\bar{\epsilon} - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}}{M}} \\ &\equiv D_\phi \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}. \end{aligned}$$

Since the shock η_{nt} does not affect $\phi_{n',m,t}$ for $n' \neq n$, its effect on the aggregate spread $\phi_t \equiv \sum_{n'=1}^N x_{n'} \sum_{m=1}^{M-1} \phi_{n',m,t}$ is

$$\hat{D}_\phi \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s},$$

where

$$\hat{D}_\phi \equiv \frac{\mu}{\bar{\epsilon}} [u\bar{\epsilon} - (f')^{-1}(\Pi)] D_\phi.$$

Therefore, the variance of the aggregate spread is $N \hat{D}_\phi^2 \left(\sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s} \right)^2 \text{Var}(\eta)$, and of $\phi_{n,m,t}$ is $D_\phi^2 \left(\sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s} \right)^2 \text{Var}(\eta)$.

In the case of integration, (A.36) implies that $\frac{\partial W_t}{\partial \eta_{nt}} = D_W$. This is the same as under segmentation, and so is the variance of the arbitrageurs' total wealth. Since W_{nt} is equal to $\frac{W_t}{N}$ under integration, its variance is $\frac{1}{N} D_W^2 \text{Var}(\eta)$, lower than under segmentation. Eqs. (A.35) and (A.36) imply that

$$\frac{\partial \phi_{n',m,t}}{\partial \eta_{nt}} = \frac{\partial \phi_{n',m,t}}{\partial W_t} \frac{\partial W_t}{\partial \eta_{nt}} = \frac{1}{N} D_\phi \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}.$$

Since this effect is independent of n' , the effect of η_{nt} on the aggregate spread ϕ_t is

$$\hat{D}_\phi \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}.$$

This is the same as under segmentation, and so is the variance of the aggregate spread. The variance of $\phi_{n,m,t}$ is $\frac{1}{N} D_\phi^2 \left(\sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s} \right)^2 \text{Var}(\eta)$, lower than under segmentation. \blacksquare

Before proving Proposition 8, we prove the following lemma:

Lemma 2. *Consider N positive scalars (a_1, \dots, a_N) that are not all equal, and an increasing and*

differentiable function $F(a)$. If the function $a^2F'(a)$ is increasing, then

$$\sum_{n=1}^N a_n^2 F(a_n) > \left(\sum_{n=1}^N a_n^2 \right) F \left(\frac{\sum_{n=1}^N a_n}{N} \right). \quad (\text{A.41})$$

Proof: Setting $\bar{a} \equiv \frac{\sum_{n=1}^N a_n}{N}$, we can write the difference between the left- and right-hand side of (A.41) as

$$\begin{aligned} \sum_{n=1}^N a_n^2 [F(a_n) - F(\bar{a})] &= \sum_{n=1}^N a_n^2 \int_{\bar{a}}^{a_n} F'(b) db \\ &> \sum_{n=1}^N \int_{\bar{a}}^{a_n} b^2 F'(b) db \\ &> \sum_{n=1}^N \int_{\bar{a}}^{a_n} \bar{a}^2 F'(\bar{a}) db \\ &= \bar{a}^2 F'(\bar{a}) \sum_{n=1}^N \int_{\bar{a}}^{a_n} db \\ &= \bar{a}^2 F'(\bar{a}) \sum_{n=1}^N (a_n - \bar{a}) = 0, \end{aligned}$$

where the second step follows because $F(a)$ is increasing, and the third because $a^2F'(a)$ is increasing. Therefore, (A.41) holds. \blacksquare

Proof of Proposition 8: Eqs. (33) and (A.36) imply that the variance of the arbitrageurs' total wealth under integration is

$$\sum_{n \in \mathcal{N}} \left(\frac{2\beta M \mu_n [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)]}{1 - \frac{\beta \sum_{n'' \in \mathcal{N}} \mu_{n''} [u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}}}{M \sum_{n'' \in \mathcal{N}} \mu_{n''} \left[u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]} \right)^2 \text{Var}(\eta). \quad (\text{A.42})$$

Likewise, (33) and (A.39) imply that the variance of the arbitrageurs' total wealth under segmen-

tation is

$$\sum_{n \in \mathcal{N}} \left(\frac{2\beta M \mu_n [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)]}{1 - \frac{\beta [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z_n^s}{(1+r)^s}}}{M \left[u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]} \right)^2 \text{Var}(\eta). \quad (\text{A.43})$$

Eq. (A.42) and (A.43) imply that the variance under segmentation exceeds that under integration if and only if

$$\sum_{n \in \mathcal{N}} \mu_n^2 a_n^2 G(a_n)^2 > \left(\sum_{n \in \mathcal{N}} \mu_n^2 a_n^2 \right) G \left(\frac{\sum_{n \in \mathcal{N}} \mu_n a_n}{\sum_{n \in \mathcal{N}} \mu_n} \right)^2, \quad (\text{A.44})$$

where

$$a_n \equiv u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) > 0, \quad (\text{A.45})$$

$$G(a) \equiv \frac{1}{1 - H(a)}, \quad (\text{A.46})$$

$$H(a) \equiv \frac{\beta a \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z(a)^s}{(1+r)^s}}{M \left[a + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]}, \quad (\text{A.47})$$

$$Z(a) \equiv \frac{\frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}{a + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}. \quad (\text{A.48})$$

When $\mu_n = \mu$ for all $n = 1, \dots, N$, (A.44) becomes equivalent to (A.41). Hence, Lemma 2 implies that (A.44) holds if $F(a) \equiv G(a)^2$ and $a^2 F'(a)$ are increasing. Since

$$F'(x) = 2G(a)G'(a) = \frac{2H'(a)}{[1 - H(a)]^3},$$

both properties will follow if we show that $H(a) < 1$ and that $H(a)$ and $a^2 H'(a)$ are increasing. (Note that $H(a) < 1$ ensures that small shocks have small effects despite the amplification, and

validates our linearization around the steady state.) We can write $H(a)$ as

$$\begin{aligned} H(a) &= \frac{\beta [1 - Z(a)]}{M} \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z(a)^s}{(1+r)^s} \\ &= \frac{\beta H_1(a)}{M} \sum_{m=1}^{M-1} \left[1 - \frac{Z(a)^m}{(1+r)^m} \right] \end{aligned} \quad (\text{A.49})$$

$$= \beta H_1(a) \left[\frac{M-1}{M} - \frac{\frac{Z(a)}{1+r} - \frac{Z(a)^M}{(1+r)^M}}{M \left[1 - \frac{Z(a)}{1+r} \right]} \right], \quad (\text{A.50})$$

where

$$H_1(a) \equiv \frac{1 - Z(a)}{1 - \frac{Z(a)}{1+r}}.$$

Since $Z(a) \in (0, 1)$, $H_1(a) \in (0, 1)$, and (A.50) implies that $H(a) < 1$. Since, in addition, $Z(a)$ is decreasing and $H_1(a)$ is increasing, (A.49) implies that $H(a)$ is increasing. Eq. (A.49) implies that

