ISSN 0956-8549-771

The Dynamics of Financially Constrained Arbitrage

By
Denis Gromb
Dimitri Vayanos

DISCUSSION PAPER NO 771

DISCUSSION PAPER SERIES

August 2017

The Dynamics of Financially Constrained Arbitrage

Denis Gromb
HEC Paris
gromb@hec.fr

Dimitri Vayanos
LSE, CEPR and NBER
d.vayanos@lse.ac.uk

August 14, 2017*

Abstract

We develop a model in which financially constrained arbitrageurs exploit price discrepancies across segmented markets. We show that the dynamics of arbitrage capital are self-correcting: following a shock that depletes capital, returns increase, and this allows capital to be gradually replenished. Spreads increase more for trades with volatile fundamentals or more time to convergence. Arbitrageurs cut their positions more in those trades, except when volatility concerns the hedgeable component. Financial constraints yield a positive cross-sectional relationship between spreads/returns and betas with respect to arbitrage capital. Diversification of arbitrageurs across markets induces contagion, but generally lowers arbitrageurs' risk and price volatility.

Keywords: Arbitrage, financial constraints, market segmentation, liquidity, contagion.

*We thank Philippe Bacchetta, Bruno Biais (the editor), Patrick Bolton, Darrell Duffie, Vito Gala, Jennifer Huang, Henri Pagès, Anna Pavlova, Matti Suominen, an anonymous associate editor and three referees, 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 so called market anomalies, notably those concerning price discrepancies between assets with almost identical payoffs. Such discrepancies arise in a variety of markets, during both crises and more tranquil times. For example, large and persistent violations of covered interest parity have been documented for all major currency pairs, both during and after the global financial crisis. Price discrepancies that are hard to reconcile with frictionless arbitrage have also been documented for stocks, government bonds, corporate bonds, and credit default swaps.[1]

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 have limited capital, and this can constrain their activity and ultimately affect market liquidity and asset prices. Empirical studies have constructed various measures of arbitrage capital and shown them to be related to the magnitude of the anomalies.

Since arbitrage capital can be targeted at multiple anomalies, the returns to investing in the anomalies are interdependent and so are arbitrageurs' positions. In this paper we develop a model to address a number of questions that this interdependence raises. How should arbitrageurs allocate their limited capital across anomalies, and how should this allocation respond to shocks to capital? Which anomalies' returns are more sensitive to changes in arbitrage capital? How do the expected returns offered by the different anomalies relate to sensitivity to arbitrage capital and other characteristics? How do the expected returns offered by anomalies evolve over time, and how do these dynamics relate to those of arbitrage capital?

We consider a discrete-time, infinite-horizon economy, with a riskless asset and a number of "arbitrage opportunities" (the anomalies within our model) each consisting of a pair of risky assets with correlated payoffs. Each risky asset is traded in a different segmented market by risk-averse investors who can trade only that asset and the riskless asset. Investors experience endowment shocks that generate a hedging demand for the risky asset in their market. Shocks are opposites within each pair, so a positive hedging demand for one asset in the pair is associated with a negative hedging demand of equal magnitude for the other. This simplifying assumption ensures that arbitrageurs trade only on the price discrepancy between the two assets. Market segmentation is exogenous in our model, but could arise because of regulation, agency problems, or lack of specialized knowledge.

[1] References to the empirical literature are in Sections 2.2.3 and 4.3.2. In these sections we also explain how to map our model and results to the empirical settings.

We make two key assumptions. First, unlike other investors, arbitrageurs can trade all assets. Thus, they have better opportunities than other investors. By exploiting price discrepancies between paired assets, they intermediate trade between otherwise segmented investors, providing them with liquidity: they buy cheap assets from investors with negative hedging demand, and sell expensive assets to investors with positive hedging demand. We term the price discrepancies that arbitrageurs seek to exploit "arbitrage spreads" and use them as an inverse measure of liquidity.

Second, we assume that 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 former 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 can offset 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 be a riskless asset position. We assume that collateral must be sufficient to protect the lender fully against default. This implies, in particular, that positions in assets with more volatile payoffs require more collateral so that lenders are protected against larger losses. The need for collateral limits the positions that an arbitrageur can establish, and that constraint is a function of his wealth. The positions that arbitrageurs can establish as a group are constrained by their aggregate wealth, which we also refer to as arbitrage capital.

When assets in each pair have identical payoffs, arbitrage is riskless. This case is a natural benchmark, and we analyze it first. If spreads are positive, then the riskless return offered by arbitrage opportunities exceeds the riskless rate. Arbitrageurs, however, may not be able to scale up their positions to exploit that return because of their financial constraint. Their 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 opportunities below the cutoff. The cutoff 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.

The inverse relationship between returns and capital implies self-correcting dynamics and a deterministic steady state. If arbitrage capital is low, then arbitrageurs hold small positions, returns are high, and capital gradually increases. Conversely, if capital is high, then returns are low and capital decreases because of arbitrageurs' consumption. In steady state, arbitrage remains

profitable enough to offset the natural depletion of capital due to consumption.

We next analyze the case where payoffs within each asset pair consist of a component that is identical across the two assets and hedgeable by arbitrageurs, and a component that differs. Because asset payoffs are not identical, arbitrage is risky. As in the riskless-arbitrage case, arbitrageurs invest in the opportunities that offer maximum return per unit of collateral. Unlike in that case, however, the relevant return is the expected return net of a risk adjustment that depends on arbitrageur risk aversion and position size. The financial constraint binds when the risk-adjusted return exceeds the riskless rate.

To compute the equilibrium under risky arbitrage in closed form, we specialize our analysis to the case were asset payoffs are near-identical and hence arbitrage risk is small. In the stochastic steady state, the financial constraint always binds and arbitrage capital follows an approximate AR(1) process. Moreover, the first-order effect of arbitrage risk on equilibrium variables operates through the financial constraint rather than through risk aversion. Indeed, price movements caused by shocks to arbitrage capital represent an additional source of risk for a collateralized position. The required collateral must then increase by an amount proportional to the standard deviation of these movements. On the other hand, the risk adjustment induced by risk aversion is proportional to the variance because it is an expectation of gains and losses weighted by marginal utility.

Using our closed-form solutions, we can determine the cross-section of expected returns and arbitrageur positions. We show that expected returns are high for arbitrage opportunities involving assets with volatile payoffs because these opportunities require more collateral. They are also high for "long-horizon" opportunities, i.e., opportunities for which price discrepancies take longer to disappear because endowment shocks have longer duration. Indeed, because spreads for these opportunities are more sensitive to shocks to arbitrageur wealth, the losses that arbitrageurs can incur are larger, implying higher collateral requirements.

The characteristics associated with high expected returns are also associated with high sensitivity of spreads to arbitrage capital, i.e., high "arbitrage-capital betas." Since opportunities with volatile payoffs require more collateral, they must offer high expected returns. Since, in addition, changes in capital impact the return per unit of collateral, arbitrage-capital betas for the same opportunities are high. In the case of long-horizon opportunities, the causal channel is different: high arbitrage-capital betas result in high collateral requirements, which in turn result in high expected returns. Our results are consistent with the relationship between expected returns or spreads on one hand and arbitrage-capital betas on the other being increasing in the cross-section, as documented in Avdjiev, Du, Koch, and Shin (2016) in the context of covered-interest arbitrage and Cho (2016) in the context of stock-market anomalies.

The cross-section of arbitrageur positions differs from that of expected returns. Arbitrageurs hold larger positions in opportunities where the hedgeable component of payoff volatility is larger, but smaller positions in opportunities where the unhedgeable component is larger or where horizon is longer. Intuitively, volatility has two countervailing effects on arbitrageur positions: it lowers them because it raises collateral requirements, but it raises them because it raises investors' hedging demand and need for intermediation. The effect of each component of volatility on collateral requirements is proportional to its standard deviation, while that on hedging demand is proportional to its variance. The former is larger in the case of small unhedgeable volatility, i.e., small arbitrage risk. The effect of volatility on the dynamics of positions parallels that on average positions. Following drops to arbitrage capital, positions in opportunities with higher unhedgeable volatility are cut by more, while positions in opportunities with higher hedgeable volatility are cut by less.

We finally use our model to study how the degree of mobility of arbitrage capital affects market stability: does capital mobility stabilize markets, or does it propagate shocks causing contagion and instability? To do so, we consider the possibility that any given arbitrageur can allocate his wealth to exploit only one opportunity. That is, arbitrage markets themselves are segmented so that arbitrage capital cannot be reallocated from one opportunity to another. For simplicity we take opportunities to be symmetric with independent payoffs. If an arbitrageur can diversify across opportunities but others remain undiversified, then the variance of his wealth decreases because spreads are independent. If instead all arbitrageurs can diversify, then spreads become perfectly correlated, as arbitrageurs act as conduits transmitting shocks in one market to all markets—a contagion effect. We show, however, that because collective diversification causes the variance of each spread to decrease, the variance of each arbitrageur's wealth decreases. In fact, collective diversification lowers wealth variance by as much as individual diversification. In that sense, capital mobility stabilizes markets.

Our paper belongs to a growing theoretical literature on the limits of arbitrage, and more precisely to its strand emphasizing arbitrageurs' financial constraints.[2] We contribute to that literature by deriving the cross section and dynamics of arbitrageur returns and positions in a setting where arbitrageurs exploit price discrepancies between assets with similar payoffs.

Shleifer and Vishny (1997) are the first to derive a 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 single arbitrage opportunity and a finite horizon. These assumptions rule out, respectively, cross-sectional effects and self-correcting dynamics.

Our result that arbitrage opportunities with higher collateral requirements offer higher returns

[2]For a survey of this literature, see Gromb and Vayanos (2010).

is related to a number of papers. In Geanakoplos (2003), Garleanu and Pedersen (2011), and Brumm, Grill, Kubler, and Schmedders (2015), multiple risky assets differ in their collateral value, i.e., the amount that agents can borrow using the asset as collateral. Assets with low collateral value must offer higher expected returns, and violations of the law of one price can arise.[3] These violations, however, are different in nature from those in our model: we assume that both assets in a pair have the same collateral value but differ in investors' hedging demand. Empirical studies confirm that hedging demand (or more generally demand unrelated to collateral value) is a key driver of arbitrage spreads.[4]

In Brunnermeier and Pedersen (2009), collateral-constrained arbitrageurs invest in assets with maximum return per unit of collateral. Since volatile assets require more collateral, their returns are higher and more sensitive to changes in arbitrage capital. In that paper, however, there is no segmentation and the law of one price holds. Moreover, the analysis does not address dynamic issues such as the effect of horizon or the recovery from shocks.

Our results on self-correcting dynamics are related to several papers. In Duffie and Strulovici (2012), capital recovers following adverse shocks because new capital enters the market. In Xiong (2001), He and Krishnamurthy (2013), and Brunnermeier and Sannikov (2014), recovery instead occurs because existing capital grows faster—the same channel as in our model. In these papers, however, arbitrageurs invest in a single risky asset. This rules out cross-sectional effects and violations of the law of one price.[5]

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) and Caballero and Simsek (2017), who show that segmentation generally raises volatility and reduces investor welfare. 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 the equilibrium when arbitrage is riskless because assets in each pair have identical payoffs. Section 4 analyzes risky arbitrage, and derives the cross-sectional properties of prices and positions, as well as the effects of capital mobility. Section 5 concludes, and proofs are in the Appendix.

[3]Detemple and Murthy (1997), Basak and Croitoru (2000, 2006), and Chabakauri (2013) derive related results for more general portfolio constraints.

[4]See, for example, the literature on covered interest arbitrage, summarized in Section 2.2.3.

[5]Kondor and Vayanos (2016) derive self-correcting dynamics in a setting where arbitrageurs can invest in multiple risky assets. Arbitrageurs in their setting, however, do not intermediate trades because there is no segmentation, and the law of one price holds. Greenwood, Hanson, and Liao (2015) assume gradual rebalancing of arbitrageur portfolios across markets, in the spirit of Duffie (2010) and Duffie and Strulovici (2012), and allow for multiple risky assets within each market. Arbitrageurs in their setting, however, face no financial constraints.

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 I of infinitely lived risky assets, all in zero supply. Risky assets come in pairs. Asset i's payoff per share in period t is

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

where $\bar{d}_i$ is a positive constant, and $\epsilon_{i,t}$ and $\eta_{i,t}$ are random variables distributed symmetrically around zero in the respective intervals $[-\bar{\epsilon}_i, \bar{\epsilon}_i]$ and $[-\bar{\eta}_i, \bar{\eta}_i]$. The other asset in i's pair is denoted by $-i$ and its payoff per share in period t is

$$d_{-i,t} \equiv \bar{d}_i + \epsilon_{i,t} - \eta_{i,t}. \tag{2}$$

If $\bar{\eta}_i = 0$, then assets i and $-i$ have identical payoffs, and a trade consisting of an one-share long position in one asset and an one-share short position in the other involves no risk. If instead $\bar{\eta}_i > 0$, then payoffs are not identical and the long-short trade is risky. In both cases, we refer to asset pair $(i, -i)$ as an arbitrage opportunity. This corresponds to textbook arbitrage when $\bar{\eta}_i = 0$ and the two assets trade at different prices.

We assume that the variables $\frac{\epsilon_{i,t}}{\bar{\epsilon}_i}$ are *i.i.d.* across time and identically distributed across asset pairs (but correlation across pairs is possible). We make the same assumption for the variables $\frac{\eta_{i,t}}{\bar{\eta}_i}$, which we also assume independent of $\frac{\epsilon_{i,t}}{\bar{\epsilon}_i}$. Because distributions are identical across asset pairs, $\bar{\epsilon}_i$ and $\bar{\eta}_i$ are proportional to the standard deviations of $\epsilon_{i,t}$ and $\eta_{i,t}$, respectively, and we refer to them as volatilities. We restrict $\bar{d}_i$ to be larger than $\bar{\epsilon}_i + \bar{\eta}_i$ so that asset payoffs are non-negative. We denote by $p_{i,t}$ the ex-dividend price of asset i in period t, and define the asset's price discount by

$$\phi_{i,t} \equiv \frac{d_i}{r} - p_{i,t}, \tag{3}$$

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

2.2 Outside Investors

2.2.1 Market Segmentation

For some agents, who we term 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

$$-\mathrm{E}_t\left[\sum_{s=t+1}^{\infty}\gamma^{s-t}\exp(-\alpha c_{i,s})\right], \tag{4}$$

where α is the coefficient of absolute risk aversion and γ is the subjective discount factor. 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 (4) subject to a budget constraint. 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 i- and $-i$-investors are identical in terms of their measure, i.e., $\mu_i = \mu_{-i}$. Negative exponential utility of outside investors simplifies our analysis because it ensures that their demand for risky assets is independent of their wealth. The only wealth effects in our model concern the arbitrageurs.

2.2.2 Endowment Shocks

Outside investors receive random endowments, which affect their appetite for risky assets. In period t each i-investor receives an endowment equal to

$$u_{i,t-1}(\epsilon_{i,t}+\eta_{i,t}), \tag{5}$$

where $u_{i,t-1}$ is known in period $t-1$. We assume that $u_{i,t}$ is equal to zero, except over a sequence of M_i periods $t \in \{h_i - M_i, .., h_i - 1\}$ during which it can become equal to a constant u_i. When this occurs, we say that i-investors experience an endowment shock of intensity u_i and duration M_i.

An endowment shock in market i is accompanied by one in market $-i$. If the shocks were identical, then assets i and $-i$ would be trading at the same price in the absence of arbitrageurs because of symmetry. To ensure a difference in prices and hence a role for arbitrageurs, we assume that endowment shocks differ. We further restrict the shocks to be opposites, i.e., $u_i = -u_{-i}$. This assumption, together with that of zero supply, ensures that the price discounts of assets i and

$-i$ are opposites in equilibrium. With opposite price discounts, arbitrageurs (described in Section 2.3) find it optimal to hold opposite positions in the two assets, hence trading only on the price discrepancy between them. This simplifies the equilibrium because arbitrageurs are not exposed to the shock $\epsilon_{i,t}$, and hence earn a riskless return when assets i and $-i$ have identical payoffs.

Arbitrageurs exploit price discrepancies, and in doing so they intermediate trade between investors and provide liquidity to them. Suppose, for example, that i-investors experience a shock $u_i > 0$. Their endowment then becomes positively correlated with $\epsilon_{i,t} + \eta_{i,t}$ and hence with asset i's payoff. As a consequence, asset i becomes less attractive to them. Conversely, asset $-i$ becomes more attractive to $-i$-investors, who experience a shock $u_{-i} < 0$. In the absence of arbitrageurs, the equilibrium price of asset i would decrease and that of asset $-i$ would increase. Arbitrageurs can exploit this price discrepancy by buying asset i from i-investors and selling asset $-i$ to $-i$-investors. In doing so, they intermediate trade between the two sets of investors, which market segmentation prevents otherwise. Because of arbitrageurs, prices are less sensitive to endowment shocks and price discrepancies are smaller.[6]

When investors i and $-i$ experience endowment shocks, we say that arbitrage opportunity $(i, -i)$ is active. We identify active opportunities with the assets with the positive endowment shocks: we set

$$A_t \equiv \{i \in I : u_{i,t} > 0\},$$

and refer to active opportunity $(i, -i)$ for $i \in A_t$ as opportunity i. We assume that the set A_t of active opportunities is finite. We also assume that the probability of an opportunity becoming active (an event that may occur in period $h_i - M_i$ for opportunity i) is arbitrarily small. This is consistent with opportunities forming a continuum and a finite number being active in each period. This also ensures that endowment shocks do not affect prices until they actually hit investors.

[6] If endowment shocks for i- and $-i$-investors were not opposites, then arbitrageurs would not hold opposite positions in assets i and $-i$. They would still intermediate trade between investors, however, if they are sufficiently risk averse. Suppose, for example, that i-investors experience a shock $u_i > 0$ but $-i$-investors do not, i.e., $u_{-i} = 0$. Arbitrageurs would buy asset i from i-investors to benefit from its positive price discount. If they are sufficiently risk averse, they would hedge that position by selling asset $-i$ to $-i$-investors, hence intermediating trade between i- and $-i$-investors. Because buying asset i yields a higher expected excess return than selling asset $-i$, arbitrageurs would choose not to be fully hedged, and hence would be exposed to the risk that $\epsilon_{i,t}$ is low.

If assets i and $-i$ were in positive rather than in zero supply, then arbitrageurs would hold a larger long position in asset i and would be more exposed to the risk that $\epsilon_{i,t}$ is low. If assets without endowment shocks were also in positive supply then they would trade at a positive price discount but one that would be smaller than asset i's. Because positions in the no-endowment-shock assets require a comparable level of collateral as in assets i and $-i$ but earn a lower expected return, arbitrageurs would not trade those assets if their wealth were small enough.

2.2.3 Interpretation

Our assumptions fit settings where assets with similar payoffs trade in partially segmented markets. These include Siamese-twin stocks, covered interest arbitrage across currencies, government bonds, corporate bonds, and credit-default swaps (CDS).

Siamese-twin stocks have identical dividend streams but differ in the country where most of their trading occurs. Rosenthal and Young (1990) find that price differences between Siamese twins can be significant. Dabora and Froot (1999) show that a stock tends to appreciate relative to its Siamese twin when the aggregate stock market in the country where that stock is mostly traded goes up. They argue that one reason why Siamese-twin stocks differ in their main trading venue is that each stock belongs to a different country's main stock index. Thus, index funds in each country can only invest in one of the stocks. Index funds in that setting correspond to our model's outside investors, flows in or out of these funds correspond to our model's endowment shocks, and market segmentation arises from restricted fund mandates (which are possibly a response to agency problems).

Covered interest arbitrage exploits violations of covered interest parity (CIP), the relationship implied by absence of arbitrage between the spot and forward exchange rates for a currency pair and the interest rates on the two currencies. Violations of CIP can be measured by the cross-currency basis (CCB). Taking one of the currencies to be the dollar, the CCB is the difference between the dollar interest rate minus its CIP-implied value. A negative CCB indicates that the dollar is cheaper in the forward market than its CIP-implied value.

