The Journal of FINANCE
The Journal of THE AMERICAN FINANCE ASSOCIATION

THE JOURNAL OF FINANCE • VOL. LXXIV, NO. 3 • JUNE 2019

Liquidity Risk and the Dynamics of Arbitrage Capital

PÉTER KONDOR and DIMITRI VAYANOS*

ABSTRACT

We develop a continuous-time model of liquidity provision in which hedgers can trade multiple risky assets with arbitrageurs. Arbitrageurs have constant relative risk-aversion (CRRA) utility, while hedgers' asset demand is independent of wealth. An increase in hedgers' risk aversion can make arbitrageurs endogenously more risk-averse. Because arbitrageurs generate endogenous risk, an increase in their wealth or a reduction in their CRRA coefficient can raise risk premia despite Sharpe ratios declining. Arbitrageur wealth is a priced risk factor because assets held by arbitrageurs offer high expected returns but suffer the most when wealth drops. Aggregate illiquidity, which declines in wealth, captures that factor.

LIQUIDITY IN FINANCIAL MARKETS IS OFTEN provided by specialized agents such as market makers, trading desks in investment banks, and hedge funds. Adverse shocks to the capital of these agents cause liquidity to decline and risk premia to increase. Conversely, movements in the prices of assets held by liquidity providers feed back into these agents' capital.[1]

*Péter Kondor is at the London School of Economics and CEPR. Dimitri Vayanos is at the London School of Economics, CEPR, and NBER. The authors thank Edina Berlinger; Bruno Biais; Itamar Dreschler; Nicolae Garleanu; Florian Heider; Pete Kyle; Stavros Panageas; Anna Pavlova; Jean-Charles Rochet; Wei Xiong; Hongjun Yan; an anonymous Associate Editor; two anonymous referees; seminar participants at Atlanta Fed, Bank of Canada, Cass, Columbia, Copenhagen, Duke, HKU, Imperial, LBS, LSE, Minneapolis Fed, Montreal, NUS, Ohio State, Tinbergen, UIUC, UNC, Wash U, Wharton, and Zurich; and participants at the AEA, AFA, Arne Ryde, BIS, Cambridge INET, CRETE, EIEF Market Microstructure, ESEM, ESSFM Gerzensee, EWFC, NBER Asset Pricing, QMUL, and SED conferences for helpful comments. They are grateful to Mihail Zervos for generous and valuable guidance with the proofs on the differential equations. Péter Kondor acknowledges financial support from the European Research Council (Starting Grant #336585). Both authors acknowledge financial support from the Paul Woolley Centre at the LSE. Both authors have read the *Journal of Finance*'s disclosure policy and have no conflicts of interest to disclose.

[1] A growing empirical literature documents the relationships between the capital of liquidity providers, the liquidity that these agents provide to other participants, and assets' risk premia. For example, Comerton-Forde et al. (2010) find that bid-ask spreads quoted by specialists in the New York Stock Exchange widen when specialists experience losses. Aragon and Strahan (2012) find that following the collapse of Lehman Brothers in 2008, hedge funds doing business with Lehman experienced a higher probability of failure and the liquidity of the stocks that they were trading declined. Jylha and Suominen (2011) find that outflows from hedge funds that perform the carry trade predict poor performance of that trade, with low interest-rate currencies appreciating

DOI: 10.1111/jofi.12757

In this paper, we study the dynamic interactions between liquidity providers' capital, the liquidity that these agents provide to other participants, and assets' risk premia. We build a framework with minimal frictions, in particular, no asymmetric information or leverage constraints. The capital of liquidity providers matters in our model only because of standard wealth effects. We depart from most frictionless asset-pricing models, however, by fixing the riskless rate and by suppressing wealth effects for agents other than the liquidity providers. These assumptions are sensible when focusing on shocks to the capital of liquidity providers in an asset class rather than in the entire asset universe.

Our combination of assumptions allows us to prove general analytical results on equilibrium prices and allocations. In particular, we characterize how liquidity providers' risk appetite, the endogenous risk that they generate, and the pricing of that risk depend on both liquidity demanders' characteristics and liquidity providers' capital. We also show that the capital of liquidity providers is the single priced risk factor, and that liquidity aggregated over the assets that we consider captures that factor because it increases in capital. Our results thus suggest that a priced liquidity risk factor may arise even with minimal frictions.

We assume a continuous-time infinite-horizon economy. There is a riskless asset with an exogenous constant return and multiple risky assets whose prices are determined endogenously in equilibrium. There are two sets of competitive agents: hedgers, who receive a risky income flow and seek to reduce their risk by participating in financial markets, and arbitrageurs, who take the other side of the trades that hedgers initiate. Arbitrageurs can be interpreted, for example, as speculators in futures markets. We consider two specifications for hedgers' preferences. Hedgers can be "long-lived" and maximize constant absolute risk-aversion (CARA) utility over an infinite consumption stream, or they can be "short-lived" and maximize a mean-variance objective over changes in wealth in the next instant. Under both specifications, hedgers' demand for insurance is independent of their wealth. In contrast, because arbitrageurs maximize constant relative risk-aversion (CRRA) utility over consumption, the supply of insurance depends on their wealth.

Arbitrageur wealth impacts equilibrium prices and allocations, and is the key state variable in our model. Solving for equilibrium amounts to solving a system of ordinary differential equations (ODEs) in wealth with boundary conditions at zero and infinity. While these ODEs include nonlinear terms, their structure makes it possible to prove general analytical results across the entire parameter space, for example, for all risk-aversion parameters of hedgers and arbitrageurs. In the case in which hedgers are short-lived, we show that a solution exists and we characterize how it depends on both wealth and the model parameters. Moreover, in both the short-lived and long-lived cases, we characterize the behavior of the solution close to the boundaries.

and high interest-rate ones depreciating. Acharya, Lochstoer, and Ramadorai (2013) find that risk premia in commodity futures markets are larger when broker-dealer balance sheets are shrinking.

Our analysis yields new insights on dynamic risk-sharing and asset pricing. We show that the risk aversion of arbitrageurs is the sum of their static CRRA coefficient and a forward-looking component that reflects intertemporal hedging. The latter component makes the risk aversion of arbitrageurs dependent on parameters of the economy that affect equilibrium prices. For example, when hedgers are more risk-averse, arbitrageurs become endogenously more risk-averse if their CRRA coefficient is less than one. This effect can be sufficiently strong to imply that more risk-averse hedgers may receive *less* insurance from arbitrageurs in equilibrium. Intuitively, when hedgers are more risk-averse, expected returns rise steeply following a decline in arbitrageur wealth. This makes arbitrageurs with CRRA coefficient less than one invest more conservatively so as to preserve wealth in bad states and earn the high returns.

On the asset-pricing side, we show that arbitrageurs generate endogenous risk, in the sense that changes in their wealth affect return variances and covariances through amplification and contagion mechanisms. Endogenous risk is small at both extremes of the wealth distribution: when wealth is close to zero, this is because arbitrageurs hold small positions and hence have a small impact on prices, whereas when wealth is close to infinity, this is because prices are insensitive to changes in wealth. The dependence of endogenous risk on arbitrageur wealth can give rise to hump-shaped patterns of variances, covariances, and correlations. It can also cause risk premia, defined as expected returns in excess of the riskless asset, to increase with arbitrageur wealth for small values of wealth even though Sharpe ratios decrease. We show that risk premia always exhibit this pattern when arbitrageurs' CRRA coefficient is small, and can exhibit it for larger values as well provided that hedgers are sufficiently risk-averse. In a similar spirit, we show that risk premia can be larger if arbitrageurs' CRRA coefficient is smaller—precisely because endogenous risk is larger.

Additional asset-pricing results concern liquidity risk and its relationship with expected returns. A large empirical literature documents that liquidity varies over time and in a correlated manner across assets within a class. Moreover, aggregate liquidity appears to be a priced risk factor and carry a positive premium: assets that underperform the most during times of low aggregate liquidity earn higher expected returns than assets with otherwise identical characteristics.[2] We map our model to this literature by defining liquidity based on the effect that hedgers have on prices. We show that liquidity is lower for assets with more volatile cash flows. It also decreases following losses by arbitrageurs, with this variation common across assets.

[2] Chordia, Roll, and Subrahmanyam (2000), Hasbrouck and Seppi (2001), and Huberman and Halka (2001) document the time variation of liquidity in the stock market and its correlation across stocks. Amihud (2002) and Hameed, Kang, and Viswanathan (2010) link time variation in aggregate liquidity to the returns on the aggregate stock market. Pastor and Stambaugh (2003) and Acharya and Pedersen (2005) find that aggregate liquidity is a priced risk factor in the stock market and carries a positive premium. Sadka (2010) and Franzoni, Nowak, and Phalippou (2012) find similar results for hedge fund and private equity returns, respectively. For more references, see Vayanos and Wang (2013), who survey the theoretical and empirical literature on market liquidity.

Expected returns in our cross-section of assets are proportional to the covariance with the portfolio of arbitrageurs, which is the single priced risk factor. This factor may be hard to measure empirically as the portfolio of arbitrageurs is unobservable. We show, however, that aggregate liquidity captures this factor. Indeed, because arbitrageurs sell a fraction of their portfolio following losses, assets that covary the most with their portfolio suffer the most when liquidity decreases. Thus, an asset's covariance with aggregate liquidity is proportional to its covariance with the portfolio of arbitrageurs. The covariances between an asset's liquidity and aggregate liquidity or return would not explain expected returns as well because they are proportional to the volatility of an asset's cash flows rather than to the asset's covariance with the arbitrageurs' portfolio. The covariance between an asset's return and other proxies for arbitrageur wealth used in recent empirical papers, such as the leverage of financial intermediaries, would also capture the true priced risk factor.[3]

We finally characterize when the long-run stationary distribution of arbitrageur wealth is nondegenerate and show that it can be bimodal. The stationary distribution can be nondegenerate because arbitrage activity is self-correcting: when wealth drops, arbitrageurs' future expected returns increase, which causes wealth to grow faster, and vice versa. The stationary density becomes bimodal when hedgers are sufficiently risk-averse. Indeed, because insurance provision in that case is more profitable, arbitrageur wealth grows fast and large values of wealth are more likely in steady state than intermediate values. At the same time, while profitability (per unit of wealth) is highest when wealth is small, wealth grows from small values slowly in absolute terms. Therefore, small values are more likely than intermediate values.

We see our work as bridging three relatively distinct streams of theoretical literature: on liquidity risk, on intermediary asset pricing, and on consumption-based asset pricing with heterogeneous agents. The first stream focuses on the pricing of liquidity risk in the cross-section of assets. In Holmstrom and Tirole (2001), firms avoid assets whose return is low when financial constraints are severe, and these assets offer high expected returns in equilibrium. The covariance between asset returns and liquidity (less severe constraints) is taken to be exogenous. This covariance is instead endogenous in our model because prices depend on arbitrageur wealth. This endogeneity is key for our results on a priced liquidity factor. In Amihud (2002) and Acharya and Pedersen (2005), illiquidity takes the form of exogenous time-varying transaction costs. An increase in the costs of trading an asset raises the expected return that investors require to hold it and lowers its price. A negative covariance between illiquidity and asset prices also arises in our model but due to an entirely different mechanism: low liquidity and low prices are endogenous symptoms of low arbitrageur wealth.

[3] Adrian, Etula, and Muir (2014) and He, Kelly, and Manela (2017) find that a single risk factor based on intermediary leverage can price a large cross-section of assets. As we argue in Section IV, our model is consistent with this finding and hence suggests that an explanation with minimal frictions may be possible.

The second stream links intermediary capital to liquidity and asset prices. In Gromb and Vayanos (2002), arbitrageurs intermediate trade between investors in segmented markets, and are subject to margin constraints. Because of the constraints, the liquidity that arbitrageurs provide to investors increases in their wealth. In Brunnermeier and Pedersen (2009), margin-constrained arbitrageurs intermediate trade in multiple assets across time periods. Assets with more volatile cash flows are more sensitive to changes in arbitrageur wealth.[4] In Kondor (2009), arbitrageurs with limited capital exploit a fundamentally riskless trading opportunity, present over an uncertain horizon. The price pressure they generate introduces risk and a corresponding premium into asset prices. Garleanu and Pedersen (2011) introduce margin constraints in an infinite-horizon setting with multiple assets. They show that assets with higher margin requirements earn higher expected returns and are more sensitive to changes in the wealth of the margin-constrained agents. This result is suggestive of a priced liquidity factor. In He and Krishnamurthy (2013), arbitrageurs can raise capital from other investors to invest in a risky asset over an infinite horizon, but this capital cannot exceed a fixed multiple of their internal capital. When arbitrageur wealth decreases, the constraint binds, and asset volatility and expected returns increase. In Brunnermeier and Sannikov (2014), arbitrageurs are more efficient holders of productive capital and can trade a risky claim to that capital with other investors. The long-run stationary distribution of arbitrageur wealth can have a bimodal density. A key difference between our paper and the above papers is that we derive the effects of arbitrage capital without imposing any constraints or contracting frictions.