$$\begin{aligned} a^2 H'(a) &= -\frac{a^2 \beta Z'(a)}{M(1+r)} \left\{ \frac{r}{\left[1 - \frac{Z(a)}{1+r} \right]^2} \sum_{m=1}^{M-1} \left[1 - \frac{Z(a)^m}{(1+r)^m} \right] + H_1(a) \sum_{m=1}^{M-1} \frac{m Z(a)^{m-1}}{(1+r)^{m-1}} \right\} \\ &= \frac{\beta [1 - Z(a)]^2}{M(1+r)} \left\{ \frac{r}{\left[1 - \frac{Z(a)}{1+r} \right]^2} \sum_{m=1}^{M-1} \left[1 - \frac{Z(a)^m}{(1+r)^m} \right] + H_1(a) \sum_{m=1}^{M-1} \frac{m Z(a)^{m-1}}{(1+r)^{m-1}} \right\} \\ &= \frac{\beta [1 - Z(a)]^2}{M(1+r)} \left\{ \frac{r}{\left[1 - \frac{Z(a)}{1+r} \right]^2} \sum_{m=1}^{M-1} \left[1 - \frac{Z(a)^m}{(1+r)^m} \right] + \frac{H_1(a) K[Z(a)]}{\left[1 - \frac{Z(a)}{1+r} \right]^2} \right\} \\ &= \frac{\beta H_1(a)^2}{M(1+r)} \left\{ r \sum_{m=1}^{M-1} \left[1 - \frac{Z(a)^m}{(1+r)^m} \right] + H_1(a) K[Z(a)] \right\}, \end{aligned} \quad (\text{A.51})$$

where

$$K(z) \equiv 1 - \frac{z^M}{(1+r)^M} - \frac{Mz^{M-1}}{(1+r)^{M-1}} \left(1 - \frac{z}{1+r} \right) = \left(1 - \frac{z}{1+r} \right)^2 \sum_{m=1}^{M-1} \frac{mz^{m-1}}{(1+r)^{m-1}}.$$

Since $Z(a) \in (0, 1)$ is decreasing, $H_1(a)$ is increasing, and $K(z)$ is decreasing, $a^2 H'(a)$ is increasing.

Therefore, if $\mu_n = \mu$ for all $n = 1, \dots, N$, then the variance of the arbitrageurs' total wealth under segmentation exceeds that under integration. To construct an example where the reverse holds if μ_n differs across n , we assume that $N = 2$, β is high enough so that $\mathcal{N} = \{1, 2\}$, and μ_2 is close to zero. Omitting terms in μ_2^2 and smaller, we can write (A.44) as

$$\begin{aligned} \mu_1^2 a_1^2 G(a_1)^2 &> \mu_1^2 a_1^2 G\left(\frac{\mu_1 a_1 + \mu_2 a_2}{\mu_1 + \mu_2}\right)^2 \\ \Leftrightarrow G(a_1) &> G\left(\frac{\mu_1 a_1 + \mu_2 a_2}{\mu_1 + \mu_2}\right) \\ \Leftrightarrow a_1 &> \frac{\mu_1 a_1 + \mu_2 a_2}{\mu_1 + \mu_2}, \end{aligned}$$

which is violated if $a_2 > a_1$.

We next determine the variance of the aggregate spread ϕ_t . Under integration,

$$\begin{aligned} \frac{\partial \phi_t}{\partial \eta_{nt}} &= \sum_{n' \in \mathcal{N}} x_{n'} \sum_{m=1}^{M-1} \frac{\partial \phi_{n',m,t}}{\partial \eta_{nt}} \\ &= \sum_{n' \in \mathcal{N}} x_{n'} \sum_{m=1}^{M-1} \frac{\partial \phi_{n',m,t}}{\partial W_t} \frac{\partial W_t}{\partial \eta_{nt}} \\ &= - \frac{\mu_n [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)] \frac{\beta \sum_{n'' \in \mathcal{N}} \mu_{n''} [u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}}{\sum_{n'' \in \mathcal{N}} \mu_{n''} \left[u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]}, \quad (\text{A.52}) \\ &\quad 1 - \frac{\beta \sum_{n'' \in \mathcal{N}} \mu_{n''} [u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}}{M \sum_{n'' \in \mathcal{N}} \mu_{n''} \left[u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]} \end{aligned}$$

where the last step follows from (A.22), (A.35) and (A.36). Under segmentation, the shock η_{nt} does not affect $\phi_{n',m,t}$ for $n' \neq n$. Therefore,

$$\begin{aligned} \frac{\partial \phi_t}{\partial \eta_{nt}} &= x_n \sum_{m=1}^{M-1} \frac{\partial \phi_{n,m,t}}{\partial \eta_{nt}} \\ &= x_n \sum_{m=1}^{M-1} \frac{\partial \phi_{n,m,t}}{\partial W_{nt}} \frac{\partial W_{nt}}{\partial \eta_{nt}} \\ &= - \frac{\mu_n [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)] \frac{\beta [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z_n^s}{(1+r)^s}}{u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}, \quad (\text{A.53}) \\ &\quad 1 - \frac{\beta [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z_n^s}{(1+r)^s}}{M \left[u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]} \end{aligned}$$

where the last step follows from (A.22), (A.38) and (A.39). Eqs. (A.52) and (A.53) imply that the variance of the aggregate spread under segmentation exceeds that under integration if and only if (A.44) holds with

$$G(a) \equiv \frac{H(a)}{1 - H(a)} \quad (\text{A.54})$$

and a_n , $H(a)$ and $Z(a)$ defined by (A.45), (A.47) and (A.48), respectively. Since the derivative of $F(a) \equiv G(a)^2$ is

$$F'(a) = 2G(a)G'(a) = \frac{2H(a)H'(a)}{[1 - H(a)]^3},$$

and $H(a)$ and $a^2H'(a)$ are increasing, $F(a)$ and $a^2F'(a)$ are also increasing. Therefore, Lemma 2 implies that if $\mu_n = \mu$ for all $n = 1, \dots, N$, then the variance of the aggregate spread under segmentation exceeds that under integration. The example constructed earlier in the proof implies that the reverse can hold if μ_n differs across n .