Violations of CIP have been small from 2000 until the global financial crisis, but have become large both during and after the crisis. Explanations of CIP violations during the crisis have focused on increased counterparty risk and difficulty to borrow in dollars.[7] These factors, however, have subsided after the crisis, and explanations of CIP violations since then have instead focused on hedging pressure in the forward market combined with financially constrained arbitrage. Borio, McCauley, McGuire, and Sushko (2016) construct measures of the hedging demand of banks, institutional investors (such as pension funds and insurance companies), and non-financial firms. Consistent with the hedging pressure explanation, they find that a negative CCB is more likely when these institutions seek to hedge against a drop in the dollar. Du, Tepper, and Verdelhan (2016) argue that the demand for hedging against a drop in the dollar should be high for currencies with low interest rates relative to the dollar, and find that a negative CCB is indeed more likely for those currencies. They also relate the CCB to measures of arbitrageurs' financial constraints.[8]

[7] See, for example, Baba and Packer (2009), Coffey, Hrung, and Sarkar (2009), and Mancini Griffoli and Ranaldo (2012).

[8] Other related work on CIP violations after the crisis includes Avdjiev, Du, Koch, and Shin (2016), Iida, Kimura,

Our model can be applied to covered interest arbitrage by interpreting the two assets in a pair as a currency forward and its synthetic counterpart. Outside investors in the forward market are the hedgers that Borio, McCauley, McGuire, and Sushko (2016) consider: these agents may lack the specialized knowledge or trading infrastructure to access synthetic forwards. Likewise, outside investors in the synthetic forward market may be prevented from trading forwards because of restricted mandates.[9]

Bonds with similar coupon rates and times to maturity can trade at significantly different yields. Fontaine and Garcia (2012) and Hu, Pan, and Wang (2013) aggregate such deviations across the nominal term structure by fitting it to a smooth curve. They find that the fit worsens when arbitrageurs' financial constraints tighten, e.g., during financial crises or when the leverage of shadow banks decreases. In that context, outside investors can represent investors who must hold bonds with specific coupon rates and times to maturity. Such investors might be insurance companies or pension funds, and their preferences could be driven by asset-liability management or tax considerations.

Fleckenstein, Longstaff, and Lustig (2014) find that nominal government bonds tend to be significantly more expensive than their synthetic counterparts formed by inflation-indexed bonds, inflation swaps, and zero-coupon bonds. Moreover, price discrepancies become larger when arbitrage capital, measured by hedge-fund assets, is depleted. The additional finding that nominal and inflation-indexed bonds are owned by different types of institutions suggests a degree of market segmentation.

Duffie (2010) documents price discrepancies between corporate bonds and matched CDS. These discrepancies became particularly large during the global financial crisis, but remained significant afterwards. One driver of market segmentation in that setting is that individual investors can trade corporate bonds but not CDS.

2.3 Arbitrageurs

2.3.1 Better Investment Opportunities

Arbitrageurs can invest in all risky assets and in the riskless asset. Hence, only they can overcome market segmentation. We assume that they are competitive and infinitely lived, form a continuum

and Sudo (2016), Liao (2016), and Sushko, Borio, McCauley, and McGuire (2016).

[9]Consider, for example, US non-financial firms that issue debt in euros to benefit from lower credit spreads in the euro area relative to the US (Borio, McCauley, McGuire, and Sushko (2016)). Those firms seek to hedge against a drop in the dollar as they earn most of their profits in dollars but must pay euro-denominated debt. They can hedge in the forward market, but trading synthetic forwards may be too complicated or impossible for them: in particular, they would have to borrow dollars without paying a credit spread. Conversely, bond mutual funds can invest in euro- or dollar-denominated bonds but may be prevented by their mandates from trading currency forwards. Liao (2016) links CIP violations to the hedging demand of non-financial firms using a segmented-markets model.

with measure one, consume in each period, and have logarithmic utility

$$\mathrm{E}_t\left[\sum_{s=t+1}^{\infty}\beta^{s-t}\log\left(c_s\right)\right], \tag{6}$$

where β is the subjective discount factor.[10] In period t, an arbitrageur chooses positions $\{x_{i,s}\}_{i\in I,s\geq t}$ in all risky assets and consumption $\{c_s\}_{s\geq t+1}$ to maximize (6). The arbitrageur is subject to a financial constraint (see section 2.3.2) and a budget constraint. 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, which we also refer to as arbitrage capital. Logarithmic utility of arbitrageurs simplifies our analysis because it ensures that their consumption is a constant fraction of their wealth.

2.3.2 Financial Constraint

We assume that agents must 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 commit to repay the cash loan. Consider next an agent who wants to establish a short position in a risky asset. The agent must borrow the asset so that he can sell it subsequently, and must post collateral to commit to return the asset. We assume that i-investors have enough wealth to collateralize any position they may want to establish, i.e., up to $\mu_i u_i$. Arbitrageurs, however, may be constrained by their wealth.[11]

Standard asset pricing models assume that agents can establish any combination of asset positions provided they have sufficient wealth to cover 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 cover. The constraint is formally equivalent to requiring wealth to be always non-negative, and is thus redundant for agents with logarithmic utility.

We assume that arbitrageurs are subject to a stronger constraint. We require them to have sufficient wealth *in each market* to cover any liabilities that their positions in that market generate.

[10] By fixing the measure of arbitrageurs, we are ruling out entry and are focusing the evolution of 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.

[11] Our 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 the case 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.

The positions of arbitrageurs in market i consist of a position in asset i and a position in the riskless asset. We require this combined position to always have non-negative value. Thus, liability is calculated market-by-market rather than by aggregating across all markets. This is in the spirit of market segmentation: the same informational or regulatory frictions that prevent i-investors for investing in risky assets other than asset i could also be preventing arbitrageurs' lenders in market i from accepting such assets as collateral.[12]

To derive the financial constraint of an arbitrageur, we denote by $x_{i,t}$ his position in asset i, by $z^0_{i,t}$ his investment in the riskless asset held in market i, and by $z_{i,t} = x_{i,t}p_{i,t} + z^0_{i,t}$ the value of his combined position in market i, all in period t. The value of the arbitrageur's combined position in market i in period $t+1$ is

$$\begin{aligned} z_{i,t+1} &= z^0_{i,t}(1+r) + x_{i,t}\,[d_{i,t+1} + p_{i,t+1}] \\ &= z_{i,t}(1+r) + x_{i,t}\,[d_{i,t+1} + p_{i,t+1} - (1+r)p_{i,t}] \end{aligned} \tag{7}$$

and must be non-negative. Requiring (7) to be non-negative for all possible realizations of uncertainty in period $t+1$ yields

$$z_{i,t} \geq \max_{\{\epsilon_{j,t+1},\eta_{j,t+1}\}_{j\in I}} \left\{ x_{i,t}\left(p_{i,t} - \frac{d_{i,t+1}+p_{i,t+1}}{1+r} \right) \right\}. \tag{8}$$

The right-hand side of (8) represents the maximum loss, in present-value terms, that the arbitrageur can realize in market i between periods t and $t+1$. This loss must 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 (8) across markets yields the financial constraint

$$W_t = \sum_{i\in I} z_{i,t} \geq \sum_{i\in I} \max_{\{\epsilon_{j,t+1},\eta_{j,t+1}\}_{j\in I}} \left\{ x_{i,t}\left(p_{i,t} - \frac{d_{i,t+1}+p_{i,t+1}}{1+r} \right) \right\} \tag{9}$$

[12]Using 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 model rules out cross-netting, it can be generalized to allow for partial cross-netting.

since the value of the arbitrageur's positions summed across markets is his wealth W_t. The constraint (9) requires the arbitrageur to have enough wealth to cover his maximum loss in each market separately.[13]

Our formulation of the financial constraint assumes that the only assets that arbitrageurs can trade with i-investors, or 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 such general contracts. The only restrictions that we are maintaining are that i-investors cannot contract directly with j-investors for $j \neq i$ (only indirectly through arbitrageurs), and that arbitrageurs cannot use contracts traded with j-investors as collateral for contracts traded with i-investors. These restrictions are consistent with the logic of market segmentation.

We show in Appendix B that without loss of generality, contracts can be assumed to be fully collateralized and hence default-free. Moreover, when assets in each pair have identical payoffs ($\eta_i = 0$) and distributions are binomial ($\epsilon_{i,t}$ binomial), contracts can be restricted to those studied in this section: only asset i and the riskless asset are traded and used as collateral in market i. This generalizes, within our setting, the no-default result of Fostel and Geanakoplos (2015), shown under the assumption that contracts extend over one period.[14] Thus, the financial constraint (9) can be derived under general contracts that are consistent with market segmentation.

2.4 Symmetric Equilibrium

We look for a competitive equilibrium that is symmetric in the sense that price discounts and agents' positions are opposites for the assets in each pair.

Defi **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:*

$$\mu_i y_{i,t} + x_{i,t} = 0. \tag{10}$$

[13]The constraint (9) can extend to a continuous-time limit of our model if that limit involves jumps. With jumps, the support of $\{\epsilon_{j,t+1}, \eta_{j,t+1}\}_{j\in I}$ does not converge to zero and neither does the maximum loss in period $t+1$. If instead the support of $\{\epsilon_{j,t+1}, \eta_{j,t+1}\}_{j\in I}$ converges to zero, as it would in a Brownian limit, then the maximum loss converges to zero and (9) is always met. For a derivation of a financial constraint in a continuous-time limit with jumps, see Chabakauri and Han (2017).

[14]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 (2015). Simsek (2013) characterizes default rates in collateral equilibrium for general distributions in a static setting. For more references on leverage and collateral equilibrium, see the survey by Fostel and Geanakoplos (2015).

Defi **2.** *A competitive equilibrium is symmetric if for the assets* $(i, -i)$ *in each pair the price discounts 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 price discount 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 price discount measures the price difference between paired assets, we also refer to it as the spread. The spread is an inverse measure of the liquidity that arbitrageurs provide to outside investors.

2.5 Optimization Problems

2.5.1 Outside Investors

The budget constraint of an i-investor is

$$w_{i,t+1} = y_{i,t}(d_{i,t+1} + p_{i,t+1}) + (1 + r)(w_{i,t} - y_{i,t}p_{i,t}) + u_{i,t}(\epsilon_{i,t+1} + \eta_{i,t+1}) - c_{i,t+1}. \tag{11}$$

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 $u_{i,t}(\epsilon_{i,t+1}+\eta_{i,t+1})$ is added to the investor's wealth in period $t+1$, while consumption $c_{i,t+1}$ lowers wealth.

We can simplify (11) by introducing the return per share of asset i in excess of the riskless asset. This excess return is

$$\begin{aligned} R_{i,t+1} &\equiv d_{i,t+1} + p_{i,t+1} - (1 + r)p_{i,t} \\ &= (1 + r)\phi_{i,t} - \phi_{i,t+1} + \epsilon_{i,t+1} + \eta_{i,t+1}, \end{aligned} \tag{12}$$

where the second step follows from (1) and (3). The expected excess return of asset i is

$$\Phi_{i,t} \equiv \mathrm{E}_t\left(R_{i,t+1}\right) = (1 + r)\phi_{i,t} - \mathrm{E}_t\left(\phi_{i,t+1}\right). \tag{13}$$

Using (12) and (13), we can write (11) as

$$w_{i,t+1} = (1+r)w_{i,t} + y_{i,t}\Phi_{i,t} + (y_{i,t}+u_{i,t})(\epsilon_{i,t+1}+\eta_{i,t+1}) + y_{i,t}\,[\mathrm{E}_t(\phi_{i,t+1}) - \phi_{i,t+1}] - c_{i,t+1}. \quad (14)$$

The investor's wealth in period $t+1$ is uncertain as of period t because of the payoff shock $\epsilon_{i,t+1}+\eta_{i,t+1}$ and the price discount $\phi_{i,t+1}$. The investor's exposure to the payoff shock is the sum of her asset position $y_{i,t}$ and endowment shock $u_{i,t}$, while her exposure to the price discount is $y_{i,t}$.

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

$$V_{i,t}(w_{i,t}) = -\exp(-Aw_{i,t} - F_{i,t}), \quad (15)$$

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

2.5.2 Arbitrageurs

The budget constraint of an arbitrageur is

$$W_{t+1} = \sum_{i\in I} x_{i,t}(d_{i,t+1}+p_{i,t+1}) + (1+r)\left(W_t - \sum_{i\in I} x_{i,t}p_{i,t}\right) - c_{t+1}. \quad (16)$$

The differences with the budget constraint (11) of an i-investor are that the arbitrageur can invest in all assets and receives no endowment. We next simplify the budget constraint (16) and the financial constraint (9) by using properties of a symmetric equilibrium.

A first simplifying property is that $\phi_{i,t} = 0$ for assets that are not part of active opportunities. This property holds in equilibrium, as we explain here and show formally in Sections 3 and 4. An implication of this property is that $\mathrm{E}_t(\phi_{i,t+1}) = \Phi_{i,t} = 0$ since the probability of an opportunity becoming active is arbitrarily small. Since $\Phi_{i,t} = 0$, investing in assets that are not part of active opportunities exposes arbitrageurs to risk that is not compensated in terms of expected excess return. Investing in those assets also tightens the financial constraint (9). Hence, the optimal position is zero. Outside investors' optimal position is also zero because they have a zero endowment and hence would hold a non-zero position only if the expected excess return were non-zero. Therefore, the markets for assets that are not part of active opportunities clear with a zero price discount, confirming our conjecture that $\phi_{i,t} = 0$. Using that property as well as (12), (13), $\epsilon_{i,t+1} = \epsilon_{-i,t+1}$,

$\eta_{i,t+1} = -\eta_{i,t+1}$, and $\phi_{i,s} = -\phi_{-i,s}$ for $s = t, t+1$, we can write the budget constraint (16) as

$$W_{t+1} = (1+r)W_t + \sum_{i \in A_t} (x_{i,t} - x_{-i,t}) [\Phi_{i,t} + \eta_{i,t+1} + \mathrm{E}_t(\phi_{i,t+1}) - \phi_{i,t+1}] + \sum_{i \in A_t} (x_{i,t} + x_{-i,t}) \epsilon_{i,t+1} - c_{t+1}, \tag{17}$$

and the financial constraint (9) as

$$W_t \geq \sum_{i \in A_t} \left[\max_{\{\epsilon_{j,t+1}, \eta_{j,t+1}\}_{j \in I}} \frac{x_{i,t}[-\Phi_{i,t} - \epsilon_{i,t+1} - \eta_{i,t+1} - \mathrm{E}_t(\phi_{i,t+1}) + \phi_{i,t+1}]}{1+r} + \max_{\{\epsilon_{j,t+1}, \eta_{j,t+1}\}_{j \in I}} \frac{x_{-i,t}[\Phi_{i,t} - \epsilon_{i,t+1} + \eta_{i,t+1} + \mathrm{E}_t(\phi_{i,t+1}) - \phi_{i,t+1}]}{1+r} \right]. \tag{18}$$

Two further simplifying properties are that $\phi_{i,t+1}$ is independent of $\epsilon_{i,t+1}$ and that $x_{i,t}$ and $x_{-i,t}$ must have opposite signs. The first property holds in equilibrium, as we show in Sections 3 and 4. Intuitively, when arbitrageurs hold opposite positions in assets i and $-i$, their wealth W_{t+1} is independent of $\epsilon_{i,t+1}$ and the same is true of spreads, which depend on wealth. The second property follows from arbitrageurs' optimization. Indeed, if $x_{i,t}$ and $x_{-i,t}$ had the same sign, then an arbitrageur would be able to reduce both in absolute value while holding $x_{i,t} - x_{-i,t}$ constant. That would reduce his risk without affecting his expected excess return, as can be seen from the budget constraint (17), and would relax the financial constraint (18). Using the two simplifying properties, we can write (18) as

$$W_t \geq \sum_{i \in A} \frac{(|x_{i,t}| + |x_{-i,t}|)\epsilon_i + 2\max_{\{\eta_{i',t+1}\}_{i'} \in I} \{(x_{i,t} - x_{-i,t}) [-\Phi_{i,t} - \eta_{i,t+1} - \mathrm{E}_t(\phi_{i,t+1}) + \phi_{i,t+1}]\}}{1+r}. \tag{19}$$

A final simplifying property is that $x_{i,t}$ and $x_{-i,t}$ must be (exact) opposites. Indeed, if $x_{i,t} + x_{-i,t} \neq 0$, then an arbitrageur could set $x_{i,t} + x_{-i,t} = 0$ while holding $x_{i,t} - x_{-i,t}$ constant. That would eliminate his exposure to $\epsilon_{i,t+1}$ without affecting his expected excess return, his exposure to
$\eta_{i,t+1}$ and $\phi_{i,t+1}$, and the financial constraint (19). Using this property, we can simplify the budget constraint (17) to

$$W_{t+1} = (1+r)W_t + 2 \sum_{i \in A_t} x_{i,t} [\Phi_{i,t} + \eta_{i,t+1} + \mathrm{E}_t(\phi_{i,t+1}) - \phi_{i,t+1}] - c_{t+1}, \tag{20}$$

and the financial constraint (19) to

$$W_t \geq 2 \sum_{i \in A} \frac{|x_{i,t}| \epsilon_i + \max_{\{\eta_{j,t+1}\}_{j \in I}} \{x_{i,t}[-\Phi_{i,t} - \eta_{i,t+1} - \mathrm{E}_t(\phi_{i,t+1}) + \phi_{i,t+1}]\}}{1+r}. \tag{21}$$

The arbitrageur's optimization problem reduces to choosing positions in assets $i \in A_t$, 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 conjecture that the value function of an arbitrageur in period t is

$$V_t(W_t) = B \log(W_t) + G_t, \tag{22}$$

where G_t is a possibly stochastic function and B is a constant.

3 Riskless Arbitrage

In this section we solve for equilibrium when assets in each pair have identical payoffs ($\eta_i = 0$). With identical payoffs, arbitrageur wealth W_t does not depend on the payoff realizations because arbitrageurs hold opposite positions in the assets in each pair. Hence, the return that arbitrageurs earn from a period to the next is riskless. That riskless return, however, could change stochastically over time. We rule out stochastic variation by assuming that the set

$$C_t \equiv \{(\epsilon_i, \eta_i, u_i, \mu_i, h_i - t) : i \in A_t\}$$

describing the characteristics of active opportunities is deterministic. Thus, while arbitrageurs are uncertain as to which opportunities will become active, they know what their overall return will be. With a deterministic C_t, the dynamics of arbitrageur wealth, arbitrageur positions, and spreads are deterministic. With deterministic spreads, the expected excess return of asset i simplifies to

$$\Phi_{i,t} = (1+r)\phi_{i,t} - \phi_{i,t+1}. \tag{23}$$

One setting that yields a deterministic C_t and that we emphasize later is as follows. The universe I of risky assets is divided into $2N$ disjoint families I_n for $n = 1, .., N$, with the assets in each family forming a continuum and having the same characteristics $(\bar{\epsilon}_i, \bar{\eta}_i, u_i, \mu_i, M_i)$. 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, C_t is not only deterministic, but constant.

The case of riskless arbitrage is a natural benchmark. It is highly tractable and yields useful results and intuitions, which also facilitate the analysis of risky arbitrage in Section 4. We start by deriving the first-order conditions of outside investors and arbitrageurs in an equilibrium of the conjectured form, i.e., symmetric with deterministic price discounts. We then impose market clearing and show that such an equilibrium exists.

3.1 First-Order Conditions

3.1.1 Outside Investors

Since $\eta_{i,t+1} = 0$ and $\phi_{i,t+1}$ is deterministic, the budget constraint (14) of an i-investor simplifies to

$$w_{i,t+1} = (1+r)w_{i,t} + y_{i,t}\Phi_{i,t} + (y_{i,t}+u_{i,t})\epsilon_{i,t+1} - c_{i,t+1}. \tag{24}$$

The only risk borne by the investor between periods t and $t+1$ is the payoff shock $\epsilon_{i,t+1}$, and the investor's exposure to that risk is $y_{i,t}+u_{i,t}$.

Proposition 1. *The value function of an i-investor in period t is given by (15), where $A = r\alpha$ and $F_{i,t}$ is deterministic. The investor's optimal position in asset i is given by the first-order condition*

$$\Phi_{i,t} - \bar{\epsilon}_i f'[(y_{i,t}+u_{i,t})\epsilon_i] = 0, \tag{25}$$

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

$$\exp\left[\frac{\alpha A f(y)}{\alpha + A}\right] \equiv E\left[\exp\left(\frac{\alpha A y \epsilon_{i,t}}{(\alpha+A)\bar{\epsilon}_i}\right)\right]. \tag{26}$$

The first-order condition (25) takes an intuitive form. The first term in the left-hand side, $\Phi_{i,t}$, is the expected excess return of asset i. The second term, $-\bar{\epsilon}_i f'[(y_{i,t}+u_{i,t})\epsilon_i]$, is a risk adjustment, reflecting the investor's risk from holding the position. Since the function $f(y)$ is convex, as shown in Lemma 1, the risk adjustment is increasing in the investor's exposure $y_{i,t}+u_{i,t}$. The investor's first-order condition amounts to setting the risk-adjusted expected excess return that she derives from asset i to zero. This 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 $p_{i,t}$. The function $\bar{\epsilon}_i f'[(y_{i,t}+u_{i,t})\epsilon_i]$ can be interpreted as a pricing function, which yields the expected excess return $\Phi_{i,t}$ as a function of the position $y_{i,t}$ of the "marginal investor."