Perhaps the closest papers to ours within the second stream of literature are Xiong (2001) and Kyle and Xiong (2001). In both papers, arbitrageurs with logarithmic utility over consumption can trade with exogenous long-term traders and noise traders over an infinite horizon.[5] The liquidity provided by arbitrageurs is increasing in their wealth, and asset volatilities are hump-shaped. Relative to these papers, we derive the demand of all traders from optimizing behavior and consider a general number of risky assets. We also show results analytically (rather than via numerical examples), and we do so for general parameter values within which logarithmic preferences are a restrictive special case. An additional related paper is Klimenko et al. (2016), in which risk-neutral banks face a downward-sloping demand for credit, and supply more credit when they have more capital. Banks in that paper parallel arbitrageurs in ours, and banks' borrowers parallel our hedgers. The closed-form solution derived in that paper when the riskless rate is zero is similar to the one derived in an early version of ours (Kondor and Vayanos (2014)) when arbitrageurs' CRRA coefficient and the riskless rate go to zero.

[4] In Gromb and Vayanos (2018), arbitrage spreads are positively related to the spreads' sensitivity to arbitrageur wealth because both characteristics are positively related to cash flow volatility and convergence horizon.

[5] Isaenko (2015) studies a related model in which long-term traders maximize CARA utility and face transaction costs.

Finally, our paper is related to the literature on consumption-based asset pricing with heterogeneous agents, for example, Dumas (1989), Wang (1996), Chan and Kogan (2002), Bhamra and Uppal (2009), Longstaff and Wang (2012), Basak and Pavlova (2013), Chabakauri (2013), Garleanu and Panageas (2015), and Ehling and Heyerdahl-Larsen (2017). In these papers, agents have CRRA utility and differ in their risk aversion. As the wealth of the less risk-averse agents increases, Sharpe ratios decrease, which can cause volatilities and correlations to be hump-shaped.[6] In contrast to these papers, we assume that only one set of agents has wealth-dependent risk aversion, and we fix the riskless rate.

We proceed as follows. In Section I, we present the model. In Sections II and III, we solve for equilibrium assuming that the risky assets are in zero supply. In Section IV, we explore the implications of our model for liquidity risk. In Section V, we show that our main results extend to positive supply. Section VI concludes, and all proofs are in the Internet Appendix.[7]

I. Model

Time t is continuous and goes from zero to infinity. Uncertainty is described by the N-dimensional Brownian motion B_t. There is a riskless asset whose instantaneous return is constant over time and equal to $r > 0$. This return is exogenous in our model and could be derived from a linear and riskless production technology. There are N risky assets with cash flows

$$dD_t = \bar{D}dt + \sigma^\top dB_t, \tag{1}$$

where $\bar{D}$ is a constant $N \times 1$ vector, σ is a constant and invertible $N \times N$ matrix, and $\top$ denotes the transpose operator. The cash flows (1) are i.i.d. The i.i.d. assumption is for simplicity; we can introduce persistence without significant changes to our analysis. We denote by S_t the $N \times 1$ vector of risky asset prices at time t, and by s the $N \times 1$ vector of asset supplies measured in terms of number of shares. The prices S_t are determined endogenously in equilibrium. We set $\Sigma \equiv \sigma^\top \sigma$.

There are two sets of agents: hedgers and arbitrageurs. Each set forms a continuum of measure one. Hedgers receive a random endowment $u^\top dD_t$ at $t + dt$, where u is a constant $N \times 1$ vector. Because the endowment is correlated with the risky assets' cash flows, it can be hedged by trading in these assets. We consider two specifications for hedgers' preferences. Under both specifications, hedgers' wealth does not affect their risk aversion or demand for insurance.

[6] Longstaff and Wang (2012) show that the hump-shaped pattern extends to expected excess returns. Garleanu and Pedersen (2011) also find hump-shaped volatilities and expected returns. These findings, however, are shown via numerical examples rather than general proofs. See also Liu et al. (2015) for a model in which arbitrageurs render anomalies they discover endogenously more correlated and this endogenous risk is hump-shaped in wealth.

[7] The Internet Appendix is available in the online version of this article on the *Journal of Finance* website.

We intentionally simplify the model in this respect, so that we can focus on the supply of insurance, which is time-varying because of the wealth-dependent risk aversion of arbitrageurs.

- *Specification* 1: *Long-lived hedgers.* Hedgers maximize negative-exponential utility over intertemporal consumption,

$$-\,\mathrm{E}_t\left(\int_t^{\infty} e^{-\frac{\alpha}{r}\bar{c}_{t'}}e^{-\bar{\rho}(t'-t)}dt'\right), \tag{2}$$

  where $\bar{c}_{t'}$ is consumption at $t' \geq t$, $\frac{\alpha}{r}$ is the coefficient of absolute risk aversion, and $\bar{\rho}$ is the subjective discount rate.
- *Specification* 2: *Short-lived hedgers.* Hedgers maximize a mean-variance objective over instantaneous changes in wealth,

$$\mathrm{E}_t(dv_t) - \frac{\alpha}{2}\mathrm{Var}_t(dv_t), \tag{3}$$

  where dv_t is the change in wealth between t and $t + dt$, and α is a risk -aversion coefficient.

The risk-aversion coefficient $\frac{\alpha}{r}$ under Specification 1 is over consumption, and it yields a risk-aversion coefficient α over wealth that is the same as under Specification 2. The interpretation of hedgers under Specification 1 is straightforward: they are infinitely lived agents. Under Specification 2, hedgers can instead be interpreted as overlapping generations living over infinitesimal periods. The generation born at time t is endowed with initial wealth $\bar{v}$ and receives the additional endowment $u^{\top}dD_t$ at $t + dt$. It consumes all of its wealth at $t + dt$ and dies. If preferences over consumption are described by the Von Neumann-Morgenstern utility U, this yields the objective (3) with the risk -aversion coefficient $\alpha = -\frac{U''(\bar{v})}{U'(\bar{v})}$.

Under the overlapping generations interpretation, Specification 2 introduces the friction that future generations of hedgers cannot trade with the current generation. Markets are therefore incomplete, although they are complete for the current generation of hedgers and for arbitrageurs because the number N of risky assets is equal to the number of Brownian motions.[8] Under Specification 1, markets are complete for all hedgers and arbitrageurs. While generating a form of incompleteness, Specification 2 has the advantage of being more tractable. We refer to the hedgers described by Specification 1 as long-lived hedgers and to those described by Specification 2 as short-lived hedgers.

Arbitrageurs maximize power utility over intertemporal consumption. When the coefficient of relative risk aversion γ is different from one, arbitrageurs' objective at time t is

$$\mathrm{E}_t\left(\int_t^{\infty} \frac{c_{t'}^{1-\gamma}}{1-\gamma}e^{-\rho(t'-t)}dt'\right), \tag{4}$$

[8] Market incompleteness prevents future generations of hedgers from hedging against the risk that, when they are born, arbitrageur wealth is low and the cost of insurance is high.

where $c_{t'}$ is consumption at $t' \geq t$ and ρ is the subjective discount rate. When $\gamma = 1$, the objective becomes

$$\mathrm{E}_t\left(\int_t^{\infty} \log(c_{t'})e^{-\rho(t'-t)}dt'\right). \tag{5}$$

We assume that $\rho > r$. As we explain in Section II.C, this assumption ensures that arbitrageurs do not accumulate infinite wealth over time, in which case wealth effects become irrelevant.

In Sections II and III, we solve for equilibrium assuming that the risky assets are in zero supply ($s = 0$). With zero supply, the payoffs that hedgers derive at $t + dt$ from their risky positions at t are the opposite of those derived by arbitrageurs. The same payoffs can be derived through suitable positions in "short-maturity" assets that pay the same cash flows as the "long-maturity" assets at $t + dt$ and zero thereafter, and are in zero supply. This is because the diffusion matrix σ of cash flows has full rank N. Therefore, an equilibrium with long-maturity assets in zero supply yields the same risk-sharing, market prices of risk, and wealth dynamics as one with short-maturity assets in zero supply. Because price dynamics with short-maturity assets are simpler than with long-maturity assets, we derive the latter equilibrium first, in Section II. In Section III, we confirm that the two equilibria are equivalent and we derive the long-maturity assets' prices, expected returns, volatilities, and correlations. Although supply is zero, there is aggregate risk because of the hedgers' endowment, and risk premia are nonzero and time varying.

We allow supply to be positive in Section V. We show that when hedgers are long-lived, risk-sharing and asset prices are the same as in the zero-supply equilibrium derived in Sections II and III, provided that we replace u by $s + u$. That is, only the aggregate of the supply u coming from hedgers and the supply s coming from issuers matters, not the relative composition. When hedgers are short-lived, this equivalence does not hold. However, key aspects of asset price behavior derived in Sections II and III generalize.

For zero supply, our model can represent futures markets, with the assets being futures contracts and the arbitrageurs being the speculators. It can also represent the market for insurance against aggregate risks, for example, weather or earthquakes, with the assets being insurance contracts and the arbitrageurs being the insurers. For positive supply, our model can represent stock or bond markets, with the arbitrageurs being hedge funds or other agents absorbing demand or supply imbalances.

II. Equilibrium with Short-Maturity Assets

A new set of N short-maturity assets can be traded at each time t. The assets available at t pay dD_t at $t + dt$ and zero thereafter. We denote by $\pi_t dt$ the $N \times 1$ vector of prices at which the assets trade at t, and by $dR_t \equiv dD_t - \pi_t dt$ the $N \times 1$ vector of returns that the assets earn between t and $t + dt$. Equation (1) implies that the instantaneous expected returns of the short-maturity assets are

$$\frac{\mathrm{E}_t(dR_t)}{dt} = \bar{D} - \pi_t, \tag{6}$$

and the instantaneous covariance matrix of returns is

$$\frac{\mathrm{Var}_t(dR_t)}{dt} = \frac{\mathrm{E}_t(dR_t dR_t^\top)}{dt} = \sigma^\top \sigma = \Sigma. \tag{7}$$

Note that dR_t is also a return in excess of the riskless asset since investing $\pi_t dt$ in the riskless asset yields return $r\pi_t(dt)^2$, which is negligible relative to dR_t. Note also that dR_t is a return per share rather than per dollar invested: computing the return per dollar would require dividing dR_t by the price $\pi_t dt$. When using dollar rather than share returns in the rest of this paper, we will be mentioning that explicitly.

A. Optimization

Consider first the optimization problem of a long-lived hedger. The hedger's budget constraint is

$$dv_t = rv_t dt + x_t^\top (dD_t - \pi_t dt) + u^\top dD_t - \bar{c}_t dt, \tag{8}$$

where x_t is the hedger's position in the risky assets at time t and v_t is the hedger's wealth. The first term in the right-hand side of (8) is the return from investing in the riskless asset, the second term is the return from investing in the risky assets, the third term is the endowment (with u characterizing the endowment's sensitivity to asset returns), and the fourth term is consumption. We solve the hedger's optimization problem using dynamic programming and conjecture the value function

$$V(v_t, w_t) = -e^{-\left[\alpha v_t + F(w_t)\right]}, \tag{9}$$

where $F(w_t)$ is a scalar function of w_t. The hedger's value function over wealth has the same negative-exponential form as the utility function over consumption, with the risk-aversion coefficient being α rather than $\frac{\alpha}{r}$. In addition, the value function depends on the wealth of arbitrageurs since the latter affects asset prices π_t. Arbitrageur wealth is the only state variable in our model.

PROPOSITION 1: *Given the value function (9), the optimal policy of a long-lived hedger at time t is to consume*

$$\bar{c}_t = rv_t + \frac{r}{\alpha}\left[F(w_t) - \log(r)\right] \tag{10}$$

and hold a position

$$x_t = \frac{\Sigma^{-1}(\bar{D} - \pi_t)}{\alpha} - u - \frac{F'(w_t) y_t}{\alpha} \tag{11}$$

in the risky assets.