We finally determine the variance of individual spreads. Eqs. (A.35) and (A.36) imply that under integration,

$$\frac{\partial \phi_{n,m,t}}{\partial \eta_{n't}} = \frac{\partial \phi_{n,m,t}}{\partial W_t} \frac{\partial W_t}{\partial \eta_{n't}} = - \frac{\frac{\beta \mu_{n'} [u_{n'} \bar{\epsilon}_{n'} - (f')^{-1}(\Pi)] \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}}{\sum_{n'' \in \mathcal{N}} \mu_{n''} \left[u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]}}{1 - \frac{\beta \sum_{n'' \in \mathcal{N}} \mu_{n''} [u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}}{M \sum_{n'' \in \mathcal{N}} \mu_{n''} \left[u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]}},$$

and hence

$$\text{Var} \left(\sum_{m=1}^{M-1} \phi_{n,m,t} \right) = \sum_{n'=1}^N \left(\frac{\frac{\beta \mu_{n'} [u_{n'} \bar{\epsilon}_{n'} - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}}{\sum_{n'' \in \mathcal{N}} \mu_{n''} \left[u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]}}{1 - \frac{\beta \sum_{n'' \in \mathcal{N}} \mu_{n''} [u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z^s}{(1+r)^s}}{M \sum_{n'' \in \mathcal{N}} \mu_{n''} \left[u_{n''} \bar{\epsilon}_{n''} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]}} \right)^2 \text{Var}(\eta). \quad (\text{A.55})$$

Eqs. (A.38) and (A.39) imply that under segmentation,

$$\frac{\partial \phi_{n,m,t}}{\partial \eta_{nt}} = \frac{\partial \phi_{n,m,t}}{\partial W_{nt}} \frac{\partial W_{nt}}{\partial \eta_{nt}} = - \frac{\frac{\beta[u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)] \sum_{s=0}^{m-1} \frac{Z_n^s}{(1+r)^s}}{u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}}{1 - \frac{\beta[u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z_n^s}{(1+r)^s}}{M \left[u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]}}.$$

Since, in addition, the shock $\eta_{n't}$ does not affect $\phi_{n,m,t}$ for $n' \neq n$,

$$\text{Var} \left(\sum_{m=1}^{M-1} \phi_{n,m,t} \right) = \left(\frac{\frac{\beta[u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z_n^s}{(1+r)^s}}{u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}}{1 - \frac{\beta[u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)] \sum_{m=1}^{M-1} \sum_{s=0}^{m-1} \frac{Z_n^s}{(1+r)^s}}{M \left[u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]} \right]}} \right)^2 \text{Var}(\eta). \quad (\text{A.56})$$

Eqs. (A.55) and (A.56) imply that the variance of $\sum_{m=1}^{M-1} \phi_{n,m,t}$ under segmentation exceeds that under integration if and only if

$$G(a_n)^2 > \frac{\sum_{n' \in \mathcal{N}} \mu_{n'}^2 a_{n'}^2}{\left(\sum_{n' \in \mathcal{N}} \mu_{n'} a_{n'} \right)^2} G \left(\frac{\sum_{n' \in \mathcal{N}} \mu_{n'} a_{n'}}{\sum_{n' \in \mathcal{N}} \mu_{n'}} \right)^2, \quad (\text{A.57})$$

where a_n , $G(a)$, $H(a)$ and $Z(a)$ are defined by (A.45), (A.54), (A.47) and (A.48), respectively.

Since $G(a)$ is increasing and $\frac{\sum_{n' \in \mathcal{N}} \mu_{n'}^2 a_{n'}^2}{\left(\sum_{n' \in \mathcal{N}} \mu_{n'} a_{n'} \right)^2} < 1$, (A.57) holds for the n that maximizes a_n . On the

other hand, (A.57) fails to hold if a_n is close to zero and $a_{n'}$ is not for $n' \neq n$. The comparisons of the variance of $\sum_{m=1}^{M-1} \phi_{n,m,t}$ under integration and segmentation extend to the variance of $\phi_{n,m,t}$ for at least one m . This is because $\phi_{n,m,t}$ are perfectly correlated for $m = 1, \dots, M-1$, under both integration and segmentation. ■

B General Contracts

B.1 Contracts and Equilibrium

A contract ω that arbitrageurs can trade with i -investors in period t is characterized by (i) payments $\pi_{\omega,t'}$ that the seller of the contract must make to the buyer in periods $t' > t$, (ii) a price $q_{\omega,t}$ that the seller of the contract receives from the buyer in period t , and (iii) collateral that the seller of the contract must post with the buyer. The payments $\pi_{\omega,t'}$ can depend on information available in

all markets including market i . We assume that payments are non-negative and are not all equal to zero. No-arbitrage then implies that the price $q_{\omega,t}$ must be positive. Collateral must be in the form of cash or other contracts. A contract ω can be traded in any period $t \in \{\underline{t}_\omega, \dots, \bar{t}_\omega - 1\}$, where \underline{t}_ω occurs before the first positive payment and \bar{t}_ω is when the last positive payment is made. The period \bar{t}_ω can be infinite, and if it is finite we set $q_{\omega,\bar{t}_\omega} = 0$. We denote by $\Omega_{i,t}$ the set of contracts that can be traded in market i and period t .

To specify how contracts can be collateralized using other contracts, we define contracts recursively. Contracts of level 1 are collateralized by the riskless asset. Contracts of level $n + 1$ are collateralized by the riskless asset and by a finite number of contracts of levels 1 up to n . For a contract $\omega \in \Omega_{i,t}$ and period t , we denote by $\psi_{\omega,t} \geq 0$ the units of the riskless asset and by $\psi_{\omega,\omega',t} \geq 0$ the units of a lower-level contract $\omega' \in \Omega_{i,t}$ that are required as collateral. We also denote by $\ell(\omega,t)$ the level of the contract. The collateral amounts $\psi_{\omega,t}$ and $\psi_{\omega,\omega',t}$ and the level $\ell(\omega,t)$ can depend on information available in all markets including market i .

We denote by $y_{\omega,t}$ the position of i -investors and $x_{\omega,t}$ the position of arbitrageurs in a contract $\omega \in \Omega_{i,t}$ and period t . Because the number of contracts is infinite, there is an infinite set of positions. We assume that only a finite number of the positions are non-zero.

The collateral that short positions require must be covered by long positions. Suppose, for example, that arbitrageurs have a short position in a contract $\omega \in \Omega_{i,t}$, which requires contract $\omega' \in \Omega_{i,t}$ as collateral. This does not necessarily imply that arbitrageurs must have an overall long position in contract ω' : they must buy contract ω' to post as collateral for the short position in contract ω , but they could undertake an additional transaction in contract ω' to establish an overall short position in that contract. We decompose the position $x_{\omega',t}$ in contract ω' into

$$x_{\omega',t} = x_{\omega',t}^c + \hat{x}_{\omega',t},$$

where $x_{\omega',t}^c \geq 0$ is collateral set aside for short positions in higher-level contracts $\omega \in \Omega_{i,t}$, and $\hat{x}_{\omega',t}$ is the remainder of the position, which can be negative. The collateral $x_{\omega',t}^c$ must satisfy

$$x_{\omega',t}^c = \sum_{\substack{\omega \in \Omega_{i,t} \\ \ell(\omega,t) > \ell(\omega',t) \text{ and } \hat{x}_{\omega,t} < 0}} (-\hat{x}_{\omega,t}) \psi_{\omega,\omega',t}. \quad (\text{B.1})$$