Lemma 1. *The function $f(y)$ is non-negative, symmetric around the vertical axis ($f(-y)=f(y)$),*

and strictly convex. It also satisfies $f'(-y) = -f'(y), f'(0) = 0$, *and* $\lim_{y\to\infty} f'(y) = 1$.

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 around the vertical axis follows because $\epsilon_{i,t}$ is distributed symmetrically around zero. Symmetry implies $f'(-y) = -f'(y)$ and $f'(0) = 0$.

The first-order condition of $-i$-investors yields an optimal position that is the opposite to that of i-investors. This follows from (25) and the observations that price discounts, expected excess returns, and endowment shocks are opposites for assets i and $-i$, and that $f'(y) = -f'(-y)$.

3.1.2 Arbitrageurs

Since $\eta_{i,t+1} = 0$ and $\phi_{i,t+1}$ is deterministic, the budget constraint (20) of an arbitrageur simplifies to

$$W_{t+1} = (1+r)W_t + 2\sum_{i\in A_t} x_{i,t}\Phi_{i,t} - c_{t+1}, \tag{27}$$

and the financial constraint (21) simplifies to

$$W_t \geq 2\sum_{i\in A_t} \left|\frac{x_{i,t}\bar{\epsilon}_i}{1+r} - \frac{x_{i,t}\Phi_{i,t}}{1+r}\right|. \tag{28}$$

Eq. (27) confirms that the dynamics of arbitrageur wealth are deterministic. The per-share return of an active opportunity i is $2\Phi_{i,t}$, i.e., twice the expected excess return $\Phi_{i,t}$ of asset i. This return is non-negative in equilibrium, as we show in Section 3.2. While i-investors earn $\Phi_{i,t}$ as compensation for risk, arbitrageurs earn it riskfree because they can combine a position in asset i with one in asset $-i$. Thus, when $\Phi_{i,t} > 0$, arbitrageurs can earn a riskless return above the riskless rate r.

Eq. (28) shows that the collateral required to hold a position in an active opportunity i is larger when asset payoffs are more volatile, i.e., $\bar{\epsilon}_i$ is larger, and when the opportunity offers a lower return, i.e., $\Phi_{i,t}$ is smaller. In both cases this is because the maximum loss of the position is larger.

Proposition 2. *The value function of an arbitrageur in period t is given by (22), where* $B = \frac{\beta}{1-\beta}$ *and* $G_{i,t}$ *is deterministic. The arbitrageur's optimal consumption is*

$$c_t = \frac{1-\beta}{\beta} W_t. \tag{29}$$

- *If all active opportunities offer a zero return, i.e., $\Phi_{i,t} = 0$ for all $i \in A_t$, then the arbitrageur is indifferent between any combination of positions in these opportunities.*

- *If instead some active opportunities offer a positive return and the remainder a zero return, then the arbitrageur holds non-zero positions only in opportunities with maximum return per unit of collateral:*

$$i \in \operatorname{argmax}_{j \in A_t} \frac{\Phi_{j,t}}{\bar{\epsilon}_j - \Phi_{j,t}}. \tag{30}$$

For these opportunities, positions are long in assets $i \in A_t$, i.e., those with positive endowment shocks. Moreover, the financial constraint (28) binds.

The arbitrageurs' optimal investment policy can be derived intuitively as follows. Substituting the optimal consumption (29) into the budget constraint (27), we can write the latter as

$$W_{t+1} = \beta \left[(1+r)W_t + 2 \sum_{i \in A_t} x_{i,t}\Phi_{i,t} \right]. \tag{31}$$

Since assets $i \in A_t$ offer non-negative expected excess returns, arbitrageurs do not benefit from shorting them. Therefore, we can write the financial constraint (28) as

$$W_t \geq 2 \sum_{i \in A_t} \frac{x_{i,t}(\epsilon_i - \Phi_{i,t})}{1+r}. \tag{32}$$

Maximizing W_{t+1} in (31) subject to (32) and $x_{i,t} \geq 0$ is a simple linear-programming problem. Arbitrageurs invest only in those opportunities that offer the maximum return $\Phi_{i,t}$ per unit of collateral $\epsilon_i - \Phi_{i,t}$. Moreover, when some opportunities offer a non-zero return, arbitrageurs "max out" their financial constraint (32) because they can earn a riskless return above 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.

3.2 Equilibrium

3.2.1 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

$$T_t \equiv \{i \in A_t : x_{i,t} > 0\},$$

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

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

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

$$T_t = \{i \in A_t : f'(u_i\bar{\epsilon}_i) > \Pi_t\}.$$

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

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

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 (25), $f'(u_i\bar{\epsilon}_i)$ is the return per unit of volatility that opportunity i would offer in the absence of arbitrageurs. Arbitrageurs trade only 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 volatility ϵ_i. 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 contagion effects derived in Section 4. Suppose, for example, that arbitrageurs experience losses in opportunity 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.

3.2.2 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. *In equilibrium, arbitrageur wealth evolves according to*

$$W_{t+1} = \beta \frac{1+r}{1-\Pi_t} W_t. \tag{33}$$

- *If* $W_t > W_{c,t} \equiv \frac{2}{1+r} \sum_{i \in A_t} \mu_i u_i \bar{\epsilon}_i$, *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 zero.*

- *If* $W_t < W_{c,t}$, *then the financial constraint binds, and arbitrageurs earn a riskless return above the riskless rate* r. *The return* Π_t *per unit of volatility offered by all traded opportunities is the unique positive solution of*

$$2\frac{1-\Pi_t}{1+r} \sum_{i \in T_t} \mu_i \left[u_i \bar{\epsilon}_i - (f')^{-1}(\Pi_t) \right] = W_t, \tag{34}$$

  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 $i \in A_t$. This happens when arbitrageurs fully absorb the endowment shocks of outside investors, i.e., $x_{i,t} = \mu_i u_i$ for all $i \in A_t$. Setting $\Phi_{i,t} = 0$ and $x_{i,t} = \mu_i u_i$ in (32), 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 above 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 for 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 shows that an equilibrium with all the properties conjectured or shown in this section exists.

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

3.3 Steady State and Convergence Dynamics

We next derive a steady state by specializing our model to the stationary "asset family" setting described at the beginning of Section 3. This steady state is deterministic, and used as a basis for constructing a stochastic steady state in Section 4. We index the $2N$ families by $n \in \{-N, .., -1, 1, .., N\}$, with the convention that for an asset in family n the other asset in its pair belongs to family $-n$, and that families $n = 1, .., N$ comprise the assets with the positive endowment shocks. We denote by $(\bar{\epsilon}_n, \bar{\eta}_n, u_n, \mu_n, M_n)$ the characteristics $(\bar{\epsilon}_i, \bar{\eta}_i, u_i, \mu_i, M_i)$ for all assets i in family n. (As in the rest of Section 3, we assume $\bar{\eta}_i = 0$.) The set of active opportunities in period t is

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

Opportunity (n, m) consists of one asset in family $n \in \{1, .., N\}$ 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. (10) and (25)). Hence, we index these quantities by the family subscript, n or $-n$, and the time subscript, t. The price discounts 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

$$T_t = \{(n, m) : n \in N_t, m \in \{1, .., M_n\}\},$$

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

Proposition 6. *In equilibrium, 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 $\bar{\Pi} \equiv \max_{n=1,\ldots,N} f'(u_n \bar{\epsilon}_n) < 1$, then W_t decreases toward $W = 0$ and Π_t increases toward $\Pi = \bar{\Pi}$.*
- *Otherwise, the steady-state values are given by*

$$W = 2\frac{1-\Pi}{1+r} \sum_{n \in N} \mu_n M_n \left[u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) \right] \in (0, W_c), \tag{35}$$

$$\Pi = 1 - \beta(1+r) \in (0, \bar{\Pi}), \tag{36}$$

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 exceeds the rate at which they consume. Arbitrageurs earn the riskless (net) return $\frac{1+r}{1-\Pi_t} - 1$, and consume at the rate $\frac{1-\beta}{\beta}$ as shown in (29).

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 riskless return above r. Decreases in W_t within that region raise Π_t and hence raise arbitrageurs' return, which 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 (25), we find that the return per unit of volatility from an active opportunity $(i, -i)$ in the absence of arbitrageurs is equal to $f'(u_i\epsilon_i)$. Therefore, the maximum value of Π_t is $\overline{\Pi} \equiv \max_{n=1,\ldots,N} f'(u_n\epsilon_n)$. 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 $\overline{\Pi}$, are constant over time.

The dynamics of wealth in the stationary case follow from the above observations. When $\beta(1 + r) > 1$, arbitrageurs consume at a rate smaller than the riskless rate. Hence, their wealth increases over time even if $\Pi_t = 0$, i.e., all active opportunities offer a zero return, and becomes arbitrarily large. When instead $\beta(1 + r) < 1 - \overline{\Pi}$, arbitrageurs consume at a rate larger than the maximum return that their trades can offer. Hence, their wealth decreases over time and converges to zero.

In the intermediate case $1 - \overline{\Pi} < \beta(1+r) < 1$, 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 close to their maximum return, and wealth increases because $1 - \overline{\Pi} < \beta(1 + r)$. 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.

4 Risky Arbitrage

Most real-life arbitrage involves some risk. To introduce arbitrage risk in our model, we allow assets in each pair to have non-identical payoffs ($\eta_i > 0$). The return that arbitrageurs earn from one period to the next is then risky and arbitrageur wealth W_t is stochastic. We look for a symmetric equilibrium in which W_t is the only stochastic state variable. We maintain the assumption that the set C_t describing the characteristics of active opportunities is deterministic. This prevents characteristics from becoming additional state variables. Because arbitrageurs hold opposite positions in

the two assets in each pair, W_t does not depend on the realizations of $\epsilon_{i,t}$ but only on those of $\eta_{i,t}$. The same is true for positions and spreads.

We start by deriving the first-order conditions of outside investors and arbitrageurs in an equilibrium of the conjectured form. We then specialize our analysis to the case where arbitrage risk is small (η_i small), and compute the equilibrium in closed form. Using our analytical solution, we study how spreads, expected excess returns, and arbitrageur positions depend on cross-sectional characteristics, how they respond to shocks, and whether the mobility of arbitrage capital across markets makes them more stable.

4.1 First-Order Conditions

4.1.1 Outside Investors

An i-investor bears more risk than in the riskless-arbitrage case ($\eta_i = 0$). This is because asset i's payoff includes the additional component $\eta_{i,t+1}$, and because the asset's price discount $\phi_{i,t+1}$ varies stochastically following changes in arbitrageur wealth. As in the riskless-arbitrage case, the investor sets the risk-adjusted expected excess return that she derives from asset i to zero. The risk adjustment, however, includes a term reflecting the additional risk. This is the third term in the left-hand side of (37).

Proposition 7. *The value function of an i-investor in period t is given by (15), where $A = r\alpha$ and $F_{i,t}$ is a function of W_t. The investor's optimal position in asset i is given by the first-order condition*

$$\Phi_{i,t} - \bar{\epsilon}_i f'\left[(y_{i,t}+u_{i,t})\epsilon_i\right] + E_t\left[\frac{M_{i,t+1}}{E_t[M_{i,t+1}]}\left(\eta_{i,t+1}+E_t(\phi_{i,t+1})-\phi_{i,t+1}\right)\right] = 0, \tag{37}$$

where

$$M_{i,t+1} \equiv \exp\left(-\frac{\alpha\{A\left[(y_{i,t}+u_{i,t})\eta_{i,t+1}-y_{i,t}\phi_{i,t+1}\right]+F_{i,t+1}\}}{\alpha+A}\right).$$

4.1.2 Arbitrageurs

An arbitrageur bears the risk represented by $\eta_{i,t+1}$ and $\phi_{i,t+1}$. Because of that risk, he may benefit from shorting assets $i \in A_t$, and so may amplify the price discrepancies induced by outside investors' endowment shocks. Short positions in assets $i \in A_t$ may be beneficial even though they earn negative expected excess returns because they hedge long positions in other such assets. In the

rest of this section we assume that short positions in assets $i \in A_t$ do not arise in equilibrium. This assumption is satisfied, for example, when arbitrage risk is small and the expected excess returns $\Phi_{i,t}$ of assets $i \in A_t$ exceed a (strictly) positive bound. This result is shown in Proposition 8 and extends a result shown in Proposition 2 under riskless arbitrage.

When arbitrageurs hold long positions in an asset $i \in A_t$, an increase in $\eta_{i,t+1}$ raises their wealth and hence lowers the spread $\phi_{i,t+1}$ (which is positive). When instead arbitrageurs hold no position in the asset, $\phi_{i,t+1}$ is independent of $\eta_{i,t+1}$. In either case, the minimum value $-\bar{\eta}_i$ of $\eta_{i,t+1}$ can be realized jointly with the maximum value of $\phi_{i,t+1}$, which we denote by $\bar{\phi}_{i,t+1}$. And conversely, the maximum value $\bar{\eta}_i$ of $\eta_{i,t+1}$ can be realized jointly with the minimum value of $\phi_{i,t+1}$, which we denote by $\underline{\phi}_{i,t+1}$. Using these observations, we can simplify the arbitrageurs' financial constraint (21) because we can determine the maximum loss scenario. A long position in asset $i \in A_t$ suffers its maximum loss when $\eta_{i,t+1} = -\bar{\eta}_i$ and $\phi_{i,t+1} = \bar{\phi}_{i,t+1}$ (maximum spread and hence maximum price discount for asset i). And conversely, a short position in the asset suffers its maximum loss when $\eta_{i,t+1} = \bar{\eta}_i$ and $\phi_{i,t+1} = \underline{\phi}_{i,t+1}$. Eq. (21) becomes

$$W_t \geq 2 \sum_{i \in A_t} \frac{|x_{i,t}|\,(\epsilon_i + \eta_i) + x_{i,t}^{+}\left[\underline{\phi}_{i,t+1} - \mathrm{E}_t(\phi_{i,t+1})\right] - x_{i,t}^{-}\left[\mathrm{E}_t(\phi_{i,t+1}) - \bar{\phi}_{i,t+1}\right] - x_{i,t}\Phi_{i,t}}{1+r}, \quad (38)$$

where $x_{i,t}^{+} = \max\{x_{i,t}, 0\}$ and $x_{i,t}^{-} = \min\{x_{i,t}, 0\}$. Moreover, when $x_{i,t} \geq 0$ for all $i \in A_t$, (38) simplifies to

$$W_t \geq 2 \sum_{i \in A_t} \frac{x_{i,t}\left[\epsilon_i + \eta_i + \bar{\phi}_{i,t+1} - \mathrm{E}_t(\phi_{i,t+1}) - \Phi_{i,t}\right]}{1+r}. \quad (39)$$

The constraint is more stringent than in the riskless-arbitrage case because of the additional risks that the arbitrageur is subject to.

Proposition 8. *The value function of an arbitrageur in period t is given by (22), where* $B = \frac{\beta}{1-\beta}$ *and* G_t *is a function of aggregate arbitrageur wealth. The arbitrageur's optimal consumption is given by (29). When short positions in assets* $i \in A_t$ *are not optimal, as is the case when* $\min_{i \in A_t} \Phi_{i,t}$ *exceeds a positive bound and* $\max_{i \in A_t}\{\bar{\eta}_i, \bar{\phi}_{i,t+1} - E_t(\phi_{i,t+1})\}$ *is small,*

- *The arbitrageur's optimal positions are non-zero only in opportunities with maximum risk-*

adjusted return per unit of collateral

$$i \in \operatorname{argmax}_{j \in A_t} \frac{\Phi_{j,t} + E_t\left[M_{t+1} \left(\eta_{j,t+1} + E_t(\phi_{j,t+1}) - \phi_{j,t+1} \right) \right]}{\bar{\epsilon}_j + \eta_j + \phi_{j,t+1} - E_t(\phi_{j,t+1}) - \Phi_{j,t}}, \tag{40}$$

where

$$M_{t+1} \equiv \frac{1}{(1+r)W_t + 2\sum_{j \in A_t} x_{j,t}[\Phi_{j,t} + \eta_{j,t+1} + E_t(\phi_{j,t+1}) - \phi_{j,t+1}]}.$$

For these opportunities, positions are long in assets $i \in A_t$.

- *The financial constraint (39) is slack if the maximum risk-adjusted return per unit of collateral is zero, and binds if it is positive.*

As in the riskless-arbitrage case, the arbitrageur invests only in those opportunities that offer the maximum return per unit of collateral. When arbitrage is risky, however, that return includes a risk adjustment that depends on arbitrageur risk aversion and position size. The risk adjustment for arbitrageurs is different than for outside investors because market segmentation implies that arbitrageurs bear different risks.

The arbitrageur's portfolio problem combines two aspects: allocate scarce capital to the most profitable opportunities, and trade off risk and return. When the financial constraint is slack, only the second aspect is present. The arbitrageur invests in any given arbitrage opportunity i to take advantage of its positive expected excess return $2\Phi_{i,t}$. As he increases his position in that opportunity, he bears more risk, and hence his risk-adjusted return decreases. The optimal position renders the risk-adjusted return equal to zero. This is the standard first-order condition of a risk-averse investor.[15]

If the optimal positions under a slack financial constraint violate that constraint, then the first aspect of the portfolio problem kicks in. To meet the financial constraint, the arbitrageur scales down his positions. As a consequence, he bears less risk and risk-adjusted returns become positive. Positions are scaled down until the constraint is met, and in such a way as to maintain risk-adjusted

[15]The standard first-order condition in a one-period setting is

$$\mathrm{E}(MR) = 0 \Leftrightarrow \mathrm{E}(R) + \mathrm{E}\left(\frac{M}{\mathrm{E}[M]}[R - \mathrm{E}(R)] \right) = 0,$$

where R is an asset's excess return over the riskless asset and M is the pricing kernel. The risk adjustment is $\mathrm{E}\left(\frac{M}{\mathrm{E}[M]}[R - \mathrm{E}(R)] \right)$. It depends on the size of the position, which is why the optimal position is interior. (That is, if the investor invests only in one asset, the risk-adjusted return of that asset declines below that of other assets, inducing the investor to also invest in the other assets.)

returns per unit of collateral equal across all opportunities for which positions remain positive.

The two aspects of the arbitrageur's portfolio problem can be completely separated in two special cases. When asset payoffs are identical, only the first aspect is present because the arbitrageur faces no risk. When instead the common payoff shock $\epsilon_{i,t+1}$ to both assets in a pair $(i, -i)$ is zero for all pairs ($\bar{\epsilon}_i = 0$), only the second aspect is present: the arbitrageur determines his positions by trading off risk and return, and the financial constraint is always slack.

Corollary 1. *When $\bar{\epsilon}_i = 0$ for all $i \in A_t$, and short positions in assets $i \in A_t$ are not optimal, the arbitrageurs' financial constraint is slack.*

The intuition for Corollary 1 is that when the common shock is not present, a long position in asset i and a short position in asset $-i$ can achieve their maximum loss at the same time. (That is not possible in the common shock's presence because the long position achieves its maximum loss when $\epsilon_{i,t+1} = -\bar{\epsilon}_i$, while the short position achieves its maximum loss when $\epsilon_{i,t+1} = \bar{\epsilon}_i$.) Hence, requiring arbitrageurs to have sufficient wealth in each market to cover any liabilities in that market is equivalent to requiring them to be able to cover any overall liability. Moreover, arbitrageurs choose their positions so to keep positive wealth and hence cover any overall liability because logarithmic utility makes zero consumption prohibitively costly.