The hedger's optimal demand for the risky assets consists of three components, which correspond to the three terms in the right-hand side of (11). The first term is a standard mean-variance demand. It consists of an investment in the tangent portfolio, scaled by the hedger's risk-aversion coefficient α. The tangent portfolio is the inverse of the covariance matrix Σ of asset returns times the vector $\bar{D} - \pi_t$ of expected returns. The second term is a demand to hedge endowment risk. It consists of a short position in the portfolio u, which characterizes the sensitivity of hedgers' endowment to asset returns. Selling short an asset n for which $u_n > 0$ yields a high payoff when dD_{nt} is low, which is when the endowment is also low. The third term is an intertemporal hedging demand as in Merton (1971). Changes in arbitrageur wealth, the only state variable in our model, affect the terms at which hedgers can obtain insurance and hence must be hedged against. Intertemporal hedging is accomplished by holding a portfolio with weights proportional to the sensitivity of arbitrageur wealth to asset returns. That sensitivity is simply the portfolio y_t of arbitrageurs. Hence, the intertemporal hedging demand is a scaled version of y_t, as the third term in (11) confirms.

When the hedger is short-lived, the budget constraint (8) does not include consumption, and the hedger's optimal demand for the risky assets does not include the intertemporal hedging component. The other two terms in (11), however, remain the same. Hence, the case of a short-lived hedger can be nested into that of a long-lived hedger by setting the function $F(w_t)$ to zero.

PROPOSITION 2: *The optimal policy of a short-lived hedger at time t is to hold a position*

$$x_t = \frac{\Sigma^{-1}(\bar{D} - \pi_t)}{\alpha} - u \tag{12}$$

in the risky assets.

Consider next the optimization problem of an arbitrageur. The arbitrageur's budget constraint is

$$dw_t = rw_t dt + y_t^\top (dD_t - \pi_t dt) - c_t dt, \tag{13}$$

where y_t is the arbitrageur's position in the risky assets at time t and w_t is the arbitrageur's wealth. The first term in the right-hand side of (13) is the return from investing in the riskless asset, the second term is the return from investing in the risky assets, and the third term is consumption. The arbitrageur's value function depends not only on his own wealth w_t, but also on the aggregate wealth of all arbitrageurs since the latter affects asset prices π_t. In equilibrium, own wealth and aggregate wealth coincide because all arbitrageurs are symmetric and in measure one. For the purposes of optimization, however, we need to make the distinction. We reserve the notation w_t for aggregate wealth and denote own wealth by $\hat{w}_t$. We likewise use (c_t, y_t) for aggregate consumption and the aggregate position in the assets, and denote own consumption and position by $(\hat{c}_t, \hat{y}_t)$. We conjecture the value function

$$V(\hat{w}_t, w_t) = q(w_t)\frac{\hat{w}_t^{1-\gamma}}{1-\gamma} \tag{14}$$

for $\gamma \neq 1$, and

$$V(\hat{w}_t, w_t) = \frac{1}{\rho}\log(\hat{w}_t) + q_1(w_t) \tag{15}$$

for $\gamma = 1$, where $q(w_t)$ and $q_1(w_t)$ are scalar functions of w_t. We set $q(w_t) = \frac{1}{\rho}$ for $\gamma = 1$.

PROPOSITION 3: *Given the value function (14) and (15), the optimal policy of an arbitrageur at time t is to consume*

$$\hat{c}_t = q(w_t)^{-\frac{1}{\gamma}}\hat{w}_t \tag{16}$$

and hold a position

$$\hat{y}_t = \frac{\Sigma^{-1}(\bar{D} - \pi_t)}{\frac{\gamma}{\hat{w}_t}} + \frac{q'(w_t)y_t}{\frac{\gamma}{\hat{w}_t}q(w_t)} \tag{17}$$

in the risky assets.

The arbitrageur's optimal consumption is proportional to his wealth $\hat{w}_t$, with the proportionality coefficient $q(w_t)^{-\frac{1}{\gamma}}$ being a function of aggregate arbitrageur wealth w_t. The arbitrageur's optimal demand for the risky assets consists of two components, which correspond to the two terms in the right-hand side of (17) and are analogous to those for hedgers. The first term consists of an investment in the tangent portfolio, scaled by the arbitrageur's coefficient of absolute risk aversion $\frac{\gamma}{\hat{w}_t}$. The second term is an intertemporal hedging demand. The arbitrageur hedges against changes in aggregate arbitrageur wealth since these affect asset prices. Intertemporal hedging is accomplished by holding a portfolio with weights proportional to the sensitivity of aggregate arbitrageur wealth to asset returns. Since this sensitivity is the aggregate portfolio y_t of arbitrageurs, the intertemporal hedging demand is a scaled version of y_t.

B. Equilibrium Characterization and Existence

Since in equilibrium all arbitrageurs are symmetric and in measure one, their aggregate position coincides with each arbitrageur's position and the same is true for wealth. Setting $\hat{y}_t = y_t$ and $\hat{w}_t = w_t$ in (17), we find that the aggregate position of arbitrageurs is

$$y_t = \frac{\Sigma^{-1}(\bar{D} - \pi_t)}{A(w_t)}, \tag{18}$$

where

$$A(w_t) \equiv \frac{\gamma}{w_t} - \frac{q'(w_t)}{q(w_t)}. \tag{19}$$

Arbitrageurs' investment in the tangent portfolio is scaled by $A(w_t)$, which can be interpreted as a coefficient of *dynamic risk aversion*. It is the sum of the static coefficient of absolute risk aversion $\frac{\gamma}{w_t}$ and the term $-\frac{q'(w_t)}{q(w_t)}$, which corresponds to the intertemporal hedging demand and hence reflects dynamic considerations. Suppose, for example, that $q(w_t)$ is decreasing, a property that holds for $\gamma < 1$, as we show in Theorem 1. Equation (17) then implies that the intertemporal hedging demand is negative and lowers the arbitrageurs' position. The negative hedging demand is reflected in (18) through a larger coefficient of dynamic risk aversion $A(w_t)$. In Section II.C.2, we provide economic intuition for the sign of $q'(w_t)$ and of the intertemporal hedging demand.

A similar calculation can be made for hedgers using the market-clearing equation

$$x_t + y_t = 0. \tag{20}$$

Setting $y_t = -x_t$ in (11), we find that the aggregate position of hedgers is

$$x_t = \frac{\Sigma^{-1}(\bar{D} - \pi_t) - \alpha u}{\alpha - F'(w_t)}. \tag{21}$$

Substituting (18) and (21) into (20), we find that asset prices π_t are

$$\pi_t = \bar{D} - \frac{\alpha A(w_t)}{\alpha + A(w_t) - F'(w_t)}\Sigma u. \tag{22}$$

Substituting (22) back into (18), we find that the arbitrageurs' position in the risky assets in equilibrium is

$$y_t = \frac{\alpha}{\alpha + A(w_t) - F'(w_t)}u. \tag{23}$$

Arbitrageurs buy a fraction of the portfolio u, which is the portfolio that hedgers want to sell to hedge their endowment risk. They buy a larger fraction of u, hence supplying more insurance to hedgers, when their coefficient of dynamic risk aversion $A(w_t)$ is smaller and hedgers' risk-aversion coefficient α is larger. The degree of insurance supplied by arbitrageurs also depends on hedgers' intertemporal hedging demand, as reflected through the function $F'(w_t)$. When hedgers are short-lived, $F(w_t) = 0$ and hence $F'(w_t) = 0$. When instead they are long-lived, our numerical solutions indicate that $F'(w_t) > 0$. Therefore, hedgers' intertemporal hedging motive makes them demand more insurance from arbitrageurs. The intuition is that hedgers seek to hedge against the possibility that arbitrageurs become poorer in the future because in that event insurance will be supplied to them at worse terms. Arbitrageurs become poorer when the assets they have bought from hedgers underperform. An individual hedger can hedge against that possibility by selling an even larger amount of those assets to arbitrageurs. As a consequence, the intertemporal hedging demand makes hedgers collectively demand more insurance from arbitrageurs.

Equation (22) implies that expected asset returns $\bar{D} - \pi_t$ are proportional to the covariance with the portfolio u, which is the single priced risk factor in our model. The risk premium $\frac{\alpha A(w_t)}{\alpha + A(w_t) - F'(w_t)}$ of that factor depends on arbitrageur wealth, and hence is time-varying. The arbitrageurs' Sharpe ratio, defined as the expected return of their portfolio divided by the portfolio's standard deviation, also depends on their wealth. Using (22) and (23), we find that the Sharpe ratio is

$$SR_t \equiv \frac{y_t^\top(\bar{D} - \pi_t)}{\sqrt{y_t^\top \Sigma y_t}} = \frac{\alpha A(w_t)}{\alpha + A(w_t) - F'(w_t)} \sqrt{u^\top \Sigma u}. \tag{24}$$

Substituting hedgers' optimal policy from Proposition 1 into their Bellman equation (equation (IA.6), derived in the proof of Proposition 2 in the Internet Appendix), and using the dynamics that arbitrageur wealth follows in equilibrium, we can derive an ODE for the function $F(w_t)$. Following the same procedure for arbitrageurs, we can derive an ODE for the function $q(w_t)$. To state these ODEs and subsequent results, we define the parameter $z > 0$ by

$$z \equiv \frac{\alpha^2 u^\top \Sigma u}{2(\rho - r)}. \tag{25}$$

This parameter is larger when hedgers' risk-aversion coefficient α or endowment variance $u^\top \Sigma u$ is larger, when arbitrageurs' subjective discount rate ρ is smaller, or when the riskless rate r is larger.

PROPOSITION 4: *In equilibrium, the function $q(w_t)$ solves the ODE*

$$1 = \frac{q(w_t)^{-\frac{1}{\gamma}} - r}{\rho - r} A(w_t) w_t - \frac{z\big(A'(w_t) + A(w_t)^2\big)}{[\alpha + A(w_t) - F'(w_t)]^2}. \tag{26}$$

The function $F(w_t)$ is equal to zero when hedgers are short-lived, and solves the ODE

$$1 = \frac{rF(w_t) - r\log(r) - \alpha u^\top \bar{D} + \rho - \bar{\rho}}{\rho - r} + \frac{q(w_t)^{-\frac{1}{\gamma}} - r}{\rho - r} F'(w_t) w_t - \frac{z\big\{F''(w_t) - A(w_t)\big[2\alpha + A(w_t) - 2F'(w_t)\big]\big\}}{\big[\alpha + A(w_t) - F'(w_t)\big]^2} \tag{27}$$

when they are long-lived.

Solving for equilibrium when hedgers are short-lived amounts to solving the ODE (26) with the function $F(w_t)$ set to zero. That ODE involves the functions $q(w_t)$ and $A(w_t)$. However, since $A(w_t)$ depends on $q(w_t)$ and $q'(w_t)$, as described in (19), (26) can be written as a second-order ODE in the single function $q(w_t)$. That ODE involves nonlinear terms in both $q(w_t)$ and $q'(w_t)$. Solving for equilibrium when hedgers are long-lived amounts to solving the ODEs (26) and (27).

These can be written as a system of second-order nonlinear ODEs in $q(w_t)$ and $F(w_t)$. We derive boundary conditions for the two ODEs through a small set of economic properties that we assume should hold when w_t goes to zero and to infinity. We next state and motivate these properties.

When w_t goes to zero and to infinity, expected asset returns should converge to finite limits: if the limits were infinite, insurance would become infinitely costly and hedgers would not only refrain from buying it but would also be willing to supply it. Since expected returns converge to finite limits and hedgers have negative exponential utility, their value function should also converge to finite limits, holding their wealth v_t constant. Equation (9) then implies that $F(w_t)$ should converge to finite limits when w_t goes to zero and to infinity.

That expected returns converge to finite limits when w_t goes to zero and to infinity does not imply that an arbitrageur's value function should do the same, holding the arbitrageur's own wealth $\hat{w}_t$ constant. Indeed, because arbitrageurs have power utility, their value function can become infinite when expected returns are large enough.[9] For w_t going to infinity, however, the finite limit of dollar expected returns is the riskless rate r, as arbitrageurs eliminate all risk premia. Moreover, because the arbitrageurs' subjective discount rate ρ exceeds r, their value function does converge to a finite limit, and so does $q(w_t)$.[10] The limit of $q(w_t)$ must further be positive: since $q(w_t)$ is the marginal utility of wealth of an arbitrageur with wealth $\hat{w}_t = 1$, and the arbitrageur can always invest in the riskless asset, $q(w_t)$ must exceed a positive bound. This is the boundary condition for $q(w_t)$ at infinity. The boundary condition at zero is through $A(w_t)$. Since arbitrageurs have power utility and risk premia are finite and nonzero, the arbitrageurs' position in the risky assets as a fraction of their wealth should converge to a finite nonzero limit (while converging to zero in absolute terms). Equation (18) then implies that $A(w_t)w_t$ should converge to a positive limit. Note that all of the boundary conditions that we impose concern the existence of finite limits rather than the limits' exact values.