The collateral $v_{i,t}$ in the riskless asset required for contracts in market i must likewise satisfy

$$v_{i,t} = \sum_{\substack{\omega \in \Omega_{i,t} \\ \hat{x}_{\omega,t} < 0}} (-\hat{x}_{\omega,t}) \psi_{\omega,t}. \quad (\text{B.2})$$

The wealth that arbitrageurs “tie up” in market i is $\sum_{\omega \in \Omega_{i,t}} x_{\omega,t} q_{\omega,t} + v_{i,t}$, the value of their positions in the contracts traded in market i and of the riskless collateral. The financial constraint of arbitrageurs requires that the sum of that quantity across markets does not exceed the arbitrageurs’ total wealth W_t :

$$W_t \geq \sum_{i \in \mathcal{I}} \left(\sum_{\omega \in \Omega_{i,t}} x_{\omega,t} q_{\omega,t} + v_{i,t} \right). \quad (\text{B.3})$$

As in Section 2.3.2, we assume that i -investors have enough wealth so that their financial constraint is never binding.

Investors and arbitrageurs can default on their short positions in the contracts. Defaulting on a unit short position in a contract $\omega \in \Omega_{i,t}$ in period $t + 1$ raises the wealth of an agent by $\pi_{\omega,t+1} + q_{\omega,t+1}$ since the agent does not make the payment $\pi_{\omega,t+1}$ and no longer has the liability $q_{\omega,t+1}$. At the same time, the agent loses the collateral associated to the position. Default is costlier to the agent than no default if

$$\pi_{\omega,t+1} + q_{\omega,t+1} \leq (1+r)\psi_{\omega,t} + \sum_{\substack{\omega' \in \Omega_{i,t} \\ \ell(\omega,t) > \ell(\omega',t)}} \psi_{\omega,\omega',t} (\pi_{\omega',t+1} + q_{\omega',t+1}), \quad (\text{B.4})$$

i.e., the amount saved by not making the payment is smaller than the value of the collateral seized. Without loss of generality, we can assume that there is no default. This is because we can replace a contract ω that involves default by one with the same collateral and with required payments equal to the actual payments (including the effects of default) under ω .

Under no default, the budget constraint of an i -investor is

$$w_{i,t+1} = \sum_{\omega \in \Omega_{i,t}} y_{\omega,t} (\pi_{\omega,t+1} + q_{\omega,t+1}) + (1+r) \left(w_{i,t} - \sum_{\omega \in \Omega_{i,t}} y_{\omega,t} q_{\omega,t} \right) + u_{i,t} \epsilon_{i,t+1} - c_{i,t+1}, \quad (\text{B.5})$$

and of an arbitrageur is

$$W_{t+1} = \sum_{i \in \mathcal{I}} \sum_{\omega \in \Omega_{i,t}} x_{\omega,t} (\pi_{\omega,t+1} + q_{\omega,t+1}) + (1+r) \left(W_t - \sum_{i \in \mathcal{I}} \sum_{\omega \in \Omega_{i,t}} x_{\omega,t} q_{\omega,t} \right) - c_{t+1}. \quad (\text{B.6})$$

Eqs. (B.5) and (B.6) are counterparts of (13) and (17), with the positions in the contracts replacing those in the risky assets.

Definition 3. *A competitive equilibrium with no default consists of prices $q_{\omega,t}$ for all contracts $\omega \in \Omega_{i,t}$, and positions in the contracts $y_{\omega,t}$ for the i -investors and $x_{\omega,t}$ for the arbitrageurs, such that (B.4) holds, positions are optimal given prices, and the markets for all contracts clear:*

$$\mu_i y_{\omega,t} + x_{\omega,t} = 0. \quad (\text{B.7})$$

B.2 Binomial Payoffs

We next assume that the variables $\epsilon_{i,t}$ have a binomial distribution. Given symmetry, this amounts to assuming that the variables $\frac{\epsilon_{i,t}}{\bar{\epsilon}_i}$ take the values 1 and -1 with probabilities one-half.

Proposition B.1. *There exists a competitive equilibrium with no default such that the dynamics of wealth of i -investors and arbitrageurs are as in Section 3.4 and the prices $q_{\omega,t}$ of all contracts $\omega \in \Omega_{i,t}$ are given by*

$$q_{\omega,t} = \frac{\exp(-\mathcal{Z}_{i,t}) E_t(\pi_{\omega,t+1} + q_{\omega,t+1} | \epsilon_{i,t+1} = \bar{\epsilon}_i) + \exp(\mathcal{Z}_{i,t}) E_t(\pi_{\omega,t+1} + q_{\omega,t+1} | \epsilon_{i,t+1} = -\bar{\epsilon}_i)}{(1+r) [\exp(-\mathcal{Z}_{i,t}) + \exp(\mathcal{Z}_{i,t})]}, \quad (\text{B.8})$$

where

$$\mathcal{Z}_{i,t} \equiv \frac{\alpha A}{\alpha + A} (y_{i,t} + u_{i,t}) \bar{\epsilon}_i$$

and $y_{i,t}$ is as in Section 3.4.

Proof of Proposition B.1: We first study optimization by i -investors. We proceed as in the proof of Proposition 1, conjecture the value function (14) with $A = r\alpha$ and $F_{i,t}$ given by (A.9), and use the budget constraint (B.5) instead of (13). Optimal consumption is given by

$$c_{i,t+1} = \frac{A \left[(1+r)w_{i,t} + \sum_{\omega \in \Omega_{i,t}} y_{\omega,t} [\pi_{\omega,t+1} + q_{\omega,t+1} - (1+r)q_{\omega,t}] + u_{i,t} \epsilon_{i,t+1} \right] + F_{i,t+1} + \log\left(\frac{\alpha}{A}\right)}{\alpha + A},$$

(B.9)

which is the counterpart of (A.4). Optimal positions in the contracts solve

$$\max_{y_{\omega,t}} \mathbb{E}_t \left\{ - \exp \left(- \frac{\alpha A}{\alpha + A} \left[(1+r)w_{i,t} + \sum_{\omega \in \Omega_{i,t}} y_{\omega,t} [\pi_{\omega,t+1} + q_{\omega,t+1} - (1+r)q_{\omega,t}] + u_{i,t} \epsilon_{i,t+1} \right] \right) \right\}, \quad (\text{B.10})$$

which is the counterpart of (A.5) after omitting terms that are known in period t . The first-order condition with respect to $y_{\omega,t}$ is

$$\mathbb{E}_t \left\{ \left[\pi_{\omega,t+1} + q_{\omega,t+1} - (1+r)q_{\omega,t} \right] \times \exp \left(- \frac{\alpha A}{\alpha + A} \left[\sum_{\omega \in \Omega_{i,t}} y_{\omega,t} [\pi_{\omega,t+1} + q_{\omega,t+1} - (1+r)q_{\omega,t}] + u_{i,t} \epsilon_{i,t+1} \right] \right) \right\} = 0. \quad (\text{B.11})$$