4.2 Equilibrium for Small Arbitrage Risk

To study the case where arbitrage risk is small, we set $\bar{\eta}_i = \lambda_i \bar{\eta}$ and take $\bar{\eta}$ to be small holding λ_i constant. We also specialize our model to the stationary "asset family" setting described at the beginning of Section 3, and focus on parameters for which the steady state derived in Section 3.3 is interior, i.e., arbitrageur wealth does not converge to zero or infinity. Since arbitrageur wealth W_t converges to the steady-state value W when $\bar{\eta} = 0$, it converges to a stationary probability distribution with support close to W when $\bar{\eta}$ is small. When W_t moves within the support of the stationary distribution, as we assume from now on, the subset N_t of families whose assets are traded in equilibrium could change over time. If we rule out, however, the non-generic case where $f'(u_n \bar{\epsilon}_n) = \Pi$ for some $n \in \{1, .., N\}$, then for $\bar{\eta}$ small enough N_t remains constant over time and so does the set T_t of traded active opportunities. We hence drop the time subscript from both sets. Since the support of the stationary distribution is close to W, expected excess returns $\Phi_{i,t}$ of assets $i \in A_t$ exceed a positive bound. Hence, short positions in assets $i \in A_t$ are not optimal, arbitrageurs' first-order condition is (40), and the financial constraint (39) binds.

We look for spreads, expected excess returns, and positions of arbitrageurs and outside investors

that take the form

$$\phi_{n,m,t} = \phi^0_{n,m,t} + \phi^1_{n,m,t}\bar{\eta} + o(\bar{\eta}), \tag{41}$$

$$\Phi_{n,m,t} = \Phi^0_{n,t} + \Phi^1_{n,m,t}\bar{\eta} + o(\bar{\eta}), \tag{42}$$

$$x_{n,m,t} = x^0_{n,t} + x^1_{n,m,t}\bar{\eta} + o(\bar{\eta}), \tag{43}$$

$$y_{n,m,t} = y^0_{n,t} + y^1_{n,m,t}\bar{\eta} + o(\bar{\eta}), \tag{44}$$

for asset (n,m). The superscript 0 denotes the value of the corresponding variable, as a function of W_t, in the case of riskless arbitrage (Section 3). The superscript 1 denotes a first-order term in $\bar{\eta}$ introduced by risk. As in Section 3.3, we drop the time subscript from zeroth- and first-order terms in (41)-(44) when these are evaluated at the steady-state value W.

We look for dynamics of wealth, within the support of the stationary distribution, that take the form

$$W_{t+1} = W + \rho(W_t - W) + v\bar{\eta} + \sum_{(n,m)\in T} \sigma_n \eta_{n,m-1,t+1} + o(\bar{\eta}), \tag{45}$$

where $(\rho, v, \{\sigma_n\}_{n\in N})$ are constants. We compute these constants, as well as the first-order terms in (42)-(44), in the Appendix (Proposition A.1). We confirm, in particular, that $\rho \in (0,1)$. Thus, to a first-order in $\bar{\eta}$, W_t follows an AR(1) process. The mean to which this process reverts is equal to W plus a first-order term in $\bar{\eta}$. Variation around the mean is caused by the shocks $\eta_{n,m-1,t+1}$ for $(n,m) \in T$. The coefficient σ_n describing how $\eta_{n,m-1,t+1}$ affects W_{t+1} is given by

$$\sigma_n = \frac{2\beta x^0_n}{1 + 2\beta \sum_{(n',m)\in T} x^0_{n'} \frac{d\phi^0_{n',m-1,t+1}}{dW_{t+1}}}, \tag{46}$$

where the derivatives in the denominator are evaluated at $W_{t+1} = W$. The intuition for (46) can be seen from the budget constraint (20). Substituting the optimal consumption (29) and changing asset subscripts from i to (n,m), we can write (20) as

$$W_{t+1} = \beta\left[(1+r)W_t + 2\sum_{(n,m)\in T} x_{n,m,t}\left[\Phi_{n,m,t} + \eta_{n,m-1,t+1} + \mathrm{E}_t(\phi_{n,m-1,t+1}) - \phi_{n,m-1,t+1}\right]\right]. \tag{47}$$

A negative shock $\eta_{n,m-1,t+1}$ reduces the wealth of arbitrageurs through the term $2\beta x_{n,m,t}\eta_{n,m-1,t+1}$.

This is the direct effect of the shock, holding spreads $\phi_{n',m-1,t+1}$ constant, and corresponds to the numerator in (46). There is also an indirect *amplification* effect, operating through a change in spreads: because W_t decreases due to the direct effect, spreads increase, and this amplifies the reduction in W_t. The indirect effect corresponds to the second term in the denominator in (46), which is negative hence lowering the denominator and raising σ_n.

4.3 Economic Implications

4.3.1 Spreads, Returns, and Positions in the Cross-Section

An arbitrage opportunity (n, m) is described by five characteristics: the volatility $\bar{\epsilon}_n$ of the payoff shock that is common to the two assets and which arbitrageurs can hedge via a long-short trade; the volatility η_n of the payoff shock which arbitrageurs cannot hedge away; the endowment shock u_n of outside investors, which determines their relative demand for the two assets; the measure μ_n of outside investors; and the horizon m of the opportunity. We examine how these characteristics affect spreads $\phi_{n,m,t}$, expected excess returns $\Phi_{n,m,t}$, and arbitrageur positions $x_{n,m,t}$.

Proposition 9. *Suppose that arbitrage risk $\bar{\eta}$ is small. Holding constant all other characteristics as well as arbitrageur wealth:*

- *An opportunity with higher hedgeable volatility ϵ_n has a higher spread and expected excess return, and attracts more investment by arbitrageurs.*
- *An opportunity with higher non-hedgeable volatility $\bar{\eta}_n$ has a higher spread and expected excess return, and attracts less investment by arbitrageurs. The same holds for an opportunity with longer horizon m.*
- *An opportunity with larger endowment shock u_n attracts more investment by arbitrageurs. It has a higher spread and expected excess return, except when the comparison is between traded opportunities, in which case spreads and expected excess returns are the same to the first order in $\bar{\eta}$.*
- *An opportunity with larger measure of outside investors μ_n attracts more investment by arbitrageurs.*

Outside investors in more volatile opportunities, i.e., those with higher $\bar{\epsilon}_n$ or $\bar{\eta}_n$, are more eager to share risk. Therefore, those opportunities offer higher expected excess returns in the arbitrageurs' absence. Their expected excess returns remain higher even if arbitrageurs invest in them. Indeed, arbitrageurs invest in opportunities with the highest return per unit of collateral, and more volatile

opportunities require more collateral. In turn, higher returns imply higher spreads. Indeed, spreads are a present value of future expected excess returns discounted at the riskless rate:

$$\phi_{n,m,t} = \mathrm{E}_t\left(\sum_{s=0}^{m-1} \frac{\Phi_{n,m-s,t+s}}{(1+\bar{r})^{s+1}}\right), \tag{48}$$

as can be seen by solving (13) backwards with the terminal condition $\phi_{n,0,t+m} = 0$.[16]

The source of volatility, $\bar{\epsilon}_n$ or $\bar{\eta}_n$, matters for arbitrageurs' positions. When hedgeable volatility is higher (higher $\bar{\epsilon}_n$), positions are larger. When instead non-hedgeable volatility is higher (higher η_n), positions are smaller. The intuition comes from higher volatility having two countervailing effects on arbitrageur positions: it raises positions because expected excess returns increase, but it lowers positions because collateral requirements increase. Expected excess returns in the arbitrageurs' absence are proportional to the return variance, which for small $\bar{\eta}$ is proportional to $\bar{\epsilon}^2_n$ plus second-order terms in $\bar{\eta}$.[17] Collateral requirements are equal to a position's maximum possible loss, which is $-\bar{\epsilon}_n + \eta_n + \bar{\phi}_{n,m-1,t+1} - \mathrm{E}_t(\phi_{n,m-1,t+1}) - \Phi_{n,m,t}$ from (39). The ratio of these quantities (return per unit of collateral) is increasing in $\bar{\epsilon}_n$ but decreasing in η_n. In particular, an increase in η_n has a second-order effect on the variance but a first-order effect on the collateral requirement.

More generally, when arbitrage risk is small, its dominant effect on equilibrium variables is through the financial constraint rather than through risk aversion. Arbitrage risk raises the maximum possible loss of a position by a first-order term in $\bar{\eta}$. The risk adjustments that it induces in (37) and (40), however, are second-order in $\bar{\eta}$, as they involve an expectation of gains and losses weighted by marginal utility. It is because of the effect through the financial constraint that the equilibrium variables in (41)-(44) include a first-order term in $\bar{\eta}$.

The effect of horizon m on spreads follows from (48). Spreads of opportunities with a longer horizon are the present value of a longer sequence of future returns, and hence are higher. The effect of horizon on expected excess returns and arbitrageur positions is more subtle. Spreads of opportunities with a longer horizon are more sensitive to shocks to arbitrageur wealth, as we show in Section 4.3.2. As a consequence, these opportunities require more collateral (term $\bar{\phi}_{n,m-1,t+1} - \mathrm{E}_t(\phi_{n,m-1,t+1})$ in (39)), and higher collateral requirements push expected excess returns up and arbitrageur positions down. Note that horizon has no effect on expected excess returns

[16] Because (13) is linear in spreads and expected excess returns, (48) is exact and not a small-risk approximation.

[17] The return variance can be computed using (12), the independence between $\epsilon_{n,t}$ and $\{\eta_{n',t}\}_{n' \in \mathcal{N}}$, the fact that W_t is the only stochastic state variable affecting spreads, and the dynamics of of W_t in (45). Note that since the effect of arbitrage risk on the return variance is second-order in $\bar{\eta}$, our analysis of the equilibrium for small arbitrage risk would not change if instead of holding $\bar{\epsilon}_n$ constant when $\bar{\eta}_n$ varies, we also vary $\bar{\epsilon}_n$ so that the return variance in the arbitrageurs' absence remains constant. Indeed, the required variation in $\bar{\epsilon}_n$ would be second-order in $\bar{\eta}$.

or positions under riskless arbitrage. The effect instead arises because arbitrage risk impacts collateral requirements, and is included in the first-order terms in (42)-(44).

The effect of the endowment shock u_n on spreads, expected excess returns, and arbitrageur positions is the same as under riskless arbitrage. In particular, spreads and expected excess returns increase with u_n for non-traded opportunities and become flat when u_n is large enough so that opportunities are traded. The flat part arises because traded opportunities offer the same return per unit of collateral and because u_n does not affect the collateral requirement. The effect of the measure μ_n of outside investors on arbitrageur positions is also the same as under riskless arbitrage: arbitrageurs hold larger positions in an opportunity where there are more outside investors.

4.3.2 Response to Shocks

We next examine how spreads, expected excess returns, and arbitrageur positions respond to shocks to arbitrageur wealth, and how these dynamics depend on the characteristics of arbitrage opportunities. We focus on traded opportunities; non-traded opportunities are not affected by shocks.

For any given traded opportunity, spreads, expected excess returns, and arbitrageur positions change over time both because arbitrageur wealth changes and because the opportunity's horizon shortens. We focus on the wealth-induced variation, keeping horizon constant. Hence, we compare spreads, expected excess returns, and arbitrageur positions across opportunities within the same family and with the same horizon (same (n, m)) but at different points in time.

Proposition 10. *Suppose that arbitrage risk $\bar{\eta}$ is small and that a shock in period t lowers arbitrageur wealth below its mean.*

- *The immediate effect is that spreads and expected excess returns increase, and arbitrageurs scale down their positions.*
- *Following this immediate reaction, there is a recovery phase, during which spreads, expected excess returns, and positions are expected to revert gradually toward their original values.*

The dynamics in Proposition 10 follow from the self-correcting dynamics (45) of arbitrageur wealth. Following a shock that lowers wealth in period t, wealth is expected to increase gradually back to its mean. Spreads and expected excess returns are decreasing functions of wealth, and hence increase instantly and then decrease gradually. By contrast, positions are increasing functions of wealth, and hence decrease instantly and then increase gradually.

Proposition 11. *Suppose that arbitrage risk $\bar{\eta}$ is small and that a shock in period t lowers arbitrageur wealth below its mean. Holding constant all other characteristics:*

- *For an opportunity with higher hedgeable volatility $\bar{\epsilon}_n$, the immediate increase in spread and expected excess return is larger, and the immediate decrease in arbitrageur positions is smaller.*
- *For an opportunity with higher non-hedgeable volatility $\bar{\eta}_n$, the immediate increase in spread and expected excess return is larger, and so is the immediate decrease in arbitrageur positions.*
- *For an opportunity with longer horizon m, the immediate increase in spread is larger. The immediate increase in expected excess return and the immediate decrease in arbitrageur positions are also larger if the steady-state value W of arbitrageur wealth W_t is large enough.*

These comparisons remain the same during the recovery phase.

Spreads and expected excess returns of more volatile opportunities, i.e., those with higher $\bar{\epsilon}_n$ or $\bar{\eta}_n$, are more sensitive to changes in arbitrageur wealth. This is because changes in wealth impact the return per unit of collateral, which arbitrageurs equalize across all opportunities in which they invest. The resulting impact on returns, and hence on spreads, is stronger for more volatile opportunities because they require more collateral.

Arbitrageur positions in more volatile opportunities can be more or less sensitive to changes in wealth, depending on the source of volatility. Positions are less sensitive when volatility is higher because of $\bar{\epsilon}_n$, but more sensitive when it is higher because of $\bar{\eta}_n$. The intuition parallels that in Proposition 9. Following a drop in arbitrageur wealth, positions in all opportunities must be scaled down, and the return per unit of collateral is equalized across opportunities at a higher level. The increase in expected excess returns resulting from a cut in arbitrageur positions (i.e., the inverse elasticity of the demand of outside investors) is proportional to the return variance, and hence to $\bar{\epsilon}_n^2$ plus second-order terms in $\bar{\eta}$. The ratio of this quantity to the collateral required $\epsilon_n + \eta_n + \bar{\phi}_{n,m-1,t+1} - \mathrm{E}_t(\phi_{n,m-1,t+1}) - \Phi_{n,m,t}$ is increasing in ϵ_n, and so positions in opportunities with higher $\bar{\epsilon}_n$ must be cut by less to achieve the new higher level of return per unit of collateral. The ratio is instead decreasing in η_n, and so positions in opportunities with higher $\bar{\eta}_n$ must be cut by more. Put differently, by cutting positions in opportunities with higher η_n, arbitrageurs save a larger amount of collateral, and the returns of these opportunities do not increase sufficiently following the cuts to compensate for that.

Spreads of opportunities with a longer horizon are more sensitive to changes in wealth because they depend on a longer sequence of future returns, all of which are sensitive to wealth changes. Because their spreads are more sensitive to changes in wealth, collateral requirements are higher for long-horizon opportunities. In turn, this implies that expected excess returns and arbitrageur positions for long-horizon opportunities tend to be more sensitive to wealth changes. At the same time, wealth changes affect the required collateral as a function of horizon because they affect the

volatility of equilibrium prices. If the incremental collateral required by longer-horizon opportunities decreases following a wealth drop, then expected excess returns and arbitrageur positions for these opportunities can be less sensitive to wealth changes. This possibility is ruled out if W is large enough.

Taken together, the results of Sections 4.3.1 and 4.3.2 imply an increasing cross-sectional relationship between spreads and expected excess returns on one hand, and betas with respect to arbitrageur wealth on the other. Indeed, Proposition 9 shows that spreads and expected excess returns are higher for more volatile opportunities and for opportunities with a longer horizon. Moreover, Proposition 11 shows that spreads for the same opportunities increase more following a drop in wealth. Hence, their realized returns decrease more and their wealth betas are higher.

The cross-sectional relationship implied by our model is consistent with recent empirical evidence. Avdjiev, Du, Koch, and Shin (2016) study the behavior of the cross-currency basis, which is defined in Section 2.2.3 and is an arbitrage spread associated with violations of covered interest parity. They find that currencies with a higher basis are also those for which the basis increases more when the dollar exchange rate increases. Moreover, increases in the exchange rate are associated with tighter funding conditions of currency-market arbitrageurs. Cho (2016) studies the relationship between alphas of a wide cross-section of stock-market anomalies and the sensitivities of these anomalies' returns to the funding conditions of broker-dealers. He finds no relationship before 1993. After 1993, however, anomalies with higher alphas are also ones that yield lower returns when funding conditions tighten. These findings are consistent with arbitrageurs becoming more active in trading the anomalies in the more recent sample.[18]

We show that an increasing cross-sectional relationship between spreads/returns and wealth betas arises because of two mutually reinforcing mechanisms. High collateral requirements cause spreads and expected excess returns to be high (through the equalization of return per unit of collateral) and wealth betas to be high (through the dynamics implied by same channel). Moreover, high wealth betas cause collateral requirements to be high (because the maximum possible loss of a position increases) and feed back into high spreads and expected excess returns. Exogenous characteristics can "activate" either of the two mechanisms, setting off the mutually reinforcing cycle. Payoff-shock volatility (ϵ_n and η_n) activates the first mechanism, as higher volatility pushes up collateral requirements. Horizon (m) activates the second mechanism, as a longer horizon renders spreads more sensitive to wealth. Versions of the first mechanism have been shown in a number of papers (e.g., Gromb and Vayanos (2002), Geanakoplos (2003), Brunnermeier and Pedersen (2009),

[18] Cho (2016) also finds that anomalies with higher profitability before 1993 tend to have higher wealth betas after 1993. This is consistent with arbitrageurs allocating more wealth to the more profitable anomalies: anomalies with small u_n do not attract any investment and are not sensitive to changes in wealth, while the opposite holds when u_n exceeds a threshold.

Garleanu and Pedersen (2011), and Brumm, Grill, Kubler, and Schmedders (2015)). The second mechanism is new to our model.

4.4 Mobility of Arbitrage Capital

Finally, we use our model to study how the mobility of arbitrage capital affects market stability. Our maintained assumption so far is that all arbitrageurs can trade all assets and hence arbitrage capital is fully mobile across markets. We contrast full mobility to the case where the assets in each family pair $(n, -n)$ for $n = 1, .., N$ are traded by a separate set of arbitrageurs. We refer to the former case as integrated arbitrage markets and to the latter as segmented arbitrage markets. These notions of integration and segmentation are distinct from the asset-level segmentation concerning outside investors and collateral requirements, which we continue to assume. Integration of arbitrage markets could be triggered, for example, by a deregulation of international capital flows.

When arbitrage is riskless, integration and segmentation of arbitrage markets yield the same steady state. Indeed, 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 2. *Suppose that $\bar{\eta} = 0$ (riskless arbitrage). In steady state, integration of arbitrage markets has no effect on spreads and returns.*

When arbitrage is risky, arbitrageurs strictly prefer diversifying across arbitrage markets as long as risks are imperfectly correlated. We assume that risks are independent across markets, i.e., the shocks $\eta_{n,m,t}$ are independent across n. But while diversification is beneficial for any given arbitrageur assuming that others do not diversify, it has a countervailing effect when they all diversify: it induces correlation between arbitrage markets, which makes diversification less effective. This is because arbitrageurs who hold positions in multiple markets react to negative shocks in one market by cutting positions in all markets—a contagion effect.

To examine whether diversification remains effective despite the correlation that it induces, we compute the variance of arbitrageur wealth. We also compute the variance of spreads, which can be interpreted a measure of market stability. We first compare integration and segmentation

when arbitrage opportunities are symmetric, i.e., $(\bar{\epsilon}_n, \bar{\eta}_n, \mu_n, u_n, M_n)$ is independent of n, and then consider the asymmetric case.

Proposition 12. *Suppose that arbitrage risk $\bar{\eta}$ is small and that arbitrage opportunities are symmetric.*

- *The variance of each arbitrageur's wealth under integration is $\frac{1}{N}$ times that under segmentation. It is also equal to the wealth variance of an arbitrageur who diversifies across markets when other arbitrageurs do not.*
- *The variance of each spread under integration is $\frac{1}{N}$ times that under segmentation.*

An arbitrageur who diversifies across markets when other arbitrageurs do not enjoys a reduction in the variance of his wealth by a factor of N. This follows from a standard result in portfolio theory: optimal diversification across N assets with *i.i.d.* returns results in a variance that is N times lower than without diversification. Surprisingly, diversification lowers the variance by a factor of N even when all other arbitrageurs diversify and hence markets become correlated. This is because while spreads become perfectly correlated across markets, their variance is divided by N.