Theorem 1 provides a comprehensive analysis of the equilibrium when hedgers are short-lived, and a partial analysis when they are long-lived. In the short-lived case, we show that a solution to the ODE (26) with $F(w_t) = 0$ and the boundary conditions on $q(w_t)$ and $A(w_t)$ exists. We also characterize monotonicity properties of the solution and show that the limits at zero and infinity are uniquely determined. In the long-lived case, we have not been able to show existence and monotonicity properties, but we do show that the limits at zero and infinity are the same as in the short-lived case.

[9] The result that the value function is finite under negative exponential utility but can become infinite under power utility when return distributions are time-invariant and expected returns finite, follows from Merton (1971).

[10] The result that the value function under power utility is finite when dollar return distributions are time-invariant, expected returns are equal to the riskless rate r, and the subjective discount rate exceeds r, follows from Merton (1971).

THEOREM 1: *When hedgers are short-lived, a solution to the ODE (26) with $F(w_t) = 0$ and positive limits of $A(w_t)w_t$ at zero and $q(w_t)$ at infinity exists. The solution has the following properties:*

- *The function $A(w_t)$ is decreasing.*
- $\lim_{w_t\to\infty} A(w_t)w_t = \gamma$ *and* $\lim_{w_t\to\infty} q(w_t) = \frac{1}{(r+\frac{\rho-r}{\gamma})\gamma}$.
- *If $\gamma < 1$, then $\frac{1}{w_t} > A(w_t) > \frac{\gamma}{w_t}$ and $q(w_t)$ is decreasing.*
- *If $\gamma > 1$, then $\frac{1}{w_t} < A(w_t) < \frac{\gamma}{w_t}$ and $q(w_t)$ is increasing.*
- *If $\gamma < K$, where $K < 1$ is the unique positive solution of*

$$G(K) \equiv 1 - K - \frac{K}{z}\left(\frac{rK}{\rho - r} + 1\right) = 0, \tag{28}$$

  then $\lim_{w_t\to 0} A(w_t)w_t = K$ *and* $\lim_{w_t\to 0} q(w_t) = \infty$.
- *If $\gamma > K$, then* $\lim_{w_t\to 0} A(w_t)w_t = \gamma$ *and* $\lim_{w_t\to 0} q(w_t) \in (0, \infty)$.

Suppose next that hedgers are long-lived, and that a solution to the system of ODEs (26) and (27) with positive limits of $A(w_t)w_t$ at zero and $q(w_t)$ at infinity, and finite limits of $F(w_t)$ at zero and infinity, exists. Suppose additionally that $F'(w_t)w_t$ and $F''(w_t)w_t^2$ have (finite or infinite) limits at zero and infinity. The solution has the following properties:

- *The limits of $A(w_t)w_t$ and $q(w_t)$ at zero and infinity are as in the case of short-lived hedgers.*
- $\lim_{w_t\to 0} F(w_t) = \log(r) + \frac{\alpha u^\top \bar{D} + \bar{\rho} - r - z(\rho - r)}{r}$ *and* $\lim_{w_t\to\infty} F(w_t) = \log(r) + \frac{\alpha u^\top \bar{D} + \bar{\rho} - r}{r}$.
- *The limits of $[A(w_t) - F'(w_t)]w_t$ at zero and infinity are the same as those of $A(w_t)w_t$.*

The basic idea of our existence proof is to start with a compact interval $[\epsilon, M]$ and show that there exists a unique solution to the ODE (26) with the limits of $A(w_t)w_t$ at zero and infinity imposed as boundary conditions at ϵ and M, respectively. We next show that when ϵ converges to zero and M to infinity, that solution converges to a solution over $(0, \infty)$. Our convergence proof uses the monotonicity of the solution with respect to ϵ and M, which in turn follows from a monotonicity property of solutions with respect to initial conditions. Our construction yields a unique solution over $(0, \infty)$, although it does not rule out the possibility that other solutions (constructed differently) may exist. Uniqueness of our constructed solution allows us to examine how that solution moves in response to exogenous parameters: we perform the comparative statics on the solution over $[\epsilon, M]$ and take the limit when ϵ converges to zero and M to infinity. Our existence proof concerns only short-lived hedgers; when stating results on long-lived hedgers in the rest of this paper, we assume that a solution to the system of ODEs (26) and (27) as described in Theorem 1 exists.

In Section II.C, we derive economic implications of the results shown in Theorem 1, as well as some additional properties. We examine how positions and returns depend on arbitrageur wealth (Section II.C.1), how dynamic risk aversion differs from its static counterpart (Section II.C.2), and the long-run dynamics of arbitrageur wealth (Section II.C.3).

C. Equilibrium Properties

C.1. Wealth Effects

Theorem 1 shows that when hedgers are short-lived, an increase in the wealth w_t of arbitrageurs causes arbitrageurs' dynamic risk aversion $A(w_t)$ to decline. A decline in $A(w_t)$ results in more insurance supplied to hedgers: arbitrageurs' positions become more positive for positive elements of u, which correspond to assets that hedgers want to sell, and more negative for negative elements of u, which correspond to assets that hedgers want to buy. Expected asset returns, which reflect the cost of the insurance, become less positive for positive elements of u and less negative for negative elements of u, and hence smaller in absolute value. The same is true for the market prices of the Brownian risks, that is, the expected returns per unit of risk exposure, and for arbitrageurs' Sharpe ratio.

COROLLARY 1: *When hedgers are short-lived, an increase in arbitrageur wealth w_t:*

(1) Raises the position of arbitrageurs in each asset in absolute value.
(2) Lowers the expected return of each asset in absolute value.
(3) Lowers the market price of each Brownian risk in absolute value.
(4) Lowers arbitrageurs' Sharpe ratio.

The results of Corollary 1 appear to be consistent with the empirical findings of Kang, Rouwenhorst, and Tang (2016). That paper finds that commodity futures in which hedgers held long positions on average over the previous year earn negative expected returns. At the same time, commodity futures in which hedgers made a net purchase over the previous week (increasing their long position or reducing their short position) earn positive expected returns. In our model, hedgers hold long positions in assets n with $u_n < 0$, and these assets earn negative expected returns. At the same time, Corollary 1 implies that hedgers increase these long positions following an increase in arbitrageur wealth w_t, and the negative expected returns become less negative. Moreover, hedgers reduce their short positions in assets n with $u_n > 0$ following a decline in w_t, and the positive expected returns of these assets become more positive. Hence, net purchases of hedgers following shocks to w_t may on average be associated with positive expected returns. Hedging demand u and arbitrageur wealth w_t may thus generate relationships between positions and expected returns that are opposite and operate at different frequencies—low frequency for u and high frequency for w_t—consistent with the empirical evidence. In a similar spirit, Cheng and Xiong (2014) find that changes in the hedging demand of commodity hedgers do not appear to be the main driver of changes to their positions at high frequency. High-frequency changes may instead be driven by arbitrageur wealth.[11]

When hedgers are long-lived, changes in the wealth of arbitrageurs affect not only their dynamic risk aversion $A(w_t)$ but also hedgers' intertemporal hedging demand, the strength of which is captured by the function $F'(w_t) > 0$. Recall that long-lived hedgers seek to hedge against the possibility that arbitrageurs will become poorer in the future because in that event insurance will be supplied to them at worse terms. Moreover, because of that intertemporal hedging motive, hedgers demand more insurance. Our numerical solutions indicate that when arbitrageur wealth w_t increases, both $A(w_t)$ and $F'(w_t)$ decrease. In the case of $F'(w_t)$, this is because the terms of insurance become less sensitive to w_t for larger values of w_t.

The interplay between increased supply of insurance (lower $A(w_t)$) and declining demand for it (lower $F'(w_t)$) can give rise to nonmonotonic patterns. Consider, for example, arbitrageur positions, which increase in absolute value when $A(w_t) - F'(w_t)$ declines, as shown in (23). When the hedgers' risk-aversion coefficient α is high, positions are hump-shaped in w_t: they increase in absolute value for small values of w_t as $A(w_t)$ declines, and decrease for larger values of w_t as $F'(w_t)$ declines. When instead α is low, the variation in $F'(w_t)$ is dominated by that in $A(w_t)$, and positions increase in absolute value for all values of w_t, as in Corollary 1.

Figure 1 illustrates the behavior of arbitrageurs' dynamic risk aversion and positions. The plots in the left column show dynamic risk aversion $A(w_t)$. The plots in the middle column show the position y_{nt} in asset n, expressed as a fraction of the position u_n that hedgers want to hedge. All of these plots concern the case in which hedgers are long-lived. The plots in the right column express positions relative to the case in which hedgers are short-lived. The blue solid line represents the baseline case, which is the same in all plots. The plots in the top row show that positions increase with w_t in the baseline case but become hump-shaped when hedgers' risk-aversion coefficient α increases. For larger values of α, arbitrageurs can overinsure hedgers, buying the full position that they want to hedge and holding an additional long position. Overinsurance introduces a large deviation relative to the case of short-lived hedgers, as the top-right figure shows. The plots in the bottom row show that arbitrageurs provide less insurance to hedgers when their risk-aversion coefficient γ increases.

C.2. Dynamic Risk Aversion

Recall from (19) that the dynamic risk aversion $A(w_t)$ of arbitrageurs is the sum of the static coefficient of absolute risk aversion $\frac{\gamma}{w_t}$ and the term

[11] Cheng and Xiong (2015) find that commodity hedgers scale down their positions when aggregate volatility (as measured by VIX) increases. This is suggestive of hedgers' positions changing in the short term for reasons unrelated to their hedging demand (which should increase when volatility increases): during volatile times, arbitrageurs may become more constrained or risk-averse, an effect derived in Corollary 2 in the case $\gamma < 1$.

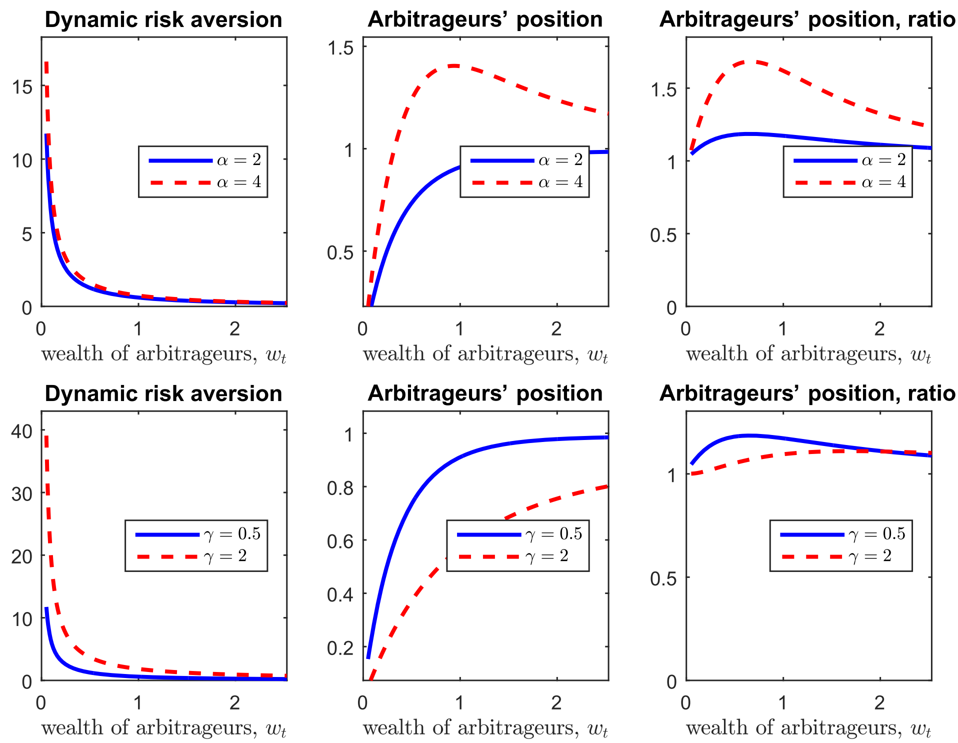

Figure 1. Arbitrageur dynamic risk aversion $A(w_t)$ and positions y_t as a function of arbitrageur wealth w_t. The plots in the left column show $A(w_t)$. The plots in the middle column show y_{nt} for asset n, expressed as a fraction of the position u_n that hedgers want to hedge. All of these plots concern the case in which hedgers are long-lived. The plots in the right column express positions relative to the case in which hedgers are short-lived. All plots assume two symmetric and uncorrelated risky assets, and set the annualized standard deviation of hedgers' endowment $\sqrt{u^\top \Sigma u}$ to 15%, arbitrageurs' subjective discount rate ρ to 4%, and the riskless rate r to 2%. In the baseline case, represented by the blue solid line in all plots, hedgers' risk-aversion coefficient α is set to 2 and arbitrageurs' relative risk-aversion coefficient γ to 0.5. (By normalizing hedgers' wealth to one via a choice of numeraire, we can interpret α as their coefficient of relative risk aversion and $\sqrt{u^\top \Sigma u}$ as the annualized standard deviation of their endowment as a fraction of their wealth.) Under these choices, the parameters z and K are 2.25 and 0.59, respectively. The plots in the top row examine the effect of raising α to 4, and the plots in the bottom row examine the effect of raising γ to 2. (Color figure can be viewed at wileyonlinelibrary.com)

$-\frac{q'(w_t)}{q(w_t)}$, which corresponds to the intertemporal hedging demand. In this section, we draw on the results of Theorem 1 and examine how dynamic and static risk aversion differ, or equivalently, the properties of intertemporal hedging demand. We start with the case of short-lived hedgers. At the end of this section, we examine how the results carry through to long-lived hedgers.