Eq. (B.8) that characterizes equilibrium prices can be written as

$$\mathbb{E}_t \left\{ \left[\pi_{\omega,t+1} + q_{\omega,t+1} - (1+r)q_{\omega,t} \right] \exp \left(- \frac{\alpha A}{\alpha + A} (y_{i,t} + u_{i,t}) \epsilon_{i,t+1} \right) \right\} = 0. \quad (\text{B.12})$$

Eqs. (B.11) and (B.12) imply that if positions in the contracts satisfy

$$\sum_{\omega \in \Omega_{i,t}} y_{\omega,t} [\pi_{\omega,t+1} + q_{\omega,t+1} - (1+r)q_{\omega,t}] = y_{i,t} \epsilon_{i,t+1} + \mathcal{G}_t, \quad (\text{B.13})$$

where \mathcal{G}_t is known in period t , then they are optimal because the first-order condition (B.11) is met. Positions satisfying (B.13) are not unique, and we present one implementation at the end of this proof. Eq. (B.13) implies that the dynamics of wealth of i -investors are the same as in Section 3.4. Indeed, multiplying (B.12) by $y_{\omega,t}$ and summing across $\omega \in \Omega_{i,t}$, we find

$$\mathbb{E}_t \left\{ \sum_{\omega \in \Omega_{i,t}} y_{\omega,t} [\pi_{\omega,t+1} + q_{\omega,t+1} - (1+r)q_{\omega,t}] \exp \left(- \frac{\alpha A}{\alpha + A} (y_{i,t} + u_{i,t}) \epsilon_{i,t+1} \right) \right\} = 0. \quad (\text{B.14})$$

Moreover, the maximization in (A.5) implies that

$$E_t \left\{ y_{i,t} (\Phi_{i,t} + \epsilon_{i,t+1}) \exp \left(-\frac{\alpha A}{\alpha + A} (y_{i,t} + u_{i,t}) \epsilon_{i,t+1} \right) \right\} = 0. \quad (\text{B.15})$$

Substituting $\sum_{\omega \in \Omega_{i,t}} y_{\omega,t} [\pi_{\omega,t+1} + q_{\omega,t+1} - (1+r)q_{\omega,t}]$ from (B.13) into (B.14), and comparing with (B.15), we find $\mathcal{G}_t = \Phi_{i,t}$. Substituting $\mathcal{G}_t = \Phi_{i,t}$ into (B.13), we find that budget constraint (B.5) of i -investors becomes identical to the budget constraint (13) in Section 3.4. Since the dynamics of the wealth of i -investors are the same as in Section 3.4, the conjectured value function (14) satisfies the Bellman equation.

We next study optimization by arbitrageurs. We proceed in two steps: in Step 1 we show that the dynamics of arbitrageur wealth are deterministic, and in Step 2 that they are as in Section 3.4.

Step 1: To show deterministic dynamics, we show that if arbitrageurs choose in period t a portfolio of contracts whose aggregate payoff in period $t+1$ is risky, then there exists another portfolio that is riskless and has a return that is at least as high as the expected return of the risky portfolio. We construct a “dominant” riskless portfolio for each market i separately, and then aggregate across markets. From the budget constraint (B.6), the (excess) return that arbitrageurs earn on their portfolio of contracts in market i is

$$\sum_{\omega \in \Omega_{i,t}} x_{\omega,t} [\pi_{\omega,t+1} + q_{\omega,t+1} - (1+r)q_{\omega,t}]. \quad (\text{B.16})$$

Consider first a market i without an endowment shock, i.e., $u_{i,t} = 0$. Since $\mathcal{Z}_{i,t} = 0$, (B.8) implies that

$$q_{\omega,t} = \frac{E_t(\pi_{\omega,t+1} + q_{\omega,t+1} | \epsilon_{i,t+1} = \bar{\epsilon}_i) + E_t(\pi_{\omega,t+1} + q_{\omega,t+1} | \epsilon_{i,t+1} = -\bar{\epsilon}_i)}{2(1+r)} = \frac{E_t(\pi_{\omega,t+1} + q_{\omega,t+1})}{1+r},$$

and hence the expected return in (B.16) is zero. A dominant riskless portfolio is one with zero positions.

Consider next a market i with an endowment shock. If the expected return in (B.16) is non-positive, then a dominant riskless portfolio is one with zero positions. If the expected return in (B.16) is positive, then we will construct a dominant riskless portfolio that involves positions in markets i and $-i$. As an intermediate step in this construction, we show that the original risky portfolio has the same expected return and ties up the same amount of arbitrageur wealth as a unit long position in a single contract $\hat{\omega}_i$ that is traded in market i and has binary payoffs. The payoffs

of $\hat{\omega}_i$ are

$$E_t \left[\sum_{\omega \in \Omega_{i,t}} x_{\omega,t} (\pi_{\omega,t+1} + q_{\omega,t+1}) \middle| \epsilon_{i,t+1} = \bar{\epsilon}_i \right] + (1+r)v_{i,t} \equiv \bar{Q}_{i,t+1} + (1+r)v_{i,t},$$

$$E_t \left[\sum_{\omega \in \Omega_{i,t}} x_{\omega,t} (\pi_{\omega,t+1} + q_{\omega,t+1}) \middle| \epsilon_{i,t+1} = -\bar{\epsilon}_i \right] + (1+r)v_{i,t} \equiv \underline{Q}_{i,t+1} + (1+r)v_{i,t},$$

in period $t+1$ and states $\epsilon_{i,t+1} = \bar{\epsilon}_i$ and $\epsilon_{i,t+1} = -\bar{\epsilon}_i$, respectively, and zero afterwards. The price of $\hat{\omega}_i$ in period t is

$$\begin{aligned} & \frac{\exp(-\mathcal{Z}_{i,t}) [\bar{Q}_{i,t+1} + (1+r)v_{i,t}] + \exp(\mathcal{Z}_{i,t}) [\underline{Q}_{i,t+1} + (1+r)v_{i,t}]}{(1+r) [\exp(-\mathcal{Z}_{i,t}) + \exp(\mathcal{Z}_{i,t})]} \\ &= \sum_{\omega \in \Omega_{i,t}} x_{\omega,t} \frac{\exp(-\mathcal{Z}_{i,t}) E_t (\pi_{\omega,t+1} + q_{\omega,t+1} | \epsilon_{i,t+1} = \bar{\epsilon}_i) + \exp(\mathcal{Z}_{i,t}) E_t (\pi_{\omega,t+1} + q_{\omega,t+1} | \epsilon_{i,t+1} = -\bar{\epsilon}_i)}{(1+r) [\exp(-\mathcal{Z}_{i,t}) + \exp(\mathcal{Z}_{i,t})]} + v_{i,t} \\ &= \sum_{\omega \in \Omega_{i,t}} x_{\omega,t} q_{\omega,t} + v_{i,t} \\ &\equiv Q_{i,t} + v_{i,t}, \end{aligned}$$

