

A Test of the Modigliani-Miller Invariance Theorem and Arbitrage in Experimental Asset Markets

GARY CHARNESS and TIBOR NEUGEBAUER

ABSTRACT

Modigliani and Miller show that the total market value of a firm is unaffected by a repackaging of asset return streams to equity and debt if pricing is arbitrage-free. We investigate this invariance theorem in experimental asset markets, finding value-invariance for assets of identical risks when returns are perfectly correlated. However, exploiting price discrepancies has risk when returns have the same expected value but are uncorrelated, in which case the law of one price is violated. Discrepancies shrink in consecutive markets, but persist even with experienced traders. In markets where overall trader acuity is high, assets trade closer to parity.

IN THEIR SEMINAL PAPER, Modigliani and Miller (1958, 1963) show mathematically that the market value of a firm is invariant to the firm's leverage—different packaging of contractual claims on the firm's assets does not impact the total market value of the firm's debt and equity. The Modigliani and Miller—henceforth MM—value-invariance theorem suggests that the law of one price prevails for assets of the same “risk class.” The core of the theorem is an arbitrage proof, whereby if two assets, one leveraged and one unleveraged, represent claims on the same cash flow, any market discrepancies that arise are arbitrated away. But due to its assumptions of perfect capital markets and the no-limits-to-arbitrage condition (which requires the perfect positive correlation

*Tibor Neugebauer is at the University of Luxembourg and is the corresponding author. Gary Charness is at University of California, Santa Barbara. The authors have no conflicts of interest to disclose. The authors obtained