For $\gamma < 1$, the intertemporal hedging demand raises $A(w_t)$ above $\frac{\gamma}{w_t}$, while the opposite is true for $\gamma > 1$. The difference between the two cases lies in the behavior of the marginal utility of arbitrageur wealth, which is $q(w_t)$ for an arbitrageur with wealth $\hat{w}_t = 1$. For $\gamma < 1$, $q(w_t)$ is decreasing in w_t, meaning that the arbitrageur has higher marginal utility in states in which aggregate arbitrageur wealth w_t is low. This is because in low-w_t states, expected returns are high and hence the arbitrageur earns a high return on wealth. By seeking to preserve wealth in those states, so as to earn the high return by investing it, the arbitrageur scales back his positions, behaving more risk-aversely than in the absence of intertemporal hedging. For $\gamma > 1$, $q(w_t)$ is instead increasing in w_t, meaning that the arbitrageur has lower marginal utility in low-w_t states. This is because the high return on wealth in those states is associated with high utility of future consumption. The return is therefore discounted by low marginal utility, with the latter effect dominating for $\gamma > 1$.

When utility is logarithmic ($\gamma = 1$), intertemporal hedging demand is zero, and hence $A(w_t) = \frac{1}{w_t}$. Although intertemporal hedging demand raises $A(w_t)$ for $\gamma < 1$, $A(w_t)$ remains smaller than for $\gamma = 1$. Conversely, although intertemporal hedging demand lowers $A(w_t)$ for $\gamma > 1$, $A(w_t)$ remains larger than for $\gamma = 1$. These comparisons do not imply, however, that the relative contribution of intertemporal hedging demand to $A(w_t)$ is small—we next point out that it is large in some cases.

When w_t goes to infinity, $A(w_t) \approx \frac{\gamma}{w_t}$. Hence, $A(w_t)$ is driven purely by the static component, and the relative contribution of the intertemporal hedging demand converges to zero. The same result holds when w_t goes to zero if γ exceeds a threshold $K \in (0, 1)$. If instead $\gamma < K$, then $A(w_t) \approx \frac{K}{w_t}$. Hence, the contribution of the intertemporal hedging demand to $A(w_t)$ is $\frac{K-\gamma}{w_t}$ in absolute terms and $\frac{K-\gamma}{K}$ in relative terms. If, in particular, γ is close to zero, $A(w_t)$ is driven almost entirely by the intertemporal hedging demand. The difference between the cases $\gamma > K$ and $\gamma < K$ lies in the behavior of the marginal utility $q(w_t)$ when w_t goes to zero. If $\gamma > K$, then $q(w_t)$ converges to a finite limit and hence the ratio of $-\frac{q'(w_t)}{q(w_t)}$ to $\frac{\gamma}{w_t}$ must converge to zero. If instead $\gamma < K$, then $q(w_t)$ converges to infinity and hence $-\frac{q'(w_t)}{q(w_t)}$ can be larger than $\frac{\gamma}{w_t}$.[12]

Equation (28) implies that the threshold K increases with hedgers' risk-aversion coefficient α and endowment variance $u^\top \Sigma u$. Hence, when hedgers are more risk-averse or their endowment is more volatile, arbitrageurs with $\gamma < K$ also become more risk-averse for w_t close to zero. Theorem 2 generalizes these results to any $\gamma < 1$ and w_t, and shows that the opposite results hold for $\gamma > 1$.

THEOREM 2: *Suppose that hedgers are short-lived. For any given level of arbitrageur wealth w_t, the following comparative statics hold:*

(1) An increase in hedgers' risk-aversion coefficient α raises arbitrageurs' dynamic risk aversion $A(w_t)$ if $\gamma < 1$, and lowers it if $\gamma > 1$.

[12] The region $\gamma < K$ in which $q(w_t)$ converges to infinity when w_t goes to zero is a subset of the region $\gamma < 1$ in which $q(w_t)$ is decreasing. The former region is also the one in which the value function of an arbitrageur facing the time-invariant expected returns that arise in equilibrium when $w_t = 0$ (no arbitrageurs) is infinite.

(2) An increase in hedgers' endowment variance $u^\top \Sigma u$ raises arbitrageurs' dynamic risk aversion $A(w_t)$ if $\gamma < 1$, and lowers it if $\gamma > 1$.

The intuition for Theorem 2 is that for larger values of α and $u^\top \Sigma u$, expected returns are more sensitive to changes in arbitrageur wealth w_t, rising more steeply when w_t declines. Hence, arbitrageurs with $\gamma < 1$ have a marginal utility that also rises more steeply following a decline in w_t. This makes them even more willing to preserve wealth in low-w_t states, and raises their dynamic risk aversion $A(w_t)$. Arbitrageurs with $\gamma > 1$, by contrast, have a marginal utility that drops more steeply following a decline in w_t, and this lowers $A(w_t)$.

The comparative statics of $A(w_t)$ yield comparative statics for arbitrageur positions and Sharpe ratios. Most surprising among these is that arbitrageurs can supply *less* insurance to hedgers when the latter become more risk-averse. Following an increase in α, supplying insurance becomes more profitable for arbitrageurs. If arbitrageur risk aversion included only the static component, which does not depend on α, arbitrageurs would supply more insurance. Because of intertemporal hedging demand, however, the dynamic risk aversion of arbitrageurs with $\gamma < 1$ increases. Moreover, this effect can dominate, inducing arbitrageurs to supply less insurance. Arbitrageurs with $\gamma \geq 1$ instead supply more insurance because their dynamic risk aversion decreases. When $\gamma \leq 1$, arbitrageurs' Sharpe ratio increases because both hedgers and arbitrageurs become more risk-averse. The comparative statics with respect to $u^\top \Sigma u$ are along similar lines.

COROLLARY 2: *Suppose that hedgers are short-lived. For any given level of arbitrageur wealth w_t, the following comparative statics hold:*

(1) An increase in hedgers' risk-aversion coefficient α raises arbitrageurs' position in absolute value when $\gamma \geq 1$ and lowers it when $\gamma < K$, $z < 1$, and w_t is small. It raises arbitrageurs' Sharpe ratio when $\gamma \leq 1$.
(2) An increase in the variance $u^\top \Sigma u$ of hedgers' endowment raises arbitrageurs' position in absolute value when $\gamma > 1$ and lowers it when $\gamma < 1$. It raises arbitrageurs' Sharpe ratio when $\gamma \leq 1$.

When hedgers are long-lived, an increase in α lowers arbitrageurs' position in absolute value when $\gamma < K$, $z < 1$, and w_t is small.

The result that an increase in α can induce arbitrageurs to supply less insurance to hedgers carries through to long-lived hedgers. This is because the asymptotic behavior of $A(w_t)$ for w_t close to zero is the same as with short-lived hedgers. Our numerical solutions indicate that the remaining comparative statics in Theorem 2 and Corollary 2 also extend to long-lived hedgers.

C.3. Stationary Distribution

We next derive the long-run dynamics of arbitrageur wealth.

PROPOSITION 5: *If $z > 1$, then arbitrageur wealth has a long-run stationary distribution with density*

$$d(w_t) = \frac{\frac{[\alpha+A(w_t)-F'(w_t)]^2}{A(w_t)} \exp\left[\int_1^{w_t}\left(A(\hat{w}_t) - \frac{[\alpha+A(\hat{w}_t)-F'(\hat{w}_t)]^2}{zA(\hat{w}_t)}\right)d\hat{w}_t\right]}{\int_0^\infty \frac{[\alpha+A(w_t)-F'(w_t)]^2}{A(w_t)} \exp\left[\int_1^{w_t}\left(A(\hat{w}_t) - \frac{[\alpha+A(\hat{w}_t)-F'(\hat{w}_t)]^2}{zA(\hat{w}_t)}\right)d\hat{w}_t\right]dw_t} \tag{29}$$

over the support $(0, \infty)$. *If $z < 1$, then wealth converges to zero in the long run.*

Arbitrageur wealth has a nondegenerate stationary density if the parameter z defined in (25) exceeds one. That is, hedgers' risk-aversion coefficient α and endowment variance $u^\top \Sigma u$ must be large enough relative to the difference between arbitrageurs' subjective discount rate ρ and the riskless rate r. Note that this result is valid both when hedgers are short-lived and when they are long-lived.

The existence of a nondegenerate stationary density is related to the dynamics of arbitrageur wealth being *self-correcting*: when wealth drops, arbitrageurs' future expected returns increase, causing wealth to grow faster, and vice versa. To explain the relationship between a nondegenerate stationary density and self-correcting wealth dynamics, and why the condition $z > 1$ is required, recall the standard Merton (1971) portfolio optimization problem in which an infinitely lived investor with CRRA coefficient γ can invest in a riskless rate r and in N risky assets whose returns have an expectation given by a vector μ and a covariance given by a matrix Σ. The investor's wealth converges to infinity in the long run when

$$r + \frac{1}{2}\mu^\top \Sigma^{-1} \mu > \rho, \tag{30}$$

that is, the riskless rate plus one-half of the squared Sharpe ratio achieved from investing in the risky assets exceeds the investor's subjective discount rate ρ. When instead (30) holds in the opposite direction, wealth converges to zero. Intuitively, wealth converges to infinity when the investor accumulates wealth at a rate that exceeds sufficiently the rate at which he consumes.

Our model differs from the Merton problem because arbitrageurs' Sharpe ratio is endogenously determined in equilibrium and decreases in their wealth (Corollary 1). Using (24) to substitute for arbitrageurs' Sharpe ratio, we can write (30) as

$$r + \frac{1}{2}\left(\frac{\alpha A(w_t)}{\alpha + A(w_t) - F'(w_t)}\right)^2 u^\top \Sigma u > \rho. \tag{31}$$

Transposing the result from the Merton problem thus suggests that there are three possibilities for the long-run dynamics of arbitrageur wealth. If (31) is satisfied for all values of w_t, then wealth converges to infinity. If (31) is violated for all values of w_t, then wealth converges to zero. If, finally, (31) is violated for large values but is satisfied for small values, convergence does not occur and wealth has a nondegenerate stationary density.

Since Theorem 1 shows that $A(w_t)$ converges to zero when w_t goes to infinity, (31) is satisfied for large values of w_t if $r > \rho$. Our assumption that $\rho > r$ thus implies that (31) is violated for large w_t and rules out the possibility that wealth converges to infinity in the long run. To examine whether (31) is satisfied for small values of w_t, recall from Theorem 1 that $A(w_t)$ is of order $\frac{1}{w_t}$ for w_t close to zero and that $\lim_{w_t \to 0} F'(w_t)w_t = 0$. Hence, (31) is satisfied for small w_t if $r + \frac{\alpha^2 u^\top \Sigma u}{2} > \rho$. This condition is equivalent to $z > 1$, which is exactly what Proposition 5 requires for a nondegenerate stationary density to exist. Proposition 6 computes the density in closed form when hedgers are short-lived and arbitrageurs have logarithmic utility ($\gamma = 1$).

PROPOSITION 6: *Suppose that hedgers are short-lived, arbitrageurs have logarithmic utility ($\gamma = 1$), and $z > 1$. The stationary density $d(w_t)$:*

(1) Is decreasing in w_t if $z < \frac{27}{8}$.
(2) Is bimodal in w_t otherwise: decreasing in w_t for $w_t \in (0, m_1)$, increasing in w_t for $w_t \in (m_1, m_2)$, and again decreasing in w_t for $w_t \in (m_2, \infty)$. The thresholds $m_1 < m_2$ are the two positive roots of

$$(\alpha w_t)^3 + 3(\alpha w_t)^2 + (3 - 2z)\alpha w_t + 1 = 0. \tag{32}$$

(3) Shifts to the right in the monotone likelihood ratio sense when α or $u^\top \Sigma u$ increase.