where the first step follows from (B.8), the second by using the definitions of $(\bar{Q}_{i,t+1}, \underline{Q}_{i,t+1})$ and rearranging terms, and the third from (B.8). Therefore, the wealth $Q_{i,t} + v_{i,t}$ that arbitrageurs tie up in market i is the same as under the original risky portfolio. The expected return from buying $\hat{\omega}_i$ is

$$\begin{aligned} & \frac{1}{2} [\bar{Q}_{i,t+1} + (1+r)v_{i,t}] + \frac{1}{2} [\underline{Q}_{i,t+1} + (1+r)v_{i,t}] - (1+r)(Q_{i,t} + v_{i,t}) \\ &= \frac{1}{2} (\bar{Q}_{i,t+1} + \underline{Q}_{i,t+1}) - (1+r)Q_{i,t} \\ &= \sum_{\omega \in \Omega_{i,t}} x_{\omega,t} \left[\frac{1}{2} [E_t (\pi_{\omega,t+1} + q_{\omega,t+1} | \epsilon_{i,t+1} = \bar{\epsilon}_i) + E_t (\pi_{\omega,t+1} + q_{\omega,t+1} | \epsilon_{i,t+1} = -\bar{\epsilon}_i)] - (1+r)q_{\omega,t} \right] \\ &= \sum_{\omega \in \Omega_{i,t}} x_{\omega,t} E_t [\pi_{\omega,t+1} + q_{\omega,t+1} - (1+r)q_{\omega,t}], \end{aligned}$$

where the third step follows from the definitions of $(Q_{i,t}, \bar{Q}_{i,t+1}, \underline{Q}_{i,t+1})$. The expected return from buying $\hat{\omega}_i$ is thus the same as under the original risky portfolio. To complete the analysis of $\hat{\omega}_i$, we

must show that it is a proper contract in the sense that its payoffs are non-negative. Multiplying (B.4) by $-\hat{x}_{\omega,t}$ for those $\omega \in \Omega_{i,t}$ for which $\hat{x}_{\omega,t} < 0$, and summing across ω , we find

$$\begin{aligned}
& \sum_{\substack{\omega \in \Omega_{i,t} \\ \hat{x}_{\omega,t} < 0}} (-\hat{x}_{\omega,t})(\pi_{\omega,t+1} + q_{\omega,t+1}) \\
& \leq (1+r) \sum_{\substack{\omega \in \Omega_{i,t} \\ \hat{x}_{\omega,t} < 0}} (-\hat{x}_{\omega,t})\psi_{\omega,t} + \sum_{\substack{\omega \in \Omega_{i,t} \\ \hat{x}_{\omega,t} < 0}} \sum_{\substack{\omega' \in \Omega_{i,t} \\ \ell(\omega,t) > \ell(\omega',t)}} (-\hat{x}_{\omega,t})\psi_{\omega,\omega',t}(\pi_{\omega',t+1} + q_{\omega',t+1}) \\
& = (1+r)v_{i,t} + \sum_{\omega' \in \Omega_{i,t}} \sum_{\substack{\omega \in \Omega_{i,t} \\ \ell(\omega,t) > \ell(\omega',t) \text{ and } \hat{x}_{\omega,t} < 0}} (-\hat{x}_{\omega,t})\psi_{\omega,\omega',t}(\pi_{\omega',t+1} + q_{\omega',t+1}) \\
& = (1+r)v_{i,t} + \sum_{\omega' \in \Omega_{i,t}} x_{\omega',t}^c(\pi_{\omega',t+1} + q_{\omega',t+1}), \tag{B.17}
\end{aligned}$$

where the second step follows from (B.1) and the third from (B.2). Eq. (B.17) implies that

$$\begin{aligned}
& \sum_{\omega \in \Omega_{i,t}} (-\hat{x}_{\omega,t})(\pi_{\omega,t+1} + q_{\omega,t+1}) \leq (1+r)v_{i,t} + \sum_{\omega \in \Omega_{i,t}} x_{\omega,t}^c(\pi_{\omega,t+1} + q_{\omega,t+1}) \\
& \Rightarrow (1+r)v_{i,t} + \sum_{\omega \in \Omega_{i,t}} x_{\omega,t}(\pi_{\omega,t+1} + q_{\omega,t+1}) \geq 0. \tag{B.18}
\end{aligned}$$

Taking expectations in (B.18) conditional on $\epsilon_{i,t+1} = \bar{\epsilon}_i$ and $\epsilon_{i,t+1} = -\bar{\epsilon}_i$, we find that $\bar{Q}_{i,t+1} + (1+r)v_{i,t}$ and $\underline{Q}_{i,t+1} + (1+r)v_{i,t}$, respectively, are non-negative.

We next combine the unit long position in the contract $\hat{\omega}_i$ with a unit short position in a contract $\hat{\omega}_{-i}$ that is traded in market $-i$, has the same payoffs as $\hat{\omega}_i$, and is collateralized with $v_{-i,t}$ units of the riskless asset. The price of $\hat{\omega}_{-i}$ in period t is

$$\begin{aligned}
& \frac{\exp(-\mathcal{Z}_{-i,t}) [\bar{Q}_{i,t+1} + (1+r)v_{i,t}] + \exp(\mathcal{Z}_{-i,t}) [\underline{Q}_{i,t+1} + (1+r)v_{i,t}]}{(1+r) [\exp(-\mathcal{Z}_{-i,t}) + \exp(\mathcal{Z}_{-i,t})]} \\
& = \frac{\exp(\mathcal{Z}_{i,t})\bar{Q}_{i,t+1} + \exp(-\mathcal{Z}_{i,t})\underline{Q}_{i,t+1}}{(1+r) [\exp(\mathcal{Z}_{i,t}) + \exp(-\mathcal{Z}_{i,t})]} + v_{i,t} \\
& = \frac{\bar{Q}_{i,t+1} + \underline{Q}_{i,t+1}}{1+r} - Q_{i,t} + v_{i,t} \\
& \equiv Q_{-i,t} + v_{i,t}, \tag{B.19}
\end{aligned}$$

where the first step follows from (B.8), the second because $Z_{-i,t} = -Z_{i,t}$, and the third from the definition of $Q_{i,t}$. The wealth that arbitrageurs tie up in market $-i$ is