The intuition why the variance of the spreads is divided by N under integration is as follows. Since the aggregate position of arbitrageurs in each market is the same under integration as under segmentation when arbitrage is riskless, the same holds for small arbitrage risk to the highest order in $\bar{\eta}$. Hence, a negative shock $\eta_{n,m,t}$ causes the same drop in aggregate arbitrageur wealth under integration as under segmentation, holding spreads constant. Under segmentation, only arbitrageurs in market n are affected, and they cut their positions in that market. Under integration, all arbitrageurs are affected, and they cut their positions in all markets. Because the drop in aggregate wealth is the same in both cases and because markets are symmetric, the cut in each market under integration is $\frac{1}{N}$ times the cut in market n under segmentation. Hence, spreads in all markets under integration increase by $\frac{1}{N}$ times the increase in the spreads in market n under segmentation. Moreover, the contribution of the shock $\eta_{n,m,t}$ to the variance of the spreads in market n under integration is $\frac{1}{N^2}$ of its contribution under segmentation. Because, however, N times as many shocks (i.e., all shocks) contribute to the variance under integration, the variance of the spreads is $N \times \frac{1}{N^2} = \frac{1}{N}$ times that under segmentation.

Proposition 13. *Suppose that arbitrage risk $\bar{\eta}$ is small and that arbitrage opportunities are asymmetric. Spreads of traded opportunities $n \in N$ for which $\frac{u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)}{\max_{n' \in N}[u_{n'} \bar{\epsilon}_{n'} - (f')^{-}(\Pi)]}$ or $\frac{\bar{\eta}_n}{\max_{n' \in N} \bar{\eta}_{n'}}$ are sufficiently small have higher variance under integration than under segmentation.*

Integration can raise the variance of spreads of opportunities with small endowment shocks (small u_n) or low volatility (small ϵ_n or η_n). For example, because arbitrageurs hold small positions in opportunities with small u_n, payoff shocks $\eta_{n,m,t}$ have small effects on their wealth and on spreads, resulting in low variances under segmentation. Instead, under integration, these opportunities are exposed to shocks coming from other markets, so variances can increase.

5 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 price discrepancies between assets traded in segmented markets, but face financial constraints limiting their ability to do so. We compute the equilibrium in closed form when arbitrage is riskless and when arbitrage risk is small. In the latter case, arbitrage capital follows AR(1) dynamics in a stochastic steady state. We determine how arbitrageurs allocate their limited capital across mispriced assets in steady state, and how this allocation changes following shocks to capital. We also determine which arbitrage trades offer the highest expected returns and how these returns relate to the trades' sensitivity to arbitrage capital and other characteristics. We finally examine how the diversification of arbitrageurs across markets affects the risk that they bear and the volatility of spreads.

When arbitrage risk is small, its first-order effect on equilibrium variables is through the financial constraint rather than through risk aversion. Hence, our results on how characteristics such as volatility, horizon, and investor demand affect expected returns and arbitrageur positions are driven by the characteristics' effects on the financial constraint. For general arbitrage risk, risk aversion would come into play, and its effects might differ from those of the financial constraint. Determining the combined effects of risk aversion and the financial constraint on the cross-section of expected returns and arbitrageur positions is a natural extension of our research. This would require computing the equilibrium for general arbitrage risk. Analysis of this equilibrium would shed light on the role of arbitrage capital as a risk factor, and on how the significance of this factor is affected by the capitalization and diversification of arbitrageurs.

APPENDIX

A Proofs

Proof of Proposition 1: The proposition follows from Proposition 7 by setting $\eta_i = 0$ and noting that arbitrageur wealth and spreads are deterministic. ∎

Proof of Lemma 1: To prove the properties in the lemma, we set $\hat{\alpha} \equiv \frac{\alpha A}{\alpha + A}$ and $\hat{\epsilon}_{i,t} \equiv \frac{\epsilon_{i,t}}{\epsilon_i}$. 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

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

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

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

and hence $f(-y) = f(y)$. Differentiating $f(-y) = f(y)$ we find $f'(-y) = -f'(y)$, and setting $y = 0$ in $f'(-y) = -f'(y)$ we find $f'(0) = 0$. To show that $f(y)$ is strictly convex, we show that $f''(y) > 0$. Since

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

differentiating once we find

$$f'(y) = -\frac{E\left[\hat{\epsilon}_{i,t} \exp\left(-\hat{\alpha} y \hat{\epsilon}_{i,t}\right)\right]}{E\left[\exp\left(-\hat{\alpha} y \hat{\epsilon}_{i,t}\right)\right]}, \tag{A.1}$$

and differentiating twice we find

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

The numerator in (A.2) 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

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

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

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

$$|f'(y) - 1| = \frac{E\left[(1+\hat{\epsilon}_{i,t})\exp(-\partial y \hat{\epsilon}_{i,t})\right]}{E\left[\exp(-\partial y \hat{\epsilon}_{i,t})\right]}$$
$$= \frac{E\left[(1+\hat{\epsilon}_{i,t})\exp(-\partial y \hat{\epsilon}_{i,t})\mathbf{1}_{\{\hat{\epsilon}_{i,t}\in[-1,-1+\eta]\}}\right]}{E\left[\exp(-\partial y \hat{\epsilon}_{i,t})\right]} + \frac{E\left[(1+\hat{\epsilon}_{i,t})\exp(-\partial y \hat{\epsilon}_{i,t})\mathbf{1}_{\{\hat{\epsilon}_{i,t}\in(-1+\eta,1]\}}\right]}{E\left[\exp(-\partial y \hat{\epsilon}_{i,t})\right]}. \tag{A.3}$$

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.3) is smaller than η. The second term can also be made smaller than η for large y. Indeed, multiplying numerator and denominator by $\exp(-\partial y(1-\eta))$, we can write this term as

$$\frac{E\left[(1+\hat{\epsilon}_{i,t})\exp(-\partial y(\hat{\epsilon}_{i,t}+1-\eta))\mathbf{1}_{\{\hat{\epsilon}_{i,t}\in(-1+\eta,1]\}}\right]}{E\left[\exp(-\partial y(\hat{\epsilon}_{i,t}+1-\eta))\right]}. \tag{A.4}$$

Since $\hat{\epsilon}_{i,t}$ in the numerator of (A.4) exceeds $-1+\eta$, the numerator remains bounded when y goes to infinity. The denominator of (A.4) converges to infinity, however, because $\epsilon_{i,t}$ takes values in $[-1, -1+\eta)$ with positive probability. ∎

Proof of Proposition 2: The results on the value function and optimal consumption follow from Proposition 8 by noting that arbitrageur wealth is deterministic. Optimal positions are derived by maximizing (A.26) with respect to $\{x_{i,t}\}_{i\in A_t}$ subject to the financial constraint (21). Since $\bar{\eta}_i = 0$

and arbitrageur wealth and spreads are deterministic, we can write (A.26) as

$$\max_{\{x_{i,t}\}_{i\in A_t}} \left\{ \beta(B+1)\log\left((1+r)W_t + 2\sum_{i\in A_t} x_{i,t}\Phi_{i,t} \right) + \beta B\log(B) - \beta(B+1)\log(B+1) + \beta G_{t+1} \right\} \tag{A.5}$$

and (21) as (28). When $\Phi_{i,t} = 0$ for all $i \in A_t$, any combination of positions in the active opportunities yields the same value for (A.5). When instead $\Phi_{j,t} > 0$ for some $j \in A_t$, (28) binds for the optimal positions because otherwise the arbitrageur could raise (A.5) by raising $x_{j,t}$. Moreover, $x_{i,t} \geq 0$ for all $i \in A_t$: if $x_{j,t} < 0$ for some $j \in A_t$, then setting $x_{j,t}$ to zero would relax (28) while also not lowering (A.5). Since $x_{i,t} \geq 0$ for all $i \in A_t$, (28) becomes (32). The maximization in (A.5) subject to (32) and $x_{i,t} \geq 0$ for all $i \in A_t$ implies that $x_{i,t} > 0$ only if $i \in A_t$ maximizes return per unit of collateral. ∎

Proof of Proposition 3: Proposition 2 implies that if arbitrageurs trade opportunity i then $\frac{\Phi_{i,t}}{\bar{\epsilon}_i}$ is equal to a value Π_t that is independent of i, 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 (10) imply that $y_{i,t} < 0$. Substituting into (25) and using the convexity of $f(y)$, we find $f'(u_i\epsilon_i) > \frac{\Phi_{i,t}}{\bar{\epsilon}_i} = \Pi_t$. In the latter case, $x_{i,t} = 0$ and (10) imply that $y_{i,t} = 0$. Substituting into (25), we find $f'(u_i\epsilon_i) = \frac{\Phi_{i,t}}{\bar{\epsilon}_i} \leq \Pi_t$. Since $f'(y) < 1$ for all y, $\Pi_t < 1$. ∎

Proof of Proposition 4: Proposition 3 shows that for all $i \in A_t$, either $x_{i,t} > 0$ and $\frac{\Phi_{i,t}}{\bar{\epsilon}_i} = \Pi_t$, or $x_{i,t} = 0$. We can hence write (31) and (32) as

$$W_{t+1} = \beta\left[(1+r)W_t + 2\Pi_t \sum_{i\in A_t} x_{i,t}\epsilon_i \right], \tag{A.6}$$

$$W_t \geq 2(1-\Pi_t)\sum_{i\in A_t} \frac{x_{i,t}\bar{\epsilon}_i}{1+r}, \tag{A.7}$$

respectively. If $\Pi_t = 0$, then (A.6) becomes (33). If instead $\Pi_t > 0$, then (A.7) holds as an equality, and substituting it into (A.6) we find again (33).

If $\Pi_t = 0$, then arbitrageurs earn the riskless rate r and Proposition 3 implies that all active opportunities are traded. To determine a lower bound on W_t, we use market clearing and the financial constraint. Eq. (25) implies that $f'[(y_{i,t}+u_i)\epsilon_i] = \frac{\Phi_{i,t}}{\bar{\epsilon}_i} = 0$ for all $i \in A_t$. 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 (10), we find $x_{i,t}=\mu_i u_i$. Substituting $x_{i,t}=\mu_i u_i$ into (A.7) and using $\Pi_t=0$, we find $W_t \geq W_{c,t}$.

If $\Pi_t > 0$, then arbitrageurs earn the riskless return $\frac{1+r}{1-\Pi_t} - 1 > r$ and Proposition 2 implies that the financial constraint binds. To determine how Π_t relates to W_t, we use market clearing and the financial constraint. Eq. (25) and Proposition 3 imply that $f'[(y_{i,t}+u_i)\epsilon_i] = \frac{\Phi_{i,t}}{\bar{\epsilon}_i} = \Pi_t$ for all $i \in T_t$. Inverting this equation yields

$$\begin{aligned} &(y_{i,t} + u_i)\epsilon_i = (f')^{-1}(\Pi_t) \\ &\Rightarrow x_{i,t} = \mu_i u_i - \mu_i \frac{(f')^{-1}(\Pi_t)}{\bar{\epsilon}_i}, \end{aligned} \tag{A.8}$$

where the second step follows from (10). Substituting (A.8) into (A.7), and recalling that $x_{i,t}=0$ for all $i \in A_t/T_t$, we find (34). Moreover, (A.7) implies that $W_t < W_{c,t}$ because $\Pi_t \in (0,1)$, $0 < x_{i,t} < \mu_i u_{i,t}$ for all $i \in T_t$ (from (A.8)), and $x_{i,t}=0$ for all $i \in A_t/T_t$.

The left-hand side of (34) decreases in Π_t because $f''(y) > 0$ implies that $f'(y)$ is increasing. Moreover, it is equal to zero for $\Pi_t = \max_{i \in A_t} f'(u_i\epsilon_i)$ and to $W_{c,t}$ for $\Pi_t = 0$. Therefore, (34) 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 (34) implicitly with respect to W_t. We find

$$\frac{d\Pi_t}{dW_t} = -\frac{2}{1+r} \sum_{i \in T_t} \mu_i \left[u_i \bar{\epsilon}_i - (f')^{-1}(\Pi_t) + \frac{1-\Pi_t}{f''[(f')^{-1}(\Pi_t)]} \right]. \tag{A.9}$$

The derivative $\frac{d\Pi_t}{dW_t}$ is continuous, except at $W_t = W_{c,t}$ and at the points where the set 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 T_t changes, this is because the denominator for the right derivative minus that for the left derivative is

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

where DT_t denotes the additional opportunities that become traded to the right of that point.

Therefore, Π_t is convex if the function

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

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

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

Differentiating (A.2) we find

$$f'''(y) = \hat{\alpha}^2 \left[2E\left[\hat{\epsilon}_{i,t}\exp\left(-\hat{\alpha}y\hat{\epsilon}_{i,t}\right)\right] \frac{E\left[\hat{\epsilon}_{i,t}^2\exp\left(-\hat{\alpha}y\hat{\epsilon}_{i,t}\right)\right]E\left[\exp\left(-\hat{\alpha}y\hat{\epsilon}_{i,t}\right)\right] - \left\{E\left[\hat{\epsilon}_{i,t}\exp\left(-\hat{\alpha}y\hat{\epsilon}_{i,t}\right)\right]\right\}^2}{\left\{E\left[\exp\left(-\hat{\alpha}y\hat{\epsilon}_{i,t}\right)\right]\right\}^3}\right.$$
$$\left. + \frac{E\left[\hat{\epsilon}_{i,t}^2\exp\left(-\hat{\alpha}y\hat{\epsilon}_{i,t}\right)\right]E\left[\hat{\epsilon}_{i,t}\exp\left(-\hat{\alpha}y\hat{\epsilon}_{i,t}\right)\right] - E\left[\hat{\epsilon}_{i,t}^3\exp\left(-\hat{\alpha}y\hat{\epsilon}_{i,t}\right)\right]E\left[\exp\left(-\hat{\alpha}y\hat{\epsilon}_{i,t}\right)\right]}{\left\{E\left[\exp\left(-\hat{\alpha}y\hat{\epsilon}_{i,t}\right)\right]\right\}^2}\right]. \tag{A.10}$$

Using (A.1), (A.2) and (A.10), we find

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

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,

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

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

and the function $G_1(y)$ becomes

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

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 $i \in A_t$ are given by Propositions 3 and 4, (ii) price discounts $\phi_{i,t}$ for assets $i \in A_t$ are given by solving (13) backwards with the terminal condition $\phi_{i,h_i} = 0$:

$$\phi_{i,t} = \sum_{s=0}^{h_i - t - 1} \frac{\Phi_{i,t+s}}{(1+r)^{s+1}}, \tag{A.11}$$

(iii) expected excess returns and price discounts for assets $-i$, $i \in A_t$, are opposites to those for assets i, and (iv) expected excess returns and price discounts for assets that are not part of active opportunities are zero. The first-order conditions of investors and arbitrageurs are then as in Section 3.1. The analysis in that section and in Section 3.2 ensures that the markets for all assets clear and that the quantities $(W_t, \phi_{i,t}, y_{i,t}, x_{i,t})$ 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 (33) implies that W_t increases to $W = \infty$.
- If $\beta(1+r) < 1 - \underline{\Pi}$, then (33) implies that W_t decreases to $W = 0$ because $\underline{\Pi}$ is the maximum value of Π_t.
- If $1 - \overline{\Pi} < \beta(1+r) < 1$, then (33) implies that W_t remains constant when Π_t is equal to the steady-state value Π given by (36). The steady-state value W of W_t is given by (35) because of (34). When $W_t < W$, (33) implies that $W_{t+1} > W_t$ because $\Pi_t > \Pi$. Conversely, when $W_t > W$, (33) 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 (33) 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 $W_t < W_c$, (34) implies that

$$F(W_t) = \frac{2}{1+r}\sum_{i \in T_t} \mu_i \left[u_i \epsilon_i - (f')^{-1}(\Pi_t)\right].$$

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 Proposition 7: The investor's Bellman equation is

$$V_{i,t}(w_{i,t}) = \max_{c_{i,t+1}, y_{i,t}} \mathrm{E}_t\{-\gamma \exp(-\alpha c_{i,t+1}) + \gamma V_{i,t+1}(w_{i,t+1})\}. \tag{A.12}$$

Substituting (14) and (15) into (A.12), we find

$$-\exp(-Aw_{i,t} - F_{i,t}) = \max_{c_{i,t+1}, y_{i,t}} \mathrm{E}_t\{-\gamma \exp(-\alpha c_{i,t+1}) - \gamma \exp\left(-A\left[(1+r)w_{i,t} + \Gamma_{i,t+1}(y_{i,t}) + (y_{i,t} + u_{i,t})\epsilon_{i,t+1} - c_{i,t+1}\right] - F_{i,t+1}\right)\}. \tag{A.13}$$

where

$$\Gamma_{i,t+1}(y_{i,t}) \equiv y_{i,t}\Phi_{i,t} + (y_{i,t} + u_{i,t})\eta_{i,t+1} + y_{i,t}\left[\mathrm{E}_t(\phi_{i,t+1}) - \phi_{i,t+1}\right].$$

The first-order condition with respect to consumption is

$$\alpha \exp(-\alpha c_{i,t+1}) = A \exp\left(-A\left[(1+r)w_{i,t} + \Gamma_{i,t+1}(y_{i,t}) + (y_{i,t}+u_{i,t})\epsilon_{i,t+1} - c_{i,t+1}\right] - F_{i,t+1}\right) \tag{A.14}$$

$$\Rightarrow c_{i,t+1} = \frac{A\left[(1+r)w_{i,t} + \Gamma_{i,t+1}(y_{i,t}) + (y_{i,t}+u_{i,t})\epsilon_{i,t+1}\right] + F_{i,t+1} + \log\left(\frac{\alpha}{A}\right)}{\alpha + A}. \tag{A.15}$$

Hence, we can write the right-hand side of (A.13) as

$$\max_{y_{i,t}} \mathrm{E}_t\left\{-\frac{\gamma(\alpha+A)}{\alpha}\exp\left(-A\left[(1+r)w_{i,t} + \Gamma_{i,t+1}(y_{i,t}) + (y_{i,t}+u_{i,t})\epsilon_{i,t+1} - c_{i,t+1}\right] - F_{i,t+1}\right)\right\}$$

$$= \max_{y_{i,t}} \mathrm{E}_t\left\{-\frac{\gamma(\alpha+A)}{\alpha}\exp\left(\frac{-\alpha\left\{A\left[(1+r)w_{i,t} + \Gamma_{i,t+1}(y_{i,t}) + (y_{i,t}+u_{i,t})\epsilon_{i,t+1}\right] + F_{i,t+1}\right\} + A\log\left(\frac{\alpha}{A}\right)}{\alpha+A}\right)\right\} \tag{A.16}$$

$$= \max_{y_{i,t}} \mathrm{E}_t\left\{-\frac{\gamma(\alpha+A)}{\alpha}\exp\left(\frac{-\alpha\left\{A\left[(1+r)w_{i,t} + \Gamma_{i,t+1}(y_{i,t}) - f[(y_{i,t}+u_{i,t})\bar{\epsilon_i}]\right] + F_{i,t+1}\right\} + A\log\left(\frac{\alpha}{A}\right)}{\alpha+A}\right)\right\}, \tag{A.17}$$

where the first step follows from (A.14), the second from (A.15), and the third from the independence of $\epsilon_{i,t+1}$ and $\Gamma_{i,t+1}(y_{i,t})$ and from (26) by setting $y \equiv (y_{i,t}+u_{i,t})\bar{\epsilon_i}$. Using the definition of $\Gamma_{i,t+1}(y_{i,t})$, we find that the first-order condition with respect to $y_{i,t}$ is

$$\mathrm{E}_t\left\{\exp\left(\frac{-\alpha\left\{A\left[(1+r)w_{i,t} + \Gamma_{i,t+1}(y_{i,t}) - f[(y_{i,t}+u_{i,t})\bar{\epsilon_i}]\right] + F_{i,t+1}\right\} + A\log\left(\frac{\alpha}{A}\right)}{\alpha+A}\right)\right.$$
$$\left.\times\left\{\Phi_{i,t} - \bar{\epsilon_i} f'[(y_{i,t}+u_{i,t})\bar{\epsilon_i}] + \eta_{i,t+1} + \mathrm{E}_t(\phi_{i,t+1}) - \phi_{i,t+1}\right\}\right\} = 0. \tag{A.18}$$

Rearranging terms, we find (37).