The stationary density has two possible shapes. When z is not much larger than one, it is decreasing in arbitrageur wealth w_t, and so values of w_t close to zero are more likely than larger values. When instead z is sufficiently larger than one, the stationary density becomes bimodal, with the two maxima being zero and an interior point m_2 of the support. Values of w_t close to the maxima are more likely than intermediate values, meaning that the economy spends more time at these values than in the middle. The intuition is that when hedgers' risk-aversion coefficient α and endowment variance $u^\top \Sigma u$ are large, arbitrageurs earn high expected returns for providing insurance and their wealth grows fast. Therefore, large values of w_t are more likely in steady state than intermediate values. At the same time, while expected returns are highest when wealth is small, wealth grows away from small values slowly in absolute terms. Therefore, small values of w_t are more likely than intermediate values.

Using the long-run stationary distribution, we can perform "unconditional" comparative statics. For example, rather than examining how arbitrageurs' Sharpe ratio depends on α conditionally on w_t, we can examine how it depends on α unconditionally, in expectation over the stationary distribution of w_t. The unconditional comparative statics can differ from the conditional ones. For example, while an increase in α and $u^\top \Sigma u$ raises the conditional Sharpe ratio when $\gamma \leq 1$ (Corollary 2), it can lower its unconditional expectation. Intuitively, for larger values of α and $u^\top \Sigma u$, arbitrageur wealth grows faster, and its stationary density shifts to the right (Proposition 6). Therefore, while the conditional Sharpe ratio increases, its unconditional expectation can decrease because large values of wealth, which yield low Sharpe ratios, become more likely.

III. Equilibrium with Long-Maturity Assets

We conjecture that in equilibrium, the price vector S_t of the long-maturity assets follows the Ito process

$$dS_t = \mu_{St}dt + \sigma_{St}^{\top}dB_t, \tag{33}$$

where μ_{St} is a $N \times 1$ vector and σ_{St} is a $N \times N$ matrix. We denote by $dR_t \equiv dS_t + dD_t - rS_t dt$ the $N \times 1$ vector of returns that the long-maturity assets earn between t and $t + dt$ in excess of the riskless asset. Note that, as in the case of short-maturity assets, dR_t is a return per share rather than per dollar invested. Equations (1) and (33) imply that the instantaneous expected excess returns of the long-maturity assets are

$$\frac{\mathrm{E}_t(dR_t)}{dt} = \mu_{St} + \bar{D} - rS_t, \tag{34}$$

and the instantaneous covariance matrix of returns is

$$\frac{\mathrm{Var}_t(dR_t)}{dt} = (\sigma_{St} + \sigma)^{\top}(\sigma_{St} + \sigma). \tag{35}$$

A. Equivalence with Short-Maturity Assets

With long-maturity assets, the budget constraint (8) of a long-lived hedger becomes

$$dv_t = rv_t dt + X_t^{\top}(dS_t + dD_t - rS_t dt) + u^{\top}dD_t - \bar{c}_t dt, \tag{36}$$

where X_t is the hedger's position in the risky assets at time t. The budget constraint of a short-lived hedger is derived from (36) by excluding consumption. The budget constraint (13) of an arbitrageur becomes

$$dw_t = rw_t dt + Y_t^{\top}(dS_t + dD_t - rS_t dt) - c_t dt, \tag{37}$$

where Y_t is the arbitrageur's position in the risky assets at time t. The market-clearing equation (20) becomes

$$X_t + Y_t = 0. \tag{38}$$

Because the diffusion matrix σ of cash flows has full rank N, an equilibrium with long-maturity assets yields the same risk-sharing, market prices of risk, and wealth dynamics as one with short-maturity assets.

LEMMA 1: *An equilibrium (S_t, X_t, Y_t) with long-maturity assets can be constructed from an equilibrium (π_t, x_t, y_t) with short-maturity assets by:*

(1) Choosing the price process S_t such that

$$\left(\sigma^\top\right)^{-1}(\bar{D} - \pi_t) = \left((\sigma_{St} + \sigma)^\top\right)^{-1}(\mu_{St} + \bar{D} - rS_t). \tag{39}$$

(2) Choosing the asset positions X_t of hedgers and Y_t of arbitrageurs such that

$$\sigma x_t = (\sigma_{St} + \sigma)X_t, \tag{40}$$

$$\sigma y_t = (\sigma_{St} + \sigma)Y_t. \tag{41}$$

In the equilibrium with long-maturity assets, the dynamics of arbitrageur wealth, the exposures of hedgers and arbitrageurs to the Brownian shocks, the market prices of the Brownian risks, and arbitrageurs' Sharpe ratio are the same as in the equilibrium with short-maturity assets.

Equations (40) and (41) construct positions of hedgers and arbitrageurs in the long-maturity assets so that the exposures to the underlying Brownian shocks are the same as with short-maturity assets. Equation (39) constructs a price process such that the market prices of the Brownian risks are also the same. Given this price process, risk exposures are optimal and clear the markets because these properties also hold with short-maturity assets.

B. Asset Prices and Returns

The prices S_t of the long-maturity assets are a function of arbitrageur wealth w_t, which is the only state variable in our model. Using Ito's lemma to compute the drift μ_{St} and diffusion σ_{St} of the price process as a function of the dynamics of w_t, and substituting into (39), we can determine $S(w_t)$ up to an ODE for a scalar function.

PROPOSITION 7: *The prices of the long-maturity assets are given by*

$$S(w_t) = \frac{\bar{D}}{r} + g(w_t)\Sigma u, \tag{42}$$

where the scalar function $g(w_t)$ solves the ODE

$$\begin{aligned}&\frac{\alpha^2 u^\top \Sigma u}{2[\alpha + A(w_t) - F'(w_t)]^2} g''(w_t) + \left(r - q(w_t)^{-\frac{1}{\gamma}}\right) g'(w_t) w_t - r g(w_t)\\ &\quad = \frac{\alpha A(w_t)}{\alpha + A(w_t) - F'(w_t)}.\end{aligned} \tag{43}$$

The assets' expected excess returns are

$$\frac{\mathrm{E}_t(dR_t)}{dt} = \frac{\alpha A(w_t)}{\alpha + A(w_t) - F'(w_t)}\left[u^\top \Sigma u f(w_t) + 1\right]\Sigma u, \tag{44}$$

and the covariance matrix of returns is

$$\frac{\mathrm{Var}_t(dR_t)}{dt} = f(w_t)\left[u^\top \Sigma u f(w_t) + 2\right]\Sigma u u^\top \Sigma + \Sigma, \tag{45}$$

where

$$f(w_t) \equiv \frac{\alpha g'(w_t)}{\alpha + A(w_t) - F'(w_t)}. \tag{46}$$

The ODE (43) is linear in $g(w_t)$. The boundary conditions that we require are that $g(w_t)$ converges to finite limits at zero and infinity. As with the ODEs in Section II, we only assume the existence of finite limits rather than the limits' exact values. Theorem 3 shows that when hedgers are short-lived, a solution $g(w_t)$ to the ODE (43) exists, is negative and increasing in w_t, and converges to $-\frac{\alpha}{r}$ (zero) when w_t goes to zero (infinity). The theorem also shows that when hedgers are long-lived, the limits are as in the short-lived case.

THEOREM 3: *When hedgers are short-lived, a solution to the ODE (43) with finite limits at zero and infinity exists. The solution has the following properties:*

- *The function $g(w_t)$ is negative and increasing.*
- $\lim_{w_t \to 0} g(w_t) = -\frac{\alpha}{r}$ *and* $\lim_{w_t \to \infty} g(w_t) = 0$.

When hedgers are long-lived, the limits of $g(w_t)$ at zero and infinity are the same as when hedgers are short-lived, provided that a solution to the ODE (43) with finite limits exists.

Proposition 7 shows that asset prices (given in (42)) are the sum of two terms. The first term, $\frac{\bar{D}}{r}$, is the present value of the assets' expected cash flows $\bar{D}$, discounted at the riskless rate r. Prices would equal that present value if arbitrageurs had infinite wealth since they would then eliminate all risk premia, rendering the expected dollar returns on all assets equal to r. The second term, $g(w_t)\Sigma u$, reflects the risk premia arising from arbitrageur wealth w_t being finite. Consider an asset n that covaries positively with the portfolio u that hedgers want to sell, that is, $(\Sigma u)_n > 0$. In the absence of arbitrageurs, that asset would trade at a discount relative to $\frac{\bar{D}_n}{r}$. Arbitrageurs cause that discount to decrease, and the more so the wealthier they are. Hence, the asset price $S_n = \frac{\bar{D}_n}{r} + g(w_t)(\Sigma u)_n$ increases in w_t and converges to $\frac{\bar{D}_n}{r}$ when w_t goes to infinity. Theorem 3 shows that $g(w_t)$ is indeed increasing and converges to zero at infinity.

Since changes in arbitrageur wealth w_t affect the prices of long-maturity assets, they also impact the assets' returns. Proposition 7 shows that the covariance matrix of asset returns (given in (45)) is the sum of a "fundamental" component Σ, which is driven purely by shocks to assets' underlying cash flows dD_t, and an "endogenous" component $f(w_t)[f(w_t)u^\top \Sigma u + 2]\Sigma u u^\top \Sigma$, which is introduced because cash flow shocks affect w_t, which affects returns. Endogenous risk does not arise with short-maturity assets since their returns are risky only because of the payoff dD_t, which is not sensitive to changes in w_t.

The endogenous covariance between an asset pair (n, n') depends on whether the corresponding components of the vector Σu have the same or opposite signs. Suppose, for example, that $(\Sigma u)_n > 0$ and $(\Sigma u)_{n'} > 0$, in which case both assets would trade at a discount in the arbitrageurs' absence. An increase in arbitrageur wealth w_t causes the prices of both assets to increase, resulting in positive endogenous covariance. Suppose instead that $(\Sigma u)_n < 0$, in which case demand from hedgers would cause asset n to trade at a premium in the arbitrageurs' absence. Asset n's price would then drop following an increase in w_t, resulting in negative endogenous covariance with asset n'.

Endogenous risk is small at both extremes of the wealth distribution and larger in the middle. When arbitrageur wealth w_t is close to zero, arbitrageurs hold small positions in absolute terms (i.e., not as a fraction of w_t). Therefore, changes in w_t are small and have a small impact on prices. Instead, when w_t is close to infinity, arbitrageurs absorb the entire portfolio u that hedgers want to sell. Changes in w_t are hence larger, but prices are insensitive to those changes. These effects can be seen in the expression (46) for $f(w_t)$: this function is small for small w_t because $A(w_t)$ is large (resulting in small positions by arbitrageurs) and for large w_t because $g'(w_t)$ is small (low price sensitivity to w_t). Since the endogenous variance is larger in the middle than in the extremes, total variance can be hump-shaped in w_t. Total covariance and correlation can be hump-shaped or inverse hump-shaped depending on whether the endogenous covariance is positive or negative, respectively. Proposition 8 confirms the hump shapes (i.e., shows that there is only one hump) in the case in which hedgers are short-lived and arbitrageurs have logarithmic utility.

PROPOSITION 8: *The effects of a change in arbitrageur wealth w_t on the volatility of asset returns and on return covariance and correlation converge to zero when w_t goes to zero and to infinity. When hedgers are short-lived and arbitrageurs have logarithmic utility, an increase in w_t has:*

(1) A hump-shaped effect on the volatility of asset returns. The hump peaks at a value that is common to all assets.

(2) The same hump-shaped effect on the covariance between the returns of assets n and n' if $(\Sigma u)_n(\Sigma u)_{n'} > 0$, and the inverse hump-shaped effect if $(\Sigma u)_n(\Sigma u)_{n'} < 0$.

(3) The same hump-shaped effect on the correlation between the returns of assets n and n' if

$$\frac{(\Sigma u)_n(\Sigma u)_{n'}\Sigma_{nn} - (\Sigma u)_n^2 \Sigma_{nn'}}{f(w_t)\big[f(w_t)u^\top \Sigma u + 2\big](\Sigma u)_n^2 + \Sigma_{nn}}$$

$$+\frac{(\Sigma u)_n(\Sigma u)_{n'}\Sigma_{n'n'}-(\Sigma u)_{n'}^2\Sigma_{nn'}}{f(w_t)\big[f(w_t)u^\top\Sigma u+2\big](\Sigma u)_{n'}^2+\Sigma_{n'n'}}>0, \tag{47}$$

and the inverse hump-shaped effect if (47) holds in the opposite direction.

The effect on correlations is more complicated than that on covariances because it includes the effect on volatilities. Suppose that changes in arbitrageur wealth move the prices of assets n and n' in the same direction and hence have a hump-shaped effect on their covariance. However, because the effect on volatilities, which are in the denominator, is also hump-shaped, the overall effect on the correlation can be inverse hump-shaped. Intuitively, arbitrageurs can cause assets to become less correlated because the increase in volatilities that they cause can swamp the increase in covariance.