$$-(Q_{-i,t} + v_{i,t}) + v_{-i,t}$$

and is equal to the wealth that they tie up in market i if

$$v_{-i,t} = Q_{i,t} + Q_{-i,t} + 2v_{i,t}. \quad (\text{B.20})$$

The expected return from shorting $\hat{\omega}_{-i}$ is

$$\begin{aligned} & -\frac{1}{2} [\bar{Q}_{i,t+1} + (1+r)v_{i,t}] - \frac{1}{2} [\underline{Q}_{i,t+1} + (1+r)v_{i,t}] + (1+r)(Q_{-i,t} + v_{i,t}) \\ & = -\frac{1}{2} (\bar{Q}_{i,t+1} + \underline{Q}_{i,t+1}) + (1+r)Q_{-i,t} \\ & = \frac{1}{2} (\bar{Q}_{i,t+1} + \underline{Q}_{i,t+1}) - (1+r)Q_{i,t}, \end{aligned}$$

where the third step follows from the definition of $Q_{-i,t}$. Therefore, the expected return of the short position in $\hat{\omega}_{-i}$ is the same as that of the long position in $\hat{\omega}_i$. To complete the analysis of $\hat{\omega}_{-i}$, we must show that arbitrageurs do not default on their short position. Eq. (B.4) implies that default does not occur if

$$\begin{aligned} & \max\{\bar{Q}_{i,t+1}, \underline{Q}_{i,t+1}\} + (1+r)v_{i,t} \leq (1+r)v_{-i,t} \\ & \Leftrightarrow \max\{\bar{Q}_{i,t+1}, \underline{Q}_{i,t+1}\} \leq (1+r)(Q_{i,t} + Q_{-i,t} + v_{i,t}) \\ & \Leftrightarrow \max\{\bar{Q}_{i,t+1}, \underline{Q}_{i,t+1}\} \leq \bar{Q}_{i,t+1} + \underline{Q}_{i,t+1} + (1+r)v_{i,t}, \end{aligned} \quad (\text{B.21})$$

where the second step follows from (B.20) and the third from the definition of $Q_{-i,t}$. Eq. (B.21) holds because the payoffs $\bar{Q}_{i,t+1} + (1+r)v_{i,t}$ and $\underline{Q}_{i,t+1} + (1+r)v_{i,t}$ of $\hat{\omega}_i$ are non-negative.

The riskless portfolio that dominates the original risky portfolio in market i consists of a half-unit long position in $\hat{\omega}_i$ and a half-unit short position in $\hat{\omega}_{-i}$. Since a unit long position in $\hat{\omega}_i$ and a unit short position in $\hat{\omega}_{-i}$ each has the same expected return as the original risky portfolio, the combination of two half-unit positions also has the same expected return. The same applies to the amount of arbitrageur wealth that is tied up: it is the same under the combination of two half-unit positions as under the original risky portfolio. Therefore, the arbitrageurs' financial constraint is still met. Finally, the portfolio is riskless because $\hat{\omega}_i$ and $\hat{\omega}_{-i}$ have the same payoffs.

Step 2: From Step 1, we can assume that the portfolio of arbitrageurs in period t is as follows: (i) in each market i with $u_{i,t} > 0$, arbitrageurs hold a long position in a contract with one-period payoffs, (ii) in each market $-i$ with $u_{i,t} < 0$, arbitrageurs hold a short position of the same size as in market i and in a contract with the same payoffs, (iii) the payoffs of the contracts in markets i and $-i$ are binary and contingent on $\epsilon_{i,t+1}$, (iv) the short position in market $-i$ is collateralized with an investment in the riskless asset such that the arbitrageur wealth tied up in market $-i$ equals that in market i , (v) in each market i with $u_{i,t} = 0$, arbitrageurs hold a zero position.

Since the long position in the contract traded in each market i with $u_{i,t} > 0$ must have positive expected return, (B.8) implies that the contract must have larger payoff when $\epsilon_{i,t+1} = \bar{\epsilon}_i$ than when $\epsilon_{i,t+1} = -\bar{\epsilon}_i$. Moreover, we can take the payoff when $\epsilon_{i,t+1} = -\bar{\epsilon}_i$ to be zero since the contract price would then be lower, and hence arbitrageurs would be able to tie up less wealth in their long position in market i . We normalize the payoff when $\epsilon_{i,t+1} = \bar{\epsilon}_i$ to $2\bar{\epsilon}_i$, and denote by ω'_i the contract in market i and by ω'_{-i} the contract in market $-i$. We also denote by $x_{i,t}$ the number of units of the long position in ω'_i and of the short position in ω'_{-i} , by $q_{i,t}$ the price of ω'_i , and by $q_{-i,t}$ the price of ω'_{-i} .

The budget constraint (B.6) of arbitrageurs can be written as

$$\begin{aligned} W_{t+1} &= (1+r)W_t + (1+r) \sum_{i \in \mathcal{A}_t} x_{i,t}(q_{-i,t} - q_{i,t}) - c_{t+1} \\ &= (1+r)W_t + 2 \sum_{i \in \mathcal{A}_t} x_{i,t} [\bar{\epsilon}_i - (1+r)q_{i,t}] - c_{t+1}, \end{aligned} \tag{B.22}$$

where the second step follows because the same calculations as in (B.19) imply that

$$q_{-i,t} = \frac{2\bar{\epsilon}_i}{1+r} - q_{i,t}.$$

Since arbitrageurs must tie up wealth $x_{i,t}q_{i,t}$ in each of markets i and $-i$, their financial constraint (B.3) becomes

$$W_t \geq 2 \sum_{i \in \mathcal{A}_t} x_{i,t}q_{i,t}. \tag{B.23}$$

Eqs. (B.22) and (B.3) become identical to (20) and (26), respectively, by setting

$$\Phi_{i,t} \equiv \bar{\epsilon}_i - (1+r)q_{i,t}.$$

Because of this equivalence, if the dynamics of $\Phi_{i,t}$ are as in Section 3.4, then arbitrageurs' optimal positions $x_{i,t}$ and the dynamics of their wealth are also as in that section. Using (B.8) to substitute for $q_{i,t}$, we find

$$\Phi_{i,t} = \bar{\epsilon}_i - \frac{2 \exp(-\mathcal{Z}_{i,t}) \bar{\epsilon}_i}{\exp(-\mathcal{Z}_{i,t}) + \exp(\mathcal{Z}_{i,t})} = \bar{\epsilon}_i \frac{\exp(\mathcal{Z}_{i,t}) - \exp(-\mathcal{Z}_{i,t})}{\exp(-\mathcal{Z}_{i,t}) + \exp(\mathcal{Z}_{i,t})}.$$

This coincides with $\Phi_{i,t}$ given by (15) when $\epsilon_{i,t+1}$ has a binomial distribution. Therefore, arbitrageurs' optimal positions $x_{i,t}$ and the dynamics of their wealth are the same as in Section 3.4. Eq. (B.13) implies that the optimal positions of i -investors are $y_{i,t}$, as in Section 3.4. Since $\mu_i y_{i,t} + x_{i,t} = 0$, markets clear. ■