Substituting (A.17) into (A.13), we find that the Bellman equation holds for all values of the single state variable $w_{i,t}$ if

$$A = \frac{\alpha A(1+r)}{\alpha + A}, \tag{A.19}$$

$$F_{i,t} = -\log \mathrm{E}_t\left\{\frac{\gamma(\alpha+A)}{\alpha}\exp\left(\frac{-\alpha\left\{A\left[\Gamma_{i,t+1}(y_{i,t}) - f[(y_{i,t}+u_{i,t})\bar{\epsilon_i}]\right] + F_{i,t+1}\right\} + A\log\left(\frac{\alpha}{A}\right)}{\alpha+A}\right)\right\}, \tag{A.20}$$

where $y_{i,t}$ denotes the optimal position. Eq. (A.19) implies that $A = r\alpha$. Substituting into (A.20), we find

$$F_{i,t} = -\log \mathrm{E}_t \left\{ \gamma(1+r) \exp\left(\frac{-r\alpha[\Gamma_{i,t+1}(y_{i,t}) - f[(y_{i,t}+u_{i,t})\epsilon_j]] - F_{i,t+1} - r\log(r)}{1+r} \right) \right\}. \quad \text{(A.21)}$$

Eq. (A.21) determines $F_{i,t}$ in terms of $y_{i,t}$ and $F_{i,t+1}$. ∎

Proof of Proposition 8: The arbitrageur's Bellman equation is

$$V_t(W_t) = \max_{c_{t+1}, \{x_{i,t}\}_{i \in A_t}} \mathrm{E}_t \{ \beta \log(c_{i,t+1}) + \beta V_{t+1}(W_{t+1}) \}. \quad \text{(A.22)}$$

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

$$B\log(W_t) + G_t = \max_{c_{t+1}, \{x_{i,t}\}_{i \in A_t}} \mathrm{E}_t \left\{ \beta \log(c_{i,t+1}) + \beta B \log\left((1+r)W_t + 2\sum_{i \in A_t} \Gamma_{i,t+1}(x_{i,t}) - c_{t+1} \right) + \beta G_{t+1} \right\}, \quad \text{(A.23)}$$

where

$$\Gamma_{i,t+1}(x_{i,t}) \equiv x_{i,t}\left[\Phi_{i,t} + \eta_{i,t+1} + \mathrm{E}_t(\phi_{i,t+1}) - \phi_{i,t+1}\right].$$

The first-order condition with respect to consumption is

$$\frac{1}{c_{i,t+1}} - \frac{B}{(1+r)W_t + 2\sum_{i \in A_t} \Gamma_{i,t+1}(x_{i,t}) - c_{t+1}} = 0$$

$$\Rightarrow c_{i,t+1} = \frac{(1+r)W_t + 2\sum_{i \in A_t} \Gamma_{i,t+1}(x_{i,t}) - c_{t+1}}{B} = \frac{W_{t+1}}{B} \quad \text{(A.24)}$$

$$\Rightarrow c_{i,t+1} = \frac{(1+r)W_t + 2\sum_{i \in A_t} \Gamma_{i,t+1}(x_{i,t})}{B+1}, \quad \text{(A.25)}$$

where the second equality in (A.24) follows from (20). Using (A.25), we can write the right-hand side of (A.23) as

$$\max_{\{x_{i,t}\}_{i \in A_t}} \mathrm{E}_t \left\{ \beta(B+1)\log\left((1+r)W_t + 2\sum_{i \in A_t} \Gamma_{i,t+1}(x_{i,t}) \right) + \beta B\log(B) - \beta(B+1)\log(B+1) + \beta G_{t+1} \right\}. \quad \text{(A.26)}$$

The maximization is subject to the financial constraint (21).

The optimal values of $\{x_{i,t}\}_{i\in A_t}$ are linear in W_t, as can be seen by setting $x_{i,t} \equiv W_t\omega_{i,t}$ and noting that the maximization objective and constraint can be written solely in terms of $\{\omega_{i,t}\}_{i\in A_t}$. Using this observation and substituting (A.26) into (A.23), we find that the Bellman equation holds for all values of the single state variable W_t if

$$B = \beta(B+1), \tag{A.27}$$

$$G_t = \mathrm{E}_t\left\{\beta(B+1)\log\left((1+r)+2\sum_{i\in A_t}\Gamma_{i,t+1}(\omega_{i,t})\right) + \beta B\log(B) - \beta(B+1)\log(B+1) + \beta G_{t+1}\right\}, \tag{A.28}$$

where $\{\omega_{i,t}\}_{i\in A_t}$ denote the optimal positions as fractions of wealth. Eq. (A.19) implies that $B = \frac{\beta}{1-\beta}$. Substituting into (A.24), we find (29). Eq. (A.28) determines G_t in terms of $\{\omega_{i,t}\}_{i\in A_t}$ and G_{t+1}.

To characterize the optimal positions $\{x_{i,t}\}_{i\in A_t}$, which are assumed to be non-negative, we first compute the derivative of (A.26) with respect to $x_{i,t}$. This derivative is

$$2\beta(B+1)\mathrm{E}_t\left\{\frac{\Phi_{i,t}+\eta_{i,t+1}+\mathrm{E}_t(\phi_{i,t+1})-\phi_{i,t+1}}{(1+r)W_t+2\sum_{j\in A_t}\Gamma_{j,t+1}(x_{j,t})}\right\} = 2\beta(B+1)\mathrm{E}_t[M_t]\,\hat{\Phi}_{i,t}, \tag{A.29}$$

where

$$\hat{\Phi}_{i,t} \equiv \Phi_{i,t} + \mathrm{E}_t\left[\frac{M_t}{\mathrm{E}_t[M_t]}\left(\eta_{i,t+1}+\mathrm{E}_t(\phi_{i,t+1})-\phi_{i,t+1}\right)\right].$$

We next show that under the optimal positions either (i) $\hat{\Phi}_{i,t}=0$ for all $i\in A_t$ and (21) is slack, or (ii) $\hat{\Phi}_{j,t}>0$ for some $j\in A_t$ and (21) binds. If $\hat{\Phi}_{j,t}>0$ for some $j\in A_t$, then (21) binds because otherwise the arbitrageur could raise (A.26) by raising $x_{j,t}$. If instead $\hat{\Phi}_{i,t}\le 0$ for all $i\in A_t$, then these inequalities must hold as equalities, and hence (21) is slack (because $\hat{\Phi}_{i,t}=0$ for all $i\in A_t$ is the first-order condition from the maximization in (A.26) when the constraint (21) is not imposed). To show that the inequalities must hold as equalities, we proceed by contradiction, assuming that $\hat{\Phi}_{j,t}<0$ for some $j\in A_t$ and distinguishing three cases. If $x_{j,t}>0$, then lowering $x_{j,t}$ would raise (A.26) while also relaxing (21). If $x_{j,t}=0$ and $x_{j',t}>0$ for some $j'\in A_t$, $j'\neq j$, then lowering $x_{j',t}$ would relax (21) while also not lowering (A.26) because $\hat{\Phi}_{j',t}\le 0$. With (21) relaxed, the

arbitrageur could raise (A.26) by lowering $x_{j,t}$ (to a negative value). If, finally, $x_{i,t} = 0$ for all $i \in A_t$, then (21) is slack, and hence lowering $x_{j,t}$ would raise (A.26).

When optimal positions $\{x_{i,t}\}_{i \in A_t}$ are non-negative, the maximization in (A.26) can be carried out subject to (39) and $x_{i,t} \geq 0$, rather than subject to (21). This maximization implies that $x_{i,t} > 0$ only if $i \in A_t$ maximizes risk-adjusted return per unit of collateral. (This is obvious in the case where $\hat{\Phi}_{i,t} = 0$ for all $i \in A_t$. In the case where $\hat{\Phi}_{j,t} > 0$ for some $j \in A_t$, it can be shown by lowering $x_{i,t} > 0$ for an $i \in A_t$ that does not maximize risk-adjusted return per unit of collateral and raising $x_{i,t} \geq 0$ for an i that does, in a way that keeps (39) binding.)

When $\min_{i \in A_t} \Phi_{i,t}$ exceeds a positive bound and $\max_{i \in A_t} \{\bar{\eta}_i, \phi_{i,t+1} - \mathrm{E}_t(\phi_{i,t+1})\}$ is small, $\hat{\Phi}_{i,t} > 0$ for all $i \in A_t$, and hence (21) binds. Moreover, short positions are not optimal: if $x_{j,t} < 0$ for some $j \in A_t$, then setting $x_{j,t}$ to zero would raise (A.26) and relax (21). ∎

Proof of Corollary 1: When optimal positions $\{x_{i,t}\}_{i \in A_t}$ are non-negative, the maximization in (A.26) can be carried out subject to (39) and $x_{i,t} \geq 0$. Consider this maximization when (39) is not imposed. Because of logarithmic utility, the solution satisfies

$$(1+r)W_t + 2\sum_{i \in A_t} \Gamma_{i,t+1}(x_{i,t}) > 0$$

$$\Leftrightarrow (1+r)W_t + 2\sum_{i \in A_t} x_{i,t}\left[\Phi_{i,t} + \eta_{i,t+1} + \mathrm{E}_t(\phi_{i,t+1}) - \phi_{i,t+1}\right] > 0 \qquad \text{(A.30)}$$

for all realizations of uncertainty in period $t+1$, including when $\eta_{i,t+1} = -\bar{\eta}_i$ for all $i \in A_t$. Under the latter realization, and because $\{x_{i,t}\}_{i \in A_t}$ are non-negative, the spreads $\phi_{i,t+1}$ for all $i \in A_t$ reach their maximum values $\bar{\phi}_{i,t+1}$. Since a long position in asset $i \in A_t$ suffers its maximum loss when $\eta_{i,t+1} = -\bar{\eta}_i$ and $\phi_{i,t+1} = \bar{\phi}_{i,t+1}$, (A.30) holds for all realizations of uncertainty in period $t+1$ if

$$(1+r)W_t + 2\sum_{i \in A_t} x_{i,t}\left[\Phi_{i,t} - \bar{\eta}_i + \mathrm{E}_t(\phi_{i,t+1}) - \bar{\phi}_{i,t+1}\right] > 0. \qquad \text{(A.31)}$$

Equation (A.31) implies that when $\bar{\epsilon}_i = 0$ for all $i \in A_t$, (39) holds as a strict inequality and hence is slack. ∎

We next state and prove Proposition A.1 which characterizes the equilibrium for small arbitrage risk. We denote by Π' the derivative of Π_t defined in (34) with respect to W_t at the steady-state value $W_t = W$.

Proposition A.1. *Suppose that $\bar{\eta}$ is small. The dynamics of arbitrageur wealth take the form (45),*

with

$$\rho = 1 + \frac{W\Pi'}{1-\Pi}, \tag{A.32}$$

$$v = 2\beta \sum_{(n,m)\in T} \left(x^0_n \Phi^1_{n,m} + x^1_{n,m} \Phi^0_n \right), \tag{A.33}$$

and σ_n *given by (46). Expected excess returns, and positions of arbitrageurs and outside investors take the form (42)-(44), with* $(\Phi^0_{n,t}, x^0_{n,t}, y^0_{n,t})$ *as in Section 3, and*

$$\Phi^1_{n,m,t} = \begin{cases} \Pi_t \left(\lambda_n - \frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{n'\in N} M_{n'} \sigma_{n',t} \lambda_{n'} \right) + \Pi^1_t \epsilon_n, & \text{if } (n,m) \in T \\ 0, & \text{otherwise} \end{cases} \tag{A.34}$$

$$x^1_{n,m,t} = -\mu_n y^1_{n,m,t}, \tag{A.35}$$

$$v^1_{n,m,t} = \frac{\epsilon^2_n f''\left[\left(y^0_{n,t} + u_n\right)\epsilon_n\right]}{\Phi^1_{n,m,t}}, \tag{A.36}$$

$$\Pi^1_t \equiv \frac{(1-\Pi_t) \sum_{(n,m)\in T} \left[x^0_{n,t} - \frac{\mu_n \Pi_t}{\epsilon_n f''\left[\left(y^0_{n,t} + u_n\right)\epsilon_n\right]} \right] \left(\lambda_n - \frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{n'\in N} M_{n'} \sigma_{n',t} \lambda_{n'} \right)}{\sum_{n\in N} \left[x^0_{n,t} \epsilon_n + (1-\Pi_t) \frac{\mu_n \Pi_t}{f''\left[\left(y^0_{n,t} + u_n\right)\epsilon_n\right]} \right]}, \tag{A.37}$$

$$\sigma_{n,t} \equiv \frac{2\beta x^0_{n,t}}{1 + 2\beta \sum_{(n',m)\in T} x^0_{n',t} \frac{d\phi^0_{n',m-1,t+1}}{dW_{t+1}}}. \tag{A.38}$$

The derivative $\frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}}$ *in (A.34), (A.37), and (A.38) is evaluated for* W_{t+1} *implied by the dynamics (33) under riskless arbitrage.*

Proof: The dynamics (45) of arbitrageur wealth can be derived from (47). Since the spreads $\{\phi_{n,m-1,t+1}\}_{(n,m)\in T}$ depend on W_{t+1}, (47) determines W_{t+1} implicitly as a function F of W_t, of the shocks $\{\eta_{n,m-1,t+1}\}_{(n,m)\in T}$, and of η. For a fixed W_t and small η, the Taylor expansion of F is

$$W_{t+1} = F + \sum_{(n,m)\in T} \frac{\partial F}{\partial \eta_{n,m-1,t+1}} \eta_{n,m-1,t+1} + \frac{\partial F}{\partial \eta} \eta + o(\eta), \tag{A.39}$$

where F and its derivatives are evaluated at $(W_t, 0, 0)$. Using (A.39), we find

$$\begin{aligned}
&\mathrm{E}_t(\phi_{n,m-1,t+1}) - \phi_{n,m-1,t+1} = \mathrm{E}_t(\phi^0_{n,m-1,t+1}) - \phi^0_{n,m-1,t+1} + o(\eta) \\
&= \mathrm{E}_t\Big[\phi^0_{n,m-1,t+1}(F) + \frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{(n',m')\in T} \frac{\partial F}{\partial \eta_{n',m'-1,t+1}} \eta_{n',m'-1,t+1} + \frac{\partial F}{\partial \bar{\eta}} \bar{\eta}\Big] \\
&\quad - \Big[\phi^0_{n,m-1,t+1}(F) - \frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{(n',m')\in T} \frac{\partial F}{\partial \eta_{n',m'-1,t+1}} \eta_{n',m'-1,t+1} + \frac{\partial F}{\partial \bar{\eta}} \bar{\eta}\Big] + o(\eta) \\
&= -\frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{(n',m')\in T} \frac{\partial F}{\partial \eta_{n',m'-1,t+1}} \eta_{n',m'-1,t+1} + o(\eta),
\end{aligned} \tag{A.40}$$

where the derivative $\frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}}$ is evaluated for $W_{t+1} = F(W_t, 0, 0)$. Substituting (41)-(43), (A.39), and (A.40) into (47), we find

$$\begin{aligned}
&F + \sum_{(n,m)\in T} \frac{\partial F}{\partial \eta_{n,m-1,t+1}} \eta_{n,m-1,t+1} + \frac{\partial F}{\partial \bar{\eta}} \bar{\eta} = \beta\Big[(1+r)W_t + 2\sum_{(n,m)\in T} (x^0_{n,t} + x^1_{n,m,t}\bar{\eta}) \\
&\times \Big[\Phi^0_{n,t} + \Phi^1_{n,m,t}\Big(\eta + \eta_{n,m-1,t+1} - \frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{(n',m')\in T} \frac{\partial F}{\partial \eta_{n',m'-1,t+1}} \eta_{n',m'-1,t+1}\Big) + o(\eta),
\end{aligned} \tag{A.41}$$

Identifying highest-order terms in $\eta_{n,m-1,t+1}$ in (A.41) we find

$$\frac{\partial F}{\partial \eta_{n,m-1,t+1}} = \frac{2\beta x^0_{n,t}}{1 + 2\beta \sum_{(n',m')\in T} x^0_{n',t} \frac{d\phi^0_{n',m'-1,t+1}}{dW_{t+1}}} = \sigma_{n,t}, \tag{A.42}$$

and identifying highest-order terms in $\bar{\eta}$ we find

$$\frac{\partial F}{\partial \bar{\eta}} = 2\beta \sum_{(n,m)\in T} (x^0_{n,t}\Phi^1_{n,m,t} + x^1_{n,m,t}\Phi^0_{n,t}). \tag{A.43}$$

When W_t lies in the support of the stationary distribution, its maximum distance from the steady-state value W is of order $\bar{\eta}$ since this is the maximum size of the shocks. Hence, we can

write (A.39) as

$$W_{t+1} = W + \frac{\partial F}{\partial W_t}(W_t - W) + \sum_{(n,m)\in T} \frac{\partial F}{\partial \eta_{n,m-1,t+1}} \eta_{n,m-1,t+1} + \frac{\partial F}{\partial \bar{\eta}} \bar{\eta} + o(\eta), \tag{A.44}$$

where F and its derivatives are now evaluated at $(W, 0, 0)$. Since $F(W_t, 0, 0) = \beta \frac{1+r}{1-\Pi_t} W_t$

$$\frac{\partial F}{\partial W_t} = \beta \frac{1+r}{1-\Pi} + \beta \frac{(1+r)\Pi'}{(1-\Pi)^2} W = 1 + \frac{W\Pi'}{1-\Pi},$$

where the second step follows from (36). Since, in addition, (A.42) and (A.43) imply

$$\frac{\partial F}{\partial \eta_{n,m-1,t+1}} = \sigma_n,$$

$$\frac{\partial F}{\partial \bar{\eta}} = v,$$

respectively, when partial derivatives are evaluated at $(W, 0, 0)$, (A.44) becomes (45). The coefficient ρ is smaller than one because Π_t decreases in W_t. It is larger than zero because the function $\frac{W_t}{1-\Pi_t}$ increases in W_t, as shown in the proof of Proposition 6.

We next derive the first-order terms in expected excess returns and positions. Substituting (42) and (44) into (37), and noting that the third term in the left-hand side is of second order in η, we find

$$\Phi^0_{n,t} + \Phi^1_{n,m,t} \underline{\eta} - \underline{\epsilon}_i f'[(y^0_{n,t} + u_{i,t})\underline{\epsilon}_i] - \underline{\epsilon}^2_i f'[(y^0_{n,t} + u_{i,t})\underline{\epsilon}_i] y^1_{n,m,t} \underline{\eta} + o(\underline{\eta}) = 0. \tag{A.45}$$

Identifying terms in $\bar{\eta}$, we find (A.36). Substituting (43) and (44) into (10) and identifying terms in η, we find (A.35). Substituting (42), (43), (A.40), and (A.42) into (40), and noting that the second term in the numerator is of second order in $\bar{\eta}$, we find that

$$\frac{\Phi^0_{n,t} + \Phi^1_{n,m,t} \eta}{\bar{\epsilon}_n + \bar{\eta}_n - \frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{n'\in N} M_{n'} \sigma_{n',t} \bar{\eta}_{n'}}$$

is identical for all $(n,m) \in T$ to a first order in $\bar{\eta}$. Since $\frac{\Phi^0_{n,t}}{\bar{\epsilon}_n} = \Pi_t$ for all $(n,m) \in T$,

$$\frac{\Phi^0_{n,t} + \Phi^1_{n,m,t}\bar{\eta}}{\bar{\epsilon}_n + \bar{\eta}_n - \frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{n' \in N} M_{n'} \sigma_{n',t} \bar{\eta}_{n'}} - \Pi_t$$
$$= \frac{\Phi^1_{n,m,t}\bar{\eta} - \Pi_t \left(\bar{\eta}_n - \frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{n' \in N} M_{n'} \sigma_{n',t} \bar{\eta}_{n'}\right)}{\bar{\epsilon}_n + \bar{\eta}_n - \frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{n' \in N} M_{n'} \sigma_{n',t} \bar{\eta}_{n'}} \equiv \Pi^1_t \bar{\eta} \tag{A.46}$$

is identical for all $(n,m) \in T$ to a first order in $\bar{\eta}$. Multiplying by the denominator in (A.46), we find

$$\Phi^1_{n,m,t}\bar{\eta} = \Pi_t \left(\bar{\eta}_n - \frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{n' \in N} M_{n'} \sigma_{n',t} \bar{\eta}_{n'}\right) + \Pi_t \epsilon_n \bar{\eta} + o(\bar{\eta}) \tag{A.47}$$

for all $(n,m) \in T$. Since $\bar{\eta}_n = \lambda_n \bar{\eta}$, (A.47) implies (A.34) for $(n,m) \in T$. For $(n,m) \notin T$, $x_{n,m,t} = 0$ and hence $x^1_{n,m,t} = 0$. Eq. (A.35) then implies $y^1_{n,m,t} = 0$, and (A.36) implies $\Phi^1_{n,m,t} = 0$.