A hump-shaped pattern is possible for expected excess returns as well. This is more surprising because Corollary 1 shows that an increase in arbitrageur wealth w_t lowers the market prices of the Brownian risks, which are equal to expected excess returns per unit of risk exposure. Offsetting this effect is the fact that for w_t to the left of the volatility hump, an increase in w_t raises volatility because it raises endogenous risk. The latter effect can dominate and cause expected excess returns to increase with w_t for small values of w_t. Proposition 9 shows that when hedgers are short-lived, the latter effect always dominates for $\gamma < K$, and can dominate for larger values of γ as well for sufficiently large z.

PROPOSITION 9: *Suppose that hedgers are short-lived. For small arbitrageur wealth w_t, an increase in w_t raises the expected excess return of each asset in absolute value if $\gamma < K$. If $\gamma > K$, then the same result holds provided that $z > \frac{\gamma(\frac{r}{\rho-r}\gamma+1)}{\gamma+1}$.*

Figure 2 plots the Sharpe ratios, expected excess returns, volatilities, and correlations of long-maturity assets as a function of arbitrageur wealth for long-lived hedgers and for $\gamma = 0.5$ and 2. Consistent with Proposition 8, shown for short-lived hedgers and arbitrageurs with logarithmic utility, volatility and correlation are hump-shaped in arbitrageur wealth. The comparison across the two values of γ is also interesting. As one would expect, Sharpe ratios increase in γ. Expected excess returns, however, can be larger for the smaller value of γ. This is because when arbitrageurs are less risk-averse, they establish larger positions and this generates more endogenous risk. As in Proposition 9, the effect of larger endogenous risk on expected excess returns can dominate that of smaller Sharpe ratios.

IV. Liquidity Risk

In this section, we explore the implications of our model for liquidity risk. We assume long-maturity assets, as in the previous section, and define liquidity based on the effect that hedgers have on prices. Consider an increase in the parameter u_n that characterizes hedgers' willingness to sell asset n. This triggers a decrease of $\frac{\partial X_{nt}}{\partial u_n}$ in the quantity of the asset held by the hedgers, and a

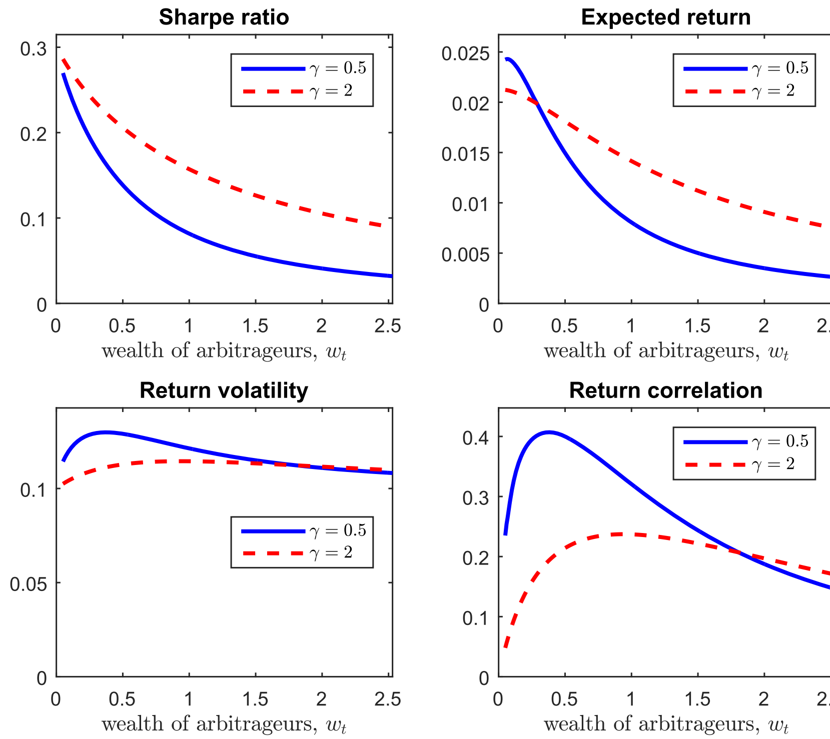

Figure 2. Assets' Sharpe ratios, expected excess returns, volatilities, and correlations as a function of arbitrageur wealth w_t. The plots assume two symmetric and uncorrelated risky assets, and set $\sqrt{u^\top \Sigma u} = 15\%$, $\alpha = 2$, $\rho = 4\%$, and $r = 2\%$. In the baseline case, represented by the blue solid line, γ is set to 0.5. The plots examine the effect of raising γ to 2. The baseline case is identical to that in Figure 1, which also examines the effect of raising γ to 2 (on dynamic risk aversion and positions). (Color figure can be viewed at wileyonlinelibrary.com)

decrease of $\frac{\partial S_{nt}}{\partial u_n}$ in the asset price. Asset n has low liquidity if the price change per unit of quantity traded is large. That is, the illiquidity of asset n is defined by

$$\lambda_{nt} \equiv \frac{\frac{\partial S_{nt}}{\partial u_n}}{\frac{\partial X_{nt}}{\partial u_n}}. \tag{48}$$

Defining illiquidity in terms of price impact follows Kyle (1985). Kyle and Xiong (2001), Xiong (2001), and Johnson (2008) use similar constructions to ours in asset-pricing settings by defining illiquidity as the derivative of price with respect to supply.

Proposition 10: *Illiquidity λ_{nt} is equal to*

$$-g(w_t)\left(1 + \frac{A(w_t) - F'(w_t)}{\alpha} + u^\top \Sigma u g'(w_t)\right)\Sigma_{nn}. \tag{49}$$

It converges to infinity when arbitrageur wealth w_t goes to zero, and to zero when w_t goes to infinity. When hedgers are short-lived, illiquidity is positive for all $w_t \in (0, \infty)$. When, in addition, arbitrageurs have logarithmic utility ($\gamma = 1$), illiquidity decreases in w_t.

Proposition 10 identifies a time series and a cross-sectional dimension of illiquidity. In the time series, illiquidity varies in response to changes in arbitrageur wealth. Our numerical solutions indicate that illiquidity is a positive and decreasing function of wealth, a finding that Proposition 10 confirms in the case in which hedgers are short-lived and arbitrageurs have logarithmic utility. The time series variation of illiquidity is common across assets and corresponds to the term multiplying Σ_{nn} in (49). In the cross-section, illiquidity is higher for assets with more volatile cash flows. The dependence of illiquidity on an asset n is through Σ_{nn}, the asset's cash flow variance. The time series and cross-sectional dimensions of illiquidity interact: assets with more volatile cash flows have higher illiquidity for any given level of wealth, and the time variation of their illiquidity is more pronounced.

We next compute the covariance between asset returns and aggregate illiquidity. Since illiquidity varies over time because of arbitrageur wealth, and with an inverse relationship, the covariance of the return vector with illiquidity is equal to the covariance with wealth times a negative coefficient. Proposition 7 implies, in turn, that the covariance of the return vector with wealth is proportional to Σu. This is the covariance between asset cash flows and the cash flows of the portfolio u, which characterizes hedgers' supply. The intuition for the proportionality is that when arbitrageurs realize losses, they sell a fraction of u and this lowers asset prices according to the covariance with u. Therefore, the covariance between asset returns and aggregate illiquidity $\Lambda_t \equiv \frac{\sum_{n=1}^{N} \lambda_{nt}}{N}$ is

$$\frac{\mathrm{Cov}_t(d\Lambda_t, dR_t)}{dt} = C^{\Lambda}(w_t)\Sigma u, \tag{50}$$

where $C^{\Lambda}(w_t)$ is a negative coefficient. Assets that suffer the most when aggregate illiquidity increases and arbitrageurs sell a fraction of the portfolio u are those corresponding to large components $(\Sigma u)_n$ of Σu. They have volatile cash flows (high Σ_{nn}), are in high supply by hedgers (high u_n), or correlate highly with assets with those characteristics.

Using Proposition 10, we can compute two additional liquidity-related covariances: the covariance between an asset's illiquidity and aggregate illiquidity, and the covariance between an asset's illiquidity and the aggregate return. We take the aggregate return to be that of the portfolio u, which characterizes hedgers' supply. Acharya and Pedersen (2005) show theoretically, within a model with exogenous transaction costs, that both covariances are linked to expected returns in the cross-section. In our model, the time-variation of the illiquidity of an asset n is proportional to the asset's cash flow variance Σ_{nn}. Therefore, the covariances between the asset's illiquidity on the one hand, and aggregate illiquidity or the aggregate return on the other, are proportional to Σ_{nn}.

COROLLARY 3: *In the cross-section of assets:*

(1) The covariance between asset n's return dR_{nt} and aggregate illiquidity Λ_t is proportional to the covariance $(\Sigma u)_n$ between the asset's cash flows and the cash flows of the hedger-supplied portfolio u.
(2) The covariance between asset n's illiquidity λ_{nt} and aggregate illiquidity Λ_t is proportional to the variance Σ_{nn} of the asset's cash flows.
(3) The covariance between asset n's illiquidity λ_{nt} and aggregate return $u^\top dR_t$ is proportional to the variance Σ_{nn} of the asset's cash flows.

When hedgers are short-lived and arbitrageurs have logarithmic utility ($\gamma = 1$), the proportionality coefficient $C^\Lambda(w_t)$ in the first relationship is negative, and those in the second and third relationships are positive and negative, respectively.

We next determine the link between liquidity-related covariances and expected returns. Equation (44) shows that the expected excess return of an asset n is proportional to $(\Sigma u)_n$. This is exactly proportional to the covariance between the asset's return and aggregate illiquidity. Thus, aggregate illiquidity is a priced risk factor that explains expected returns perfectly. Intuitively, assets are priced by the portfolio of arbitrageurs, who are the marginal agents. Moreover, the covariance between asset returns and aggregate illiquidity identifies that portfolio perfectly. This is because (i) changes in aggregate illiquidity are driven by arbitrageur wealth, and (ii) the portfolio of trades that arbitrageurs perform when their wealth changes is proportional to their existing portfolio and impacts returns proportionately to the covariance with that portfolio.

The covariances between an asset's illiquidity on the one hand, and aggregate illiquidity or the aggregate return on the other, are less informative about expected returns. Indeed, these covariances are proportional to cash flow variance Σ_{nn}, which is only a component of $(\Sigma u)_n$. Thus, these covariances proxy for the true priced risk factor but imperfectly so.

COROLLARY 4: *In the cross-section of assets, expected excess returns are proportional to the covariance between returns and aggregate illiquidity. When hedgers are short-lived and arbitrageurs have logarithmic utility ($\gamma = 1$), the proportionality coefficient $\Pi^\Lambda(w_t)$ in this relationship is negative.*

The premium associated with the illiquidity risk factor is the expected excess return per unit of covariance with the factor. Hence, it coincides with the proportionality coefficient $\Pi^\Lambda(w_t)$ in the relationship between expected excess returns and the covariance of returns with aggregate illiquidity.

Both the premium of the illiquidity risk factor and the covariance of returns with aggregate illiquidity vary over time in response to changes in arbitrageur wealth w_t. When w_t is small, illiquidity is high and highly sensitive to changes in wealth. As a result, assets' covariance with illiquidity is large in absolute value when w_t is small and decreases when w_t increases. Conversely, because the premium of the illiquidity risk factor is the expected excess return per unit of covariance, it is small in absolute value when w_t is small and increases when w_t increases.[13]

PROPOSITION 11: *The common component $C^{\Lambda}(w_t)$ of assets' covariance with aggregate illiquidity and the premium $\Pi^{\Lambda}(w_t)$ of the illiquidity risk factor have the following properties:*

(1) $C^{\Lambda}(w_t)$ converges to minus infinity when arbitrageur wealth w_t goes to zero, and to zero when w_t goes to infinity.
(2) $\Pi^{\Lambda}(w_t)$ converges to zero when w_t goes to zero. When hedgers are short-lived, $\Pi^{\Lambda}(w_t)$ converges to minus infinity when w_t goes to infinity.

Aggregate illiquidity explains expected returns perfectly in our model because it is a monotone function of arbitrageur wealth w_t. Hence, any other monotone function of w_t would also have this property. Recent empirical papers on intermediary asset pricing, such as Adrian, Etula, and Muir (2014) and He, Kelly, and Manela (2017), use the leverage of specific groups of financial intermediaries as a risk factor, and show that it can price a large cross-section of assets. The findings of these papers are consistent with our model, insofar as leverage is a monotone function of w_t.[14] Our modeling approach suggests that these findings may be explained even with minimal frictions, for example, no leverage constraints. On the other hand, introducing frictions may give better guidance on which function of w_t to use as a risk factor.