An alternative implementation of the equilibrium derived in Proposition B.1 is through the contracts assumed in Section 2. Two contracts are traded in market i . The first is asset i , with short positions in that contract being collateralized by the riskless asset. The second is a contract with a riskless payoff, with short positions in that contract being collateralized by asset i . The first contract is level 1, and the second is level 2. The collateral for each contract is the minimum required so that the no-default condition (B.4) is met. A short position of arbitrageurs in the first contract, combined with the required collateral, yields zero if $\epsilon_{t+1} = \bar{\epsilon}_i$ and $2\bar{\epsilon}_i$ if $\epsilon_{t+1} = -\bar{\epsilon}_i$. A short position of arbitrageurs in the second contract, combined with the required collateral, yields $2\bar{\epsilon}_i$ if $\epsilon_{t+1} = \bar{\epsilon}_i$ and zero if $\epsilon_{t+1} = -\bar{\epsilon}_i$. The former is equivalent to the short position in ω'_{-i} , and the latter is equivalent to the long position in ω'_i .

References

- Amihud, Yakov, and Haim Mendelson, 1991, Liquidity, maturity, and the yield on US treasury securities, *Journal of Finance* 46, 479–486.
- Basak, Suleyman, and Benjamin Croitoru, 2000, Equilibrium mispricing in a capital market with portfolio constraints, *Review of Financial Studies* 13, 715–748.
- , 2006, On the role of arbitrageurs in rational markets, *Journal of Financial Economics* 81, 143–173.
- Brumm, Johannes, Michael Grill, Felix Kubler, and Karl Schmedders, 2011, Collateral requirements and asset prices, working paper Swiss Finance Institute.
- Brunnermeier, Markus, and Lasse Pedersen, 2009, Market liquidity and funding liquidity, *Review of Financial Studies* 22, 2201–2238.
- Brunnermeier, Markus, and Yuliy Sannikov, 2014, A macroeconomic model with a financial sector, *American Economic Review* 104, 379–421.
- Comerton-Forde, Carole, Terrence Hendershott, Charles Jones, Pamela Moulton, and Mark Seasholes, 2010, Time variation in liquidity: The role of market-maker inventories and revenues, *Journal of Finance* 65, 295–331.
- Coval, Joshua, and Erik Stafford, 2007, Asset fire sales (and purchases) in equity markets, *Journal of Financial Economics* 86, 475–512.
- Dabora, Emil, and Kenneth Froot, 1999, How are stock prices affected by the location of trade?, *Journal of Financial Economics* 53, 189–216.
- Detemple, Jerome, and Shashidhar Murthy, 1997, Equilibrium asset prices and no-arbitrage with portfolio constraints, *Review of Financial Studies* 10, 1133–1174.
- Duffie, Darrell, and Bruno Strulovici, 2012, Capital mobility and asset pricing, *Econometrica* 80, 2469–2509.
- Fostel, Ana, and John Geanakoplos, 2013a, Leverage and default in binomial economies: A complete characterization, discussion paper Cowles Foundation, Yale University.
- , 2013b, Reviewing the leverage cycle, discussion paper Cowles Foundation, Yale University.
- Garleanu, Nicolae, and Lasse Pedersen, 2011, Margin-based asset pricing and deviations from the law of one price, *Review of Financial Studies* 24, 1980–2022.

- Geanakoplos, John, 1997, Promises, promises, in W. Brian Arthur, Stephen Durlauf, and David Lane, ed.: *The Economy as an Evolving Complex System II* . pp. 285–320 (Addison-Wesley: Reading, MA).
- , 2003, Liquidity, default and crashes: Endogenous contracts in general equilibrium, in Mathias Dewatripont, Lars Hansen, and Stephen Turnovsky, ed.: *Advances in Economics and Econometrics: Theory and Applications II, Econometric Society Monographs: Eighth World Congress* . pp. 170–205 (Cambridge University Press: Cambridge, UK).
- Gottardi, Piero, and Felix Kubler, 2014, Dynamic competitive economies with complete markets and collateral constraints, *Review of Economic Studies* forthcoming.
- Gromb, Denis, and Dimitri Vayanos, 2002, Equilibrium and welfare in markets with financially constrained arbitrageurs, *Journal of Financial Economics* 66, 361–407.
- , 2010, Limits of arbitrage, *Annual Review of Financial Economics* 2, 251–275.
- , 2015, Financially constrained arbitrage and cross-market contagion, working paper INSEAD.
- Guembel, Alexander, and Oren Sussman, 2015, A welfare analysis of fragmented liquidity markets, working paper Toulouse School of Economics.
- Hameed, Allaudeen, Wenjin Kang, and S. Viswanathan, 2010, Stock market declines and liquidity, *Journal of Finance* 65, 257–294.
- He, Zhiguo, and Arvind Krishnamurthy, 2013, Intermediary asset pricing, *American Economic Review* 103, 1–43.
- Jylha, Petri, and Matti Suominen, 2011, Speculative capital and currency carry trades, *Journal of Financial Economics* 99, 60–75.
- Kondor, Peter, and Dimitri Vayanos, 2014, Liquidity risk and the dynamics of arbitrage capital, working paper Central European University.
- Krishnamurthy, Arvind, 2002, The bond/old-bond spread, *Journal of Financial Economics* 66, 463–506.
- Kyle, Albert, and Wei Xiong, 2001, Contagion as a wealth effect, *Journal of Finance* 56, 1401–1440.
- Nagel, Stefan, 2012, Evaporating liquidity, *Review of Financial Studies* 25, 2005–2039.
- Pavlova, Anna, and Roberto Rigobon, 2008, The role of portfolio constraints in the international propagation of shocks, *Review of Economic Studies* 75, 1215–1256.

- Rosenthal, Leonard, and Colin Young, 1990, The seemingly anomalous price behavior of Royal Dutch/Shell and Unilever N.V./PLC, *Journal of Financial Economics* 26, 123–141.
- Shen, Ji, Hongjun Yan, and Jinfan Zhang, 2014, Collateral-motivated financial innovation, *Review of Financial Studies* forthcoming.
- Shleifer, Andrei, and Robert Vishny, 1997, The limits of arbitrage, *Journal of Finance* 52, 35–55.
- Simsek, Alp, 2013, Belief disagreements and collateral constraints, *Econometrica* 81, 1–53.
- Wagner, Wolf, 2011, Systemic liquidation risk and the diversity-diversification trade-off, *Journal of Finance* 66, 1141–1175.
- Warga, Arthur, 1992, Bond returns, liquidity, and missing data, *Journal of Financial and Quantitative Analysis* 27, 605–617.
- Xiong, Wei, 2001, Convergence trading with wealth effects: An amplification mechanism in financial markets, *Journal of Financial Economics* 62, 247–292.