To compute Π^1_t, we use the financial constraint (39), which binds. Substituting (42), (43), (A.40), and (A.42) into (39), we find

$$W_t = 2 \sum_{(n,m) \in T} \frac{\left(x^0_{n,t} + x^1_{n,m,t}\bar{\eta}\right)\left(\bar{\epsilon}_n + \bar{\eta}_n - \frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{n' \in N} M_{n'} \sigma_{n',t} \bar{\eta}_{n'}\right) - \left(\Phi^0_{n,t} + \Phi^1_{n,m,t}\bar{\eta}\right)}{1 + r} + o(\eta). \tag{A.48}$$

Identifying first-order terms in $\bar{\eta}$, we find

$$\sum_{(n,m) \in T} \left[x^0_{n,t}\left(\bar{\eta}_n - \frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{n' \in N} M_{n'} \sigma_{n',t} \bar{\eta}_{n'}\right) - \Phi^1_{n,m,t}\bar{\eta} + x^1_{n,m,t}\left(\bar{\epsilon}_n - \Phi^0_{n,t}\right)\right] = 0$$
$$\Leftrightarrow (1 - \Pi_t) \sum_{(n,m) \in T} \left[x^0_{n,t}\left(\lambda_n - \frac{d\phi^0_{n,m-1,t+1}}{dW_{t+1}} \sum_{n' \in N} M_{n'} \sigma_{n',t} \lambda_{n'}\right) + x^1_{n,m,t}\bar{\epsilon}_n\right] - \Pi^1_t \sum_{(n,m) \in T} x^0_{n,t}\bar{\epsilon}_n = 0, \tag{A.49}$$

where the second step follows from (A.47) and $\frac{\Phi^0_{n,t}}{\bar{\epsilon}_n} = \Pi_t$ for all $(n, m) \in T$. Noting from (A.34)-

(A.36) that

$$x^1_{n,m,t} = -\mu_n y^1_{n,m,t}$$

$$= -\frac{\mu_n \Phi^1_{n,m,t}}{\bar{\epsilon}_n^2 f''\left[\left(y^0_{n,t} + u_n\right)\bar{\epsilon}_n\right]}$$

$$= -\frac{\mu_n \left[\Pi_t \left(\lambda_n - \frac{\partial \phi^0_{n,m,t}}{\partial M_{t+1}} \frac{M_{t+1}}{\phi^0_{n,m,t}}\right) \sum_{n' \in N} M_{n'} \sigma_{n',t} \lambda_{n'} + \Pi^1_t \epsilon_n\right]}{\bar{\epsilon}_n^2 f''\left[\left(y^0_{n,t} + u_n\right)\bar{\epsilon}_n\right]},$$

and substituting into (A.49), we find a linear equation in Π^1_t whose solution is (A.37). ∎

Proof of Proposition 9: For each characteristic, we compare two arbitrage opportunities that differ only in that characteristic, and we assume that arbitrage risk is small. We can perform the comparison by examining how spreads, expected excess returns, and positions corresponding to one arbitrage opportunity (n, m) depend on that characteristic, holding aggregate variables constant. If zeroth-order terms in spreads, expected excess returns, and positions depend on the characteristic, then we use those to determine the characteristic's effect; otherwise, we use first- or second-order terms.

The comparative statics with respect to $\bar{\epsilon}_n$ follow from the zeroth-order terms. Proposition 3 implies that $\Phi^0_{n,t}$ increases in $\bar{\epsilon}_n$: it does so both in the region where the opportunity is not traded because $\bar{\epsilon}_n$ is below a threshold and in the region where the opportunity is traded because $\bar{\epsilon}_n$ is above the threshold. Since $\Phi^0_{n,t}$ increases in $\bar{\epsilon}_n$, so does $\phi^0_{n,m,t}$ because of (48). The arbitrageur position $x_{n,m,t}$ increases in $\bar{\epsilon}_n$ because it is zero in the region where the opportunity is not traded and because (A.8) implies that $x^0_{n,t}$ increases in $\bar{\epsilon}_n$ in the region where the opportunity is traded.

The comparative statics with respect to $\bar{\eta}_n$ follow from the first-order terms, or the second-order terms, or are trivial. Suppose that the opportunity is traded. Eq. (A.34) implies that $\Phi^1_{n,m,t}$ increases in λ_n and hence in $\bar{\eta}_n$. Eq. (48) then implies that $\phi^1_{n,m,t}$ increases in η_n, and (A.35) and (A.36) imply that $x^1_{n,m,t}$ decreases in η_n. Suppose next that the opportunity is not traded. Eq. (A.34) implies that $\Phi^1_{n,m,t} = 0$, so we need to consider the second-order term, which we can derive from the first-order condition of outside investors. Since the opportunity is not traded, the presence of arbitrageurs is immaterial, so spreads and the outside investors' value function are independent of arbitrageur wealth. Using these observations, the independence of the payoff shocks $(\epsilon_{n,m,t}, \eta_{n,m,t})$, and the same arguments as in the proof of Proposition 7, we find the first-order condition

$$\Phi_{n,m,t} - \bar{\epsilon}_n f'(u_n \bar{\epsilon}_n) - \bar{\eta}_n f'(u_n \bar{\eta}_n) = 0. \tag{A.50}$$

Eq. (A.50) implies that $\Phi_{n,m,t}$ increases in $\bar{\eta}_n$. Eq. (48) then implies that $\phi_{n,m,t}$ also increases in $\bar{\eta}_n$. The arbitrageur position $x_{n,m,t}$ is zero, and hence decreases weakly in $\bar{\eta}_n$.

The comparative statics of spreads with respect to m follow from the zeroth-order term. Those of expected excess returns and positions follow from the first-order terms or are trivial. Identifying zeroth-order terms in (48) and noting that at the zeroth order the dynamics of W_t are deterministic, we find

$$\phi^0_{n,m,t} = \sum_{s=0}^{m-1} \frac{\Phi^0_{n,t+s}}{(1+r)^{s+1}} \tag{A.51}$$

$$\Rightarrow \frac{d\phi^0_{n,m,t}}{dW_t} = \sum_{s=0}^{m-1} \frac{1}{(1+r)^{s+1}} \frac{d\Phi^0_{n,t+s}}{dW_t}$$

$$\Rightarrow \frac{d\phi^0_{n,m,t}}{dW_t} = \sum_{s=0}^{m-1} \frac{1}{(1+r)^{s+1}} \frac{d\Phi^0_{n,t+s}}{dW_{t+s}} \frac{dW_{t+s}}{dW_t}. \tag{A.52}$$

Subtracting (A.51) and (A.52) from their counterparts for $m+1$, we find

$$\phi^0_{n,m+1,t} - \phi^0_{n,m,t} = \frac{\Phi^0_{n,t+m}}{(1+r)^{m+1}}, \tag{A.53}$$

$$\frac{d\phi^0_{n,m+1,t}}{dW_t} - \frac{d\phi^0_{n,m,t}}{dW_t} = \frac{1}{(1+r)^{m+1}} \frac{d\Phi^0_{n,t+m}}{dW_{t+m}} \frac{dW_{t+m}}{dW_t}, \tag{A.54}$$

respectively. Since $\Phi^0_{n,t+m} > 0$, (A.53) implies that $\phi^0_{n,m,t}$ increases in m. Since the function $\frac{W_t}{1-\Pi_t}$ is increasing in W_t (as shown in the proof of Proposition 6), Proposition 4 implies that W_{t+s} increases in W_t for $s \geq 1$. Since, in addition, Π_t decreases in W_t (Proposition 4), Proposition 3 implies that $\Phi^0_{n,t}$ for a traded opportunity decreases in W_t. Hence, (A.52) implies that $\phi^0_{n,m,t}$ decreases in W_t. Moreover, (A.54) implies that $\frac{d\phi^0_{n,m,t}}{dW_t}$ decreases in m, becoming more negative for larger m. Eq. (A.34) then implies that $\Phi^1_{n,m,t}$ increases in m, and (A.35) and (A.36) imply that $x^1_{n,m,t}$ decreases in m. Eq. (A.50) implies that $\Phi_{n,m,t}$ is independent of m for a non-traded opportunity, and hence increases weakly in m. The arbitrageur position $x_{n,m,t}$ is zero, and hence decreases weakly in m.

The comparative statics with respect to u_n follow from the zeroth-order terms, or cannot be determined from the zeroth- and first-order terms, or are trivial. Proposition 3 implies that $\Phi^0_{n,t}$ increases in u_n in the region where the opportunity is not traded because u_n is below a threshold. Eq. (A.34) then implies that $\phi^0_{n,m,t}$ increases in u_n. In the region where the opportunity is traded because u_n is above the threshold, both zeroth- and first-order terms in expected excess returns

are independent of u_n. The arbitrageur position $x_{n,m,t}$ increases in u_n because it is zero in the region where the opportunity is not traded and because (A.8) implies that $x^0_{n,t}$ increases in u_n in the region where the opportunity is traded.

The comparative statics with respect to μ_n follow from the zeroth-order terms or are trivial. For a traded opportunity, (A.8) implies that $x^0_{n,t}$ increases in μ_n. For a non-traded opportunity $x_{n,m,t}$ is zero, and hence increases weakly in μ_n. ∎

Proof of Proposition 10: The proposition follows by combining the AR(1) dynamics (45) of arbitrageur wealth W_t with the dependence of spreads, expected excess returns, and arbitrageur positions on W_t. This dependence can be deduced from the zeroth-order terms. As shown in the proof of Proposition 9, $\Phi^0_{n,t}$ and $\phi^0_{n,m,t}$ decrease in W_t. Moreover, since Π_t decreases in W_t (A.8) implies that $x^0_{n,t}$ increases in W_t. ∎

Proof of Proposition 11: The comparative statics with respect to $\bar{\epsilon}_n$ follow from the zeroth-order terms. Since $\Phi^0_{n,t} = \Pi_t \bar{\epsilon}_n$ for traded opportunities, $\Phi^0_{n,t}$ is more sensitive to changes in W_t for an opportunity with higher $\bar{\epsilon}_n$. The higher sensitivity of $\Phi^0_{n,t}$ translates to a higher sensitivity of $\phi^0_{n,m,t}$ because of (A.52). Eq. (A.8) implies that $x^0_{n,t}$ is less sensitive to changes in W_t.

The comparative statics with respect to $\bar{\eta}_n$ follow from the first-order terms. Since (A.34) implies that $\frac{\partial^2 \Phi^1_{n,m,t}}{\partial \Pi_t \partial \lambda_n} > 0$, $\Phi^1_{n,m,t}$ is more sensitive to changes in W_t for a traded opportunity with higher η_n. The higher sensitivity of $\Phi^1_{n,m,t}$ translates to a higher sensitivity of $\phi^1_{n,m,t}$ because of (48), and to a higher sensitivity of $x^0_{n,m,t}$ because of (A.35) and (A.36).

The comparative statics of spreads with respect to m follow from $\frac{d\phi^0_{n,m+1,t}}{dW_t} < \frac{d\phi^0_{n,m,t}}{dW_t} < 0$. Those of expected excess returns and positions follow from the first-order terms. Eq. (A.34) implies that

$$\Phi^1_{n,m+1,t} - \Phi^1_{n,m,t} = -\Pi_t \frac{d(\phi^0_{n,m,t+1} - \phi^0_{n,m-1,t+1})}{dW_{t+1}} \sum_{n' \in N} M_{n'} \sigma_{n',t} \lambda_{n'}$$

$$= -\frac{\Pi_t}{(1+r)^{m+1}} \frac{d\Phi^0_{n,t+m+1}}{dW_{t+1}} \sum_{n' \in N} M_{n'} \sigma_{n',t} \lambda_{n'}$$

$$\Rightarrow \frac{d\Phi^1_{n,m+1,t}}{dW_t} - \frac{d\Phi^1_{n,m,t}}{dW_t} = -\frac{\frac{d\Pi_t}{dW_t}}{(1+r)^{m+1}} \frac{d\Phi^0_{n,t+m+1}}{dW_{t+1}} \sum_{n' \in N} M_{n'} \sigma_{n',t} \lambda_{n'}$$

$$- \frac{\Pi_t}{(1+r)^{m+1}} \frac{d}{dW_t} \left(\frac{d\Phi^0_{n,t+m+1}}{dW_{t+1}} \sum_{n' \in N} M_{n'} \sigma_{n',t} \lambda_{n'} \right), \tag{A.55}$$

where the second step follows from (A.51). If Π_t is positive but close to zero (as is the case when the steady-state value W of W_t is smaller than but close to W_c), then the first term in the right-hand side of (A.55) is negative and bounded away from zero, while the second term is close to zero. Hence, $\frac{d\Phi^1_{n,m+1,t}}{dW_t} < \frac{d\Phi^1_{n,m,t}}{dW_t}$. Since the zeroth-order term in $\Phi_{n,m,t}$ is independent of m and $\frac{d\Phi^0_{n,t}}{dW_t} < 0$, $\frac{d\Phi^1_{n,m+1,t}}{dW_t} < \frac{d\Phi^1_{n,m,t}}{dW_t}$ implies $\frac{d\Phi_{n,m+1,t}}{dW_t} < \frac{d\Phi_{n,m,t}}{dW_t} < 0$. Hence, $\Phi_{n,m,t}$ is more sensitive to changes in W_t for a traded opportunity with higher m. To show that positions are also more sensitive, we note from (A.35) and (A.36) that

$$\frac{dx^1_{n,m+1,t}}{dW_t} - \frac{dx^1_{n,m,t}}{dW_t} = -\frac{\mu_n \left(\frac{d\Phi^1_{n,m+1,t}}{dW_t} - \frac{d\Phi^1_{n,m,t}}{dW_t}\right)}{\bar{\epsilon}_n^2 f''\left[\left(y^0_{n,t} + u_n\right)\bar{\epsilon}_n\right]} + \frac{\mu_n \left(\Phi^1_{n,m+1,t} - \Phi^1_{n,m,t}\right)}{\bar{\epsilon}_n f''\left[\left(y^0_{n,t} + u_n\right)\bar{\epsilon}_n\right]} f'''\left[\left(y^0_{n,t} + u_n\right)\bar{\epsilon}_n\right] \frac{dy^0_{n,t}}{dW_t}. \tag{A.56}$$

If Π_t is close to zero, so is $\Phi^0_{n,t}$ (Proposition 3) and hence $y^0_{n,t} + u_n$ (Proposition 1). Since $f(y)$ is symmetric around the vertical axis, $f'''(0) = 0$ and hence $f'''\left[\left(y^0_{n,t} + u_n\right)\bar{\epsilon}_n\right]$ is close to zero when $y^0_{n,t} + u_n$ is close to zero. Since the first term in the right-hand side of (A.56) is positive and the second is close to zero, $\frac{dx^1_{n,m+1,t}}{dW_t} > \frac{dx^1_{n,m,t}}{dW_t}$. Since the zeroth-order term in $x_{n,m,t}$ is independent of m and $\frac{dx^0_{n,t}}{dW_t} > 0$, $\frac{dx^1_{n,m+1,t}}{dW_t} > \frac{dx^1_{n,m,t}}{dW_t}$ implies $\frac{dx_{n,m+1,t}}{dW_t} > \frac{dx_{n,m,t}}{dW_t} > 0$. Hence, $x_{n,m,t}$ is more sensitive to changes in W_t for a traded opportunity with higher m. ∎

Proof of Corollary 2: The proof follows from the argument in the paragraph just before the proposition. ∎

Proof of Proposition 12: Using (A.8), we can write (46) for a traded opportunity (n, m) as

$$\begin{aligned}\sigma_n &= \frac{2\beta u_n \left[u_n - \frac{(f')^{-1}(\epsilon_n(\Pi))}{(f')^{-1}(\Pi)}\right]}{1 + 2\beta \sum_{(n',m)\in T} \mu_{n'} \left(u_{n'} - \frac{(f')^{-1}(\epsilon_{n'}(\Pi))}{(f')^{-1}(\Pi)}\right)^2} \frac{d\phi^0_{n',m-1,t+1}}{dW_{t+1}} \\ &= \frac{2\beta u_n \left[u_n - \frac{(f')^{-1}(\epsilon_n(\Pi))}{(f')^{-1}(\Pi)}\right]}{1 + 2\beta \sum_{n'\in N} \mu_{n'} \left(u_{n'} - \frac{(f')^{-1}(\epsilon_{n'}(\Pi))}{(f')^{-1}(\Pi)}\right)^2} \sum_{m=1}^{M_n - 1} \frac{d\phi^0_{n',m,t}}{dW_t}\end{aligned} \tag{A.57}$$

Since $\Phi^0_{n,t} = \Pi_t \epsilon_n$, (A.52) implies that

$$\frac{d\phi^0_{n,m,t}}{dW_t} = \frac{d\Pi_t}{dW_t}\epsilon_n \sum_{s=0}^{m-1} \frac{1}{(1+r)^{s+1}} \frac{dW_{t+s}}{dW_t}$$
$$= \frac{d\Pi_t}{dW_t}\epsilon_n \sum_{s=0}^{m-1} \frac{1}{(1+r)^{s+1}} \prod_{s'=0}^{s-1} \frac{dW_{t+s'+1}}{dW_{t+s'}}. \tag{A.58}$$

Eq. (A.9) implies that in steady state

$$\frac{\partial \Pi_t}{\partial W_t} = -\frac{1}{\frac{2}{1+r}\sum_{n\in N}\mu_n M_n \left[u_n\epsilon_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}. \tag{A.59}$$

The argument used in the proof of Proposition A.1 to show the AR(1) dynamics of W_t implies that in steady state and for $s \geq 0$

$$\frac{dW_{t+s+1}}{dW_{t+s}} = \frac{dW_{t+1}}{dW_t} = \rho. \tag{A.60}$$

Using (A.32), we find

$$\rho = 1 + \frac{W\Pi'}{1-\Pi}$$
$$= 1 - \frac{\frac{2}{1+r}\sum_{n\in N}\mu_n M_n \left[u_n\epsilon_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}{\frac{2}{1+r}\sum_{n\in N}\mu_n M_n \left[u_n\epsilon_n - (f')^{-1}(\Pi)\right]}$$
$$= \frac{\sum_{n\in N}\mu_n M_n \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}{\sum_{n\in N}\mu_n M_n \left[u_n\epsilon_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]} \tag{A.61}$$

where the second step follows from (A.59). Using (A.59)-(A.61), we can write (A.58) as

$$\frac{\partial \phi^0_{n,m,t}}{\partial W_t} = -\frac{\epsilon_n \sum_{s=0}^{m-1}\frac{\rho^s}{(1+r)^s}}{2\sum_{n'\in N}\mu_{n'}M_{n'}\left[u_{n'}\epsilon_{n'} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}, \tag{A.62}$$

and (A.57) as

$$\sigma_n = \frac{2\beta\mu_n u_n - \left[(f')^{-1}(\Pi) - \bar{\epsilon}_n\right]\sum_{m=1}^{M_n}\sum_{s=0}^{m-1}\frac{\rho^s}{(1+r)^s}}{\beta\sum_{n'\in N}\mu_{n'}\left[u_{n'}\bar{\epsilon}_{n'} - (f')^{-1}(\Pi)\right]M_{n'}} \cdot \tag{A.63}$$
$$1 - \frac{\left[(f')^{-1}(\Pi) - \bar{\epsilon}_n\right]\frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}{\sum_{n'\in N}\mu_{n'}M_{n'}\left[u_{n'}\bar{\epsilon}_{n'} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}$$

Under segmentation, the wealth $W_{n,t}$ of arbitrageurs in market $n \in N$ evolves according to the AR(1) process

$$W_{n,t+1} = W_n + \rho_n(W_{n,t} - W_n) + v_n\bar{\eta} + \sigma_{n,n}\sum_{m=1}^{M_n}\eta_{n,m-1,t+1} + o(\bar{\eta}). \tag{A.64}$$

The counterparts of (A.61)-(A.63) can be obtained by removing the summation over N, and are

$$\rho_n = \frac{\frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}{u_n\bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}, \tag{A.65}$$

$$\frac{\partial\phi^0_{n,m,t}}{\partial W_{n,t}} = -\frac{\left[\bar{\epsilon}_n - \sum_{s=0}^{m-1}\frac{\rho^s}{(1+r)^s}\right]}{2\mu_n M_n\left[u_n\bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}, \tag{A.66}$$

$$\sigma_{n,n} = \frac{2\beta u_n u_n - \left[(f')^{-1}(\Pi) - \bar{\epsilon}_n\right]\sum_{m=1}^{M_n-1}\sum_{s=0}^{m-1}\frac{\rho^s}{(1+r)^s}}{\beta\left[u_n\bar{\epsilon}_n - (f')^{-1}(\Pi)\right]M_n}, \tag{A.67}$$
$$1 - \frac{\left[(f')^{-1}(\Pi) - \bar{\epsilon}_n\right]\frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}{M_n\left[u_n\bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}$$

respectively.