V. Positive Supply

Our analysis so far assumes that the long-maturity assets are in zero supply ($s = 0$). Proposition 12 shows that this assumption is without loss of generality when hedgers are long-lived: risk-sharing and asset prices are the same as when assets are in zero supply, provided that we replace u by $s + u$. The intuition is that the stream of random endowments $u^{\top} dD_t$ that a long-lived hedger receives over time is equivalent to an endowment of u shares in the long-maturity assets. Thus, hedgers generate a supply u, which is added to the

[13] Since the premium of the illiquidity risk factor depends on w_t, it can be viewed as a function of illiquidity itself, which is a monotone decreasing function of w_t.

[14] A natural measure of leverage in our model is the total risk exposure of arbitrageurs as a fraction of their wealth. Since the exposure of arbitrageurs to the Brownian risks is $\sigma y_t = (\sigma_{St} + \sigma) Y_t$, (18) implies that leverage is

$$\frac{1^{\top} \sigma y_t}{w_t} = \frac{\alpha 1^{\top} \sigma u}{\alpha w_t + A(w_t) w_t - F'(w_t) w_t}.$$

Lemma IA.8, stated and proven in the Internet Appendix, shows that this measure of leverage is decreasing in w_t when hedgers are short-lived and $\gamma \leq 1$. Our numerical solutions suggest that this result holds more generally. A countercyclical leverage is consistent with the empirical finding of He, Kelly, and Manela (2017), and the theory of He and Krishnamurthy (2013) and Brunnermeier and Sannikov (2014), but is inconsistent with Adrian, Etula, and Muir (2014).

supply s coming from the asset issuers. If instead asset issuers generate no supply and hedgers generate $s+u$, then hedgers reduce their demand by s. Since the reduction in hedgers' demand exactly offsets the reduction in asset issuers' supply, equilibrium prices remain the same and so does risk-sharing.

PROPOSITION 12: *Suppose that hedgers are long-lived and long-maturity assets are in positive supply s. If* (S_t, X_t, Y_t) *is an equilibrium, then* $(S_t, X_t - s, Y_t)$ *is an equilibrium when assets are in zero supply* ($s = 0$) *and u is replaced by* $s + u$. *In both equilibria, the exposures of hedgers and arbitrageurs to the Brownian shocks and the prices of the assets are the same.*

When hedgers are short-lived, the equivalence between positive-supply and zero-supply equilibria does not hold. This is because any given short-lived hedger receives a random endowment only in the next instant, so that endowment is not equivalent to one in the long-maturity assets. Prices and expected returns, however, have the same general form as with zero supply.

PROPOSITION 13: *Suppose that hedgers are short-lived and that long-maturity assets are in positive supply s. Then the prices of the assets are given by*

$$S(w_t) = \frac{\bar{D}}{r} + g(w_t)\Sigma(s+u), \tag{51}$$

where the scalar function $g(w_t)$ *solves the ODE (IA.159). The assets' expected excess returns are*

$$\frac{\mathrm{E}_t(dR_t)}{dt} = \frac{\alpha A(w_t)}{\alpha + A(w_t) - \alpha g'(w_t)(s+u)^\top \Sigma s}\left[(s+u)^\top \Sigma (s+u) f(w_t) + 1\right]\Sigma(s+u), \tag{52}$$

where

$$f(w_t) \equiv \frac{\alpha g'(w_t)}{\alpha + A(w_t) - \alpha g'(w_t)(s+u)^\top \Sigma s}.$$

The price of an asset n, given in (51), is the sum of the price $\frac{\bar{D}_n}{r}$ that would arise if arbitrageurs had infinite wealth and of a discount that is proportional to the asset's covariance $(\Sigma(s+u))_n$ with aggregate supply $s+u$. Changes in arbitrageur wealth affect the discount, and hence their effect is proportional to the covariance. Since the asset's expected return, given in (52), is proportional to the same covariance, aggregate illiquidity is a priced risk factor and explains expected returns perfectly, as in the case of zero supply. The equivalence between positive and zero supply does not hold for short-lived hedgers because the proportionality coefficients (e.g., the function $g(w_t)$) depend on both s and u rather than only on their sum.

VI. Conclusion

We develop a dynamic model of liquidity provision in which hedgers can trade multiple risky assets with arbitrageurs. Arbitrageurs have CRRA utility over consumption, and their capital matters because of wealth effects. We strip out frictions such as asymmetric information and leverage constraints. At the same time, we depart from most frictionless asset-pricing models by fixing the riskless rate and suppressing wealth effects for the arbitrageurs' counterparties. Under this combination of assumptions, we prove general analytical results on equilibrium prices and allocations. In particular, we characterize how arbitrageurs' risk aversion, the endogenous risk that they generate, and the pricing of that risk depend on arbitrageur wealth and hedger characteristics. We also show that arbitrageur wealth is the single priced risk factor, and that aggregate illiquidity, which declines in wealth, captures that factor.

One important extension of our model is to allow the supply u coming from hedgers to be time-varying and stochastic. Such an extension would give our measure of illiquidity (48) stronger empirical content because that measure is based on supply shocks. We could also compare our measure to measures commonly used in empirical work, for example, Amihud (2002) and Pastor and Stambaugh (2003), and identify their properties when volume arises because of both supply shocks affecting liquidity demanders and wealth effects affecting liquidity providers. In a similar spirit, supply shocks may generate a tighter relationship between volatility and our measure of illiquidity. Indeed, volatility in our model is driven by wealth effects of arbitrageurs and is hump-shaped in wealth, while illiquidity is defined based on supply shocks affecting hedgers and is decreasing in wealth.

Extending our model to stochastic u would strengthen our interpretation of arbitrageurs as specialized liquidity providers. Indeed, a common view of liquidity providers (e.g., Grossman and Miller (1988)) is that they absorb temporary imbalances between demand and supply. A natural interpretation of these imbalances within our model is as shocks to u. In the presence of these shocks, liquidity provision could become distinct from sharing the aggregate risk in the economy, which also includes the supply s coming from issuers. Modeling this idea would require introducing additional agents to the model who absorb s but are unable to identify shocks to u or trade on them.

Initial submission: January 25, 2016; Accepted: March 29, 2018
Editors: Stefan Nagel, Philip Bond, Amit Seru, and Wei Xiong

REFERENCES

Acharya, Viral, Lars Lochstoer, and Tarun Ramadorai, 2013, Limits to arbitrage and hedging: Evidence from commodity markets, *Journal of Financial Economics* 109, 441–465.

Acharya, Viral, and Lasse Pedersen, 2005, Asset pricing with liquidity risk, *Journal of Financial Economics* 77, 375–410.

Adrian, Tobias, Erkko Etula, and Tyler Muir, 2014, Financial intermediaries and the cross-section of asset returns, *Journal of Finance* 69, 2557–2596.

Amihud, Yakov, 2002, Illiquidity and stock returns: Cross-section and time-series effects, *Journal of Financial Markets* 5, 31–56.

Aragon, George, and Philip Strahan, 2012, Hedge funds as liquidity providers: Evidence from the Lehman bankruptcy, *Journal of Financial Economics* 103, 570–587.

Basak, Suleyman, and Anna Pavlova, 2013, Asset prices and institutional investors, *American Economic Review* 103, 1728–1758.

Bhamra, Harjoat, and Raman Uppal, 2009, The effect of introducing a non-redundant derivative on the volatility of stock-market returns when agents differ in risk aversion, *Review of Financial Studies* 22, 2303–2330.

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.

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

Chan, Lewis, and Leonid Kogan, 2002, Catching up with the Joneses: Heterogeneous preferences and the dynamics of asset prices, *Journal of Political Economy* 110, 1255–1285.

Cheng, Ing-Haw, and Wei Xiong, 2014, Why do hedgers trade so much? *Journal of Legal Studies* 43, S183–S207.

Cheng, Ing-Haw, and Wei Xiong, 2015, Convective risk flows in commodity futures markets, *Review of Finance* 19, 1733–1781.

Chordia, Tarun, Richard Roll, and Avanidhar Subrahmanyam, 2000, Commonality in liquidity, *Journal of Financial Economics* 56, 3–28.

Comerton-Forde, Carole, Terrence Hendershott, Charles Jones, Pamela Moulton, and Mark Seasholes, 2010, Time variation in liquidity: The role of market-maker inventories and revenues, *Journal of Finance* 65, 295–331.

Dumas, Bernard, 1989, Two-person dynamic equilibrium in the capital market, *Review of Financial Studies* 2, 157–188.

Ehling, Paul, and Christian Heyerdahl-Larsen, 2017, Correlations, *Management Science* 63, 1919–1937.

Franzoni, Francesco, Eric Nowak, and Ludovic Phalippou, 2012, Private equity performance and liquidity risk, *Journal of Finance* 67, 2341–2373.

Garleanu, Nicolae, and Stavros Panageas, 2015, Young, old, conservative, and bold: The implications of heterogeneity and finite lives for asset pricing, *Journal of Political Economy* 123, 670–685.

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.

Gromb, Denis, and Dimitri Vayanos, 2002, Equilibrium and welfare in markets with financially constrained arbitrageurs, *Journal of Financial Economics* 66, 361–407.

Gromb, Denis, and Dimitri Vayanos, 2018, The dynamics of financially constrained arbitrage, *Journal of Finance* 73, 1713–1750.

Grossman, Sanford, and Merton Miller, 1988, Liquidity and market structure, *Journal of Finance* 43, 617–637.

Hameed, Allaudeen, Wenjin Kang, and S. Viswanathan, 2010, Stock market declines and liquidity, *Journal of Finance* 65, 257–293.

Hasbrouck, Joel, and Duane Seppi, 2001, Common factors in prices, order flows and liquidity, *Journal of Financial Economics* 59, 383–411.

He, Zhiguo, Bryan Kelly, and Asaf Manela, 2017, Intermediary asset pricing: New evidence from many asset classes, *Journal of Financial Economics* 126, 1–35.

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

Holmstrom, Bengt, and Jean Tirole, 2001, LAPM: A liquidity-based asset pricing model, *Journal of Finance* 56, 1837–1867.

Huberman, Gur, and Dominika Halka, 2001, Systematic liquidity, *Journal of Financial Research* 24, 161–178.

Isaenko, Sergey, 2015, Equilibrium theory of stock market crashes, *Journal of Economic Dynamics and Control* 60, 73–94.

Johnson, Timothy, 2008, Volume, liquidity, and liquidity risk, *Journal of Financial Economics* 87, 388–417.

Jylha, Petri, and Matti Suominen, 2011, Speculative capital and currency carry trades, *Journal of Financial Economics* 99, 60–75.

Kang, Wenjin, Geert Rouwenhorst, and Ke Tang, 2016, A tale of two premiums: The role of hedgers and speculators in commodity futures markets, Working paper, Yale School of Management.

Klimenko, Nataliya, Sebastian Pfeil, Jean-Charles Rochet, and Gianni De Nicolo, 2016, Aggregate bank capital and credit dynamics, Working paper No. 16-42, Swiss Finance Institute.

Kondor, Peter, 2009, Risk in dynamic arbitrage: Price effects of convergence trading, *Journal of Finance* 64, 638–658.

Kondor, Peter, and Dimitri Vayanos, 2014, Liquidity risk and the dynamics of arbitrage capital, Working paper No. 38, Paul Woolley Centre, London School of Economics.

Kyle, Albert, 1985, Continuous auctions and insider trading, *Econometrica* 53, 1315–1336.

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

Liu, Qi, Lei Lu, Bo Sun, and Hongjun Yan, 2015, A model of anomaly discovery, Working paper, DePaul University.

Longstaff, Francis, and Jiang Wang, 2012, Asset pricing and the credit market, *Review of Financial Studies* 25, 3169–3215.

Merton, Robert, 1971, Optimum consumption and portfolio rules in a continuous-time model, *Journal of Economic Theory* 3, 373–413.

Pastor, Lubos, and Robert Stambaugh, 2003, Liquidity risk and expected stock returns, *Journal of Political Economy* 111, 642–685.

Sadka, Ronnie, 2010, Liquidity risk and the cross-section of hedge fund returns, *Journal of Financial Economics* 98, 54–71.

Vayanos, Dimitri, and Jiang Wang, 2013, Market liquidity—Theory and empirical evidence, in George Constantinides, Milton Harris, and Rene Stulz, eds.: *Handbook of the Economics of Finance* (North Holland, Amsterdam).

Wang, Jiang, 1996, The term structure of interest rates in a pure exchange economy with heterogeneous investors, *Journal of Financial Economics* 41, 75–110.

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

Supporting Information

Additional Supporting Information may be found in the online version of this article at the publisher's website:

Appendix S1: Internet Appendix.