Suppose next that opportunities are symmetric, and denote by $(\bar{\epsilon}, \bar{\eta}, \mu, u, M)$ the common values of $(\bar{\epsilon}_n, \bar{\eta}_n, \mu_n, u_n, M_n)$ and by σ^2_η the common variance of $\eta_{n,m,t}$. Symmetry and an interior steady

state imply that all opportunities are traded, i.e., $N = \{1, .., N\}$. Moreover, for all $n = 1, .., N$,

$$\rho_n = \rho = \frac{\frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}{u\bar{\epsilon} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}} \tag{A.68}$$

$$\frac{\partial \phi^0_{n,m,t}}{\partial W_t} = \frac{1}{N}\frac{\partial \phi^0_{n,m,t}}{\partial W_{n,t}} = -\frac{\left[\bar{\epsilon} - \sum_{s=0}^{m-1}\frac{\rho^s}{(1+r)^s}\right]}{2\mu M\left[u\bar{\epsilon} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]} \equiv \phi_{W,m} \tag{A.69}$$

$$\sigma_n = \sigma_{n,n} = \frac{2\beta\mu\left[u - \frac{(f')^{-1}(\Pi)}{\bar{\epsilon}}\sum_{m=1}^{M-1}\sum_{s=0}^{m-1}\frac{\rho^s}{(1+r)^s}\right]}{\beta[u\bar{\epsilon} - (f')^{-1}(\Pi)]\left[1 - \frac{\sum_{m=1}^{M-1}\sum_{s=0}^{m-1}\frac{\rho^s}{(1+r)^s}}{M\left[u\bar{\epsilon} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}\right]} \equiv \sigma, \tag{A.70}$$

because of (A.61) and (A.65), (A.62) and (A.66), and (A.63) and (A.67), respectively.

The variance of arbitrageur wealth under integration follows from the mean-reverting dynamics (45), and is

$$\mathrm{Var}(W_t) = \frac{\sum_{(n,m)\in T}\sigma_n^2 \mathrm{Var}(\eta_{n,m,t})}{1-\rho^2} + o\left(\eta^2\right). \tag{A.71}$$

Likewise, the variance of arbitrageur wealth under segmentation can be computed from (A.64), and is

$$\mathrm{Var}(W_{n,t}) = \frac{\sum_{m=1}^{M_n}\sigma_{n,n}^2 \mathrm{Var}(\eta_{n,m,t})}{1-\rho_n^2} + o\left(\eta^2\right). \tag{A.72}$$

Under symmetry, the highest-order term in (A.71) is $\frac{NM\sigma^2\sigma_\eta^2}{1-\rho^2}$, and that in (A.72) is $\frac{M\sigma^2\sigma_\eta^2}{1-\rho^2}$. Since the wealth of the arbitrageurs who are in market n under segmentation is $\frac{W_t}{N}$ under integration, the highest-order term in the variance of those arbitrageurs' wealth is $\frac{M\sigma^2\sigma_\eta^2}{N(1-\rho^2)}$ under integration and $\frac{M\sigma^2\sigma_\eta^2}{1-\rho^2}$ under segmentation. Therefore, the variance of each arbitrageur's wealth under integration is N times smaller than under segmentation. The variance of wealth of an arbitrageur who diversifies across all opportunities equally under segmentation is N times smaller than without diversification because spreads are independent across n.

The variance of the spread associated to opportunity (n,m) is

$$\mathrm{Var}(\phi_{n,m,t}) = \left(\frac{d\phi_{n,m,t}}{dW_t}\right)^2 \mathrm{Var}(W_t) + o\left(\bar{\eta}^2\right) \tag{A.73}$$

under integration, and

$$\mathrm{Var}(\phi_{n,m,t}) = \left(\frac{d\phi_{n,m,t}}{dW_{n,t}}\right)^2 \mathrm{Var}(W_{n,t}) + o\left(\bar{\eta}^2\right) \tag{A.74}$$

under segmentation, where the derivative is evaluated at $W_t = W$. Under symmetry, the highest-order term in (A.73) is $\frac{NM\sigma^2\sigma_\eta^2\phi_{W,m}^2}{1-\rho^2}$, and that in (A.74) is $\frac{N^2M\sigma^2\sigma_\eta^2\phi_{W,m}^2}{1-\rho^2}$. Therefore, the variance of spreads under integration is N times smaller than under segmentation. ∎

Proof of Proposition 13: Eqs. (A.61), (A.63), and (A.71) imply that the variance of the arbitrageur wealth under integration is

$$\mathrm{Var}(W_t) = \frac{4\beta^2 \sum_{n\in N} \mu_n^2 M_n \left[u_n - \frac{(f')^{-1}(\Pi)}{\bar{\epsilon}_n}\right]^2 \sigma_{\eta,n}^2}{\left(\beta \sum_{n\in N} \mu_n [u_n \bar{\epsilon}_n - (f')^{-1}(\Pi)] \sum_{m=1}^{M_n - 1} \sum_{s=0}^{m-1} \frac{\sum_{n\in N} \mu_n M_n \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}{(1+r)\left[\sum_{n\in N} \mu_n M_n \left[u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]\right]}\right)^2}{\left(1 - \frac{\left[\sum_{n\in N} \mu_n M_n \left[u_n \bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]\right]^2}{\sum_{n\in N} \mu_n M_n \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}\right)^2} + o\left(\bar{\eta}^2\right), \tag{A.75}$$

where we denote by $\sigma_{\eta,n}^2$ the common variance of $\eta_{n,m,t}$ across $m = 1,..,M_n$. Likewise, (A.65), (A.67), and (A.72) imply that in a segmented market n with traded opportunities, the variance of

the wealth of arbitrageurs is

$$\mathrm{Var}(W_{n,t}) = \frac{4\beta^2\mu_n^2 M_n^2 \left[u_n - \frac{(f')^{-1}(\Pi)}{\bar{\epsilon}_n}\right]^2 \sigma^2_{\eta,n}}{\left(1 - \frac{\beta\left[u_n\bar{\epsilon}_n - (f')^{-1}(\Pi)\right]\sum_{m=1}^{M_n-1}\sum_{s=0}^{m-1}\left(\frac{1}{(1+r)}\right)^s \left[u_n\bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}{M_n\left[u_n\bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}\right)^2} + o\left(\sigma_\eta^2\right). \quad \text{(A.76)}$$

Combining (A.75) with (A.62) and (A.73), we find that the variance of the spread associated to a traded opportunity (n,m) under integration is

$$\mathrm{Var}(\phi_{n,m,t}) = \frac{\beta^2\bar{\epsilon}_n^2\left[\sum_{n'\in N}\mu_{n'}M_{n'}\left[u_{n'} - \frac{(f')^{-1}(\Pi)}{\bar{\epsilon}_{n'}}\right]\right]^2\sigma^2_{\eta,n'}\sum_{s=0}^{m-1}\left(\frac{\sum_{n'\in N}\mu_{n'}M_{n'}\frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}{(1+r)\sum_{n'\in N}\mu_{n'}M_{n'}\left[u_{n'}\bar{\epsilon}_{n'} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}\right)^s}{\left(1 - \frac{\beta\sum_{n'\in N}\mu_{n'}\left[u_{n'}\bar{\epsilon}_{n'} - (f')^{-1}(\Pi)\right]\sum_{m=1}^{M_n-1}\sum_{s=0}^{m-1}\left(\frac{1}{(1+r)}\right)^s\left[\sum_{n'\in N}\mu_{n'}M_{n'}\left[u_{n'}\bar{\epsilon}_{n'} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]\right]}{\sum_{n'\in N}\mu_{n'}M_{n'}\left[u_{n'}\bar{\epsilon}_{n'} - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}\right)^2} + o\left(\sigma_\eta^2\right). \quad \text{(A.77)}$$

Likewise, combining (A.76) with (A.66) and (A.74), we find that the variance of the spread associated to opportunity (n,m) under segmentation is

$$\mathrm{Var}(\phi_{n,m,t}) = \frac{\beta^2\mu_n^2 M_n\left[u_n\bar{\epsilon}_n - (f')^{-1}(\Pi)\right]^2\sigma^2_{\eta,n}\sum_{s=0}^{m-1}\left(\frac{\frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}}{(1+r)\left[u_n\bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}\right)^s}{\left(1 - \frac{\beta\left[u_n\bar{\epsilon}_n - (f')^{-1}(\Pi)\right]\sum_{m=1}^{M_n-1}\sum_{s=0}^{m-1}\left(\frac{1}{(1+r)}\right)^s\left[u_n\bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}{M_n\left[u_n\bar{\epsilon}_n - (f')^{-1}(\Pi) + \frac{1-\Pi}{f''[(f')^{-1}(\Pi)]}\right]}\right)^2} + o\left(\sigma_\eta^2\right). \quad \text{(A.78)}$$

(A.78)

If $u_n\bar{\epsilon}_n - (f')^{-1}(\Pi)$ or $\sigma_{\eta,n}/\eta$ are close to zero for n but are sufficiently large for $n' \in N \setminus \{n\}$, then $\text{Var}(\phi_{n,m,t})$ is close to zero under segmentation but not under integration. Hence, $\text{Var}(\phi_{n,m,t})$ is larger under integration. ∎

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 $t \in \{\underline{t}_\omega, .., \bar{t}_\omega - 1\}$, where $\underline{t}_\omega$ occurs before the first positive payment and $\bar{t}_\omega$ is when the last positive payment is made. The period $\bar{t}_\omega$ can be infinite, and if it is finite we set $q_{\bar{\omega},t_\omega} = 0$. 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

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

where $x^c_{\omega',t} \geq 0$ 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 $x^c_{\omega',t}$ must satisfy

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

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

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

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:

$$W_t \geq \sum_{i \in I} \left[\sum_{\omega \in \Omega_{i,t}} x_{\omega,t} q_{\omega,t} + v_{i,t} \right]. \tag{B.3}$$

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

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

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

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

Eqs. (B.5) and (B.6) are counterparts of (14) and (16), with the positions in the contracts replacing those in the risky assets.

Defi 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 and the variables $\eta_{i,t}$ are equal to zero. Given symmetry, the binomial assumption implies 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 3.2 and the prices $q_{\omega,t}$ of all contracts $\omega \in \Omega_{i,t}$ are given by*

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

where

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

and $y_{i,t}$ is as in Section 3.2.

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 (15) with $A = r\alpha$ and $F_{i,t}$ given by (A.21), and use the budget constraint (B.5) instead of (14). Optimal consumption is given by

$$c_{i,t+1} = \frac{A\left[(1+r)w_{i,t} + \sum_{\omega\in\Omega_{i,t}} y_{\omega,t}[\pi_{\omega,t+1} + q_{\omega,t+1} - (1+r)q_{\omega,t}] + u_{i,t}\epsilon_{i,t+1}\right] + F_{i,t+1} + \log\left(\frac{\alpha}{A}\right)}{\alpha + A}, \tag{B.9}$$

which is the counterpart of (A.15). Optimal positions in the contracts solve

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

which is the counterpart of (A.16) after omitting terms that are known in period *t*. The first-order condition with respect to $y_{\omega,t}$ is

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

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

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

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

$$\sum_{\omega\in\Omega_{i,t}} y_{\omega,t}[\pi_{\omega,t+1}+q_{\omega,t+1}-(1+r)q_{\omega,t}]=y_{i,t}\epsilon_{i,t+1}+G_t, \tag{B.13}$$

where 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 3.2. Indeed, multiplying (B.12) by $y_{\omega,t}$ and summing across $\omega \in \Omega_{i,t}$, we find

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

Moreover, the maximization in (A.16) implies that

$$\mathrm{E}_t\left\{y_{i,t}(\Phi_{i,t}+\epsilon_{i,t+1})\exp\left(-\frac{\alpha A}{\alpha+A}(y_{i,t}+u_{i,t})\epsilon_{i,t+1}\right)\right\}=0. \tag{B.15}$$

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 $G_t = \Phi_{i,t}$. Substituting $G_t = \Phi_{i,t}$ into (B.13), we find that budget constraint (B.5) of i-investors becomes identical to the budget constraint (14) in Section 3.2. Since the dynamics of the wealth of i-investors are the same as in Section 3.2, the conjectured value function (15) 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 3.2.

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

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

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

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

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 $\hat{\omega}$ that is traded in market i and has binary payoffs. The payoffs of $\hat{\omega}$ are

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

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

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 $\hat{\omega}$ in period t is

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

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 $\hat{\omega}$ is

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

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

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

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

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

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 $\overline{Q}_{i,t+1} + (1+r)v_{i,t}$ 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

$$\begin{aligned}
&\frac{\exp(-Z_{-i,t})\left[\overline{Q}_{i,t+1}+(1+r)v_{i,t}\right]+\exp(Z_{-i,t})\left[\underline{Q}_{i,t+1}+(1+r)v_{i,t}\right]}{(1+r)\left[\exp(-Z_{-i,t})+\exp(Z_{-i,t})\right]}\\
&=\frac{\exp(Z_{i,t})\overline{Q}_{i,t+1}+\exp(-Z_{i,t})\underline{Q}_{i,t+1}}{(1+r)\left[\exp(Z_{i,t})+\exp(Z_{i,t})\right]}+v_{i,t}\\
&=\frac{\overline{Q}_{i,t+1}+\underline{Q}_{i,t+1}}{1+r}-Q_{i,t}+v_{i,t}\\
&\equiv Q_{-i,t}+v_{i,t}, \qquad \text{(B.19)}
\end{aligned}$$

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

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

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

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

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

$$\max\{Q_{i,t+1}, \underline{Q}_{i,t+1}\} + (1+r)v_{i,t} \leq (1+r)v_{-i,t}$$

$$\Leftrightarrow \max\{Q_{i,t+1}, \underline{Q}_{i,t+1}\} \leq (1+r)(Q_{i,t} + Q_{-i,t} + v_{i,t})$$

$$\Leftrightarrow \max\{Q_{i,t+1}, \underline{Q}_{i,t+1}\} \leq \bar{Q}_{i,t+1} + \underline{Q}_{i,t+1} + (1+r)v_t, \tag{B.21}$$

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 $\hat{\omega}$ are non-negative.

The riskless portfolio that dominates the original risky portfolio in market i consists of a half-unit long position in $\hat{\omega}$ and a half-unit short position in $\hat{\omega}_{-i}$. Since a unit long position in $\hat{\omega}$ 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 $\hat{\omega}$ 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

$$W_{t+1} = (1+r)W_t + (1+r)\sum_{i\in A_t} x_{i,t}(q_{-i,t} - q_{i,t}) - c_{t+1}$$
$$= (1+r)W_t + 2\sum_{i\in A_t} x_{i,t}[\epsilon_{\bar{i}} - (1+r)q_{i,t}] - c_{t+1}, \tag{B.22}$$

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

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

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

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

Eqs. (B.22) and (B.3) become identical to (27) and (32), respectively, by setting

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

Because of this equivalence, if the dynamics of $\Phi_{i,t}$ are as in Section 3.2, 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

$$\Phi_{i,t} = \epsilon_i - \frac{2\exp(-Z_{i,t})\epsilon_i}{\exp(-Z_{i,t})+\exp(Z_{i,t})} = \epsilon_i\frac{\exp(Z_{i,t}) - \exp(-Z_{i,t})}{\exp(-Z_{i,t})+\exp(Z_{i,t})}.$$

This coincides with $\Phi_{i,t}$ given by (25) 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 3.2. Eq. (B.13) implies that the optimal positions of i-investors are $y_{i,t}$, as in Section 3.2. 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 $\epsilon_{t+1} = \epsilon_i$ and $2\epsilon_i$ if $\epsilon_{t+1} = -\epsilon_i$. 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

Avdjiev, Stefan, Wenxin Du, Catherine Koch, and Huyn Song Shin, 2016, The dollar, bank leverage and the deviation from covered interest parity, working paper 592 Bank of International Settlements.

Baba, Naohiko, and Frank Packer, 2009, From turmoil to crisis: Dislocations in the fx swap market before and after the failure of Lehman Brothers, *Journal of International Money and Finance* 28, 1350–1374.

Borio, Claudio, Robert McCauley, Patrick McGuire, and Vladyslav Sushko, 2016, Covered interest parity lost: Understanding the cross-currency basis, *BIS Quarterly Review* September, 45–64.

Brumm, Johannes, Michael Grill, Felix Kubler, and Karl Schmedders, 2015, Collateral requirements and asset prices, *International Economic Review* 56, 1–25.

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.

Caballero, Ricardo, and Alp Simsek, 2017, A model of fickle capital flows and retrenchment, working paper MIT.

Chabakauri, Georgy, 2013, Asset pricing with heterogeneous investors and portfolio constraints, *Review of Financial Studies* 26, 3104–3141.

——— , and Brandon Han, 2017, Capital requirements and asset prices, working paper London School of Economics.

Cho, Thummim, 2016, Turning alphas into betas: Arbitrage and endogenous risk, working paper Harvard University.

Coffey, Niall, Warren Hrung, and Asani Sarkar, 2009, Capital constraints, counterparty risk, and deviations from covered interest rate parity, staff report 393 Federal Reserve Bank of New York.

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.

Du, Wenxin, Alexander Tepper, and Adrien Verdelhan, 2016, Deviations from covered interest rate parity, working paper MIT.

Duffie, Darrell, 2010, Presidential address: Asset price dynamics with slow-moving capital, *Journal of Finance* 65, 1237–1267.

——— , and Bruno Strulovici, 2012, Capital mobility and asset pricing, *Econometrica* 80, 2469–2509.

Fleckenstein, Matthias, Francis Longstaff, and Hanno Lustig, 2014, The tips-treasury bond puzzle, *Journal of Finance* 69, 2151–2197.

Fontaine, Jean-Sebastien, and Rene Garcia, 2012, Bond liquidity premia, *Review of Financial Studies* 25, 1207–1254.

Fostel, Ana, and John Geanakoplos, 2015, Leverage and default in binomial economies: A complete characterization, *Econometrica* 83, 2191–2229.

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, 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, 2015, Dynamic competitive economies with complete markets and collateral constraints, *Review of Economic Studies* 82, 1119–1153.

Greenwood, Robin, Samuel Hanson, and Gordon Liao, 2015, Price dynamics in partially segmented markets, working paper Harvard Business School.

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.

Guembel, Alexander, and Oren Sussman, 2015, A welfare analysis of fragmented liquidity markets, working paper Toulouse School of Economics.

He, Zhiguo, and Arvind Krishnamurthy, 2013, Intermediary asset pricing, *American Economic Review* 103, 1–43.

Hu, Xing, Jun Pan, and Jiang Wang, 2013, Noise as information for illiquidity, *Journal of Finance* 68, 2341–2382.

Iida, Tomoyuki, Takeshi Kimura, and Nao Sudo, 2016, An upsurge in a cip deviation during the noncrisis period and the role of divergence in monetary policy, working paper 16-E-14 Bank of Japan.

Kondor, Peter, and Dimitri Vayanos, 2016, Liquidity risk and the dynamics of arbitrage capital, working paper London School of Economics.

Kyle, Albert, and Wei Xiong, 2001, Contagion as a wealth effect, *Journal of Finance* 56, 1401–1440.

Liao, Gordon, 2016, Credit migration and covered interest rate parity, working paper 17 Harvard Business School.

Mancini Griffoli, Tommaso, and Angelo Ranaldo, 2012, Limits to arbitrage during the crisis: Funding liquidity constraints and covered interest parity, working paper 1212 University of St Gallen School of Finance.

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* 27, 2961–2997.

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.

Sushko, Vladyslav, Claudio Borio, Robert McCauley, and Patrick McGuire, 2016, Failure of covered interest parity: FX hedging demand and costly balance sheets, working paper Bank of International Settlements.

Wagner, Wolf, 2011, Systemic liquidation risk and the diversity-diversification trade-off, *Journal of Finance* 66, 1141–1175.

Xiong, Wei, 2001, Convergence trading with wealth effects: An amplification mechanism in financial markets, *Journal of Financial Economics* 62, 247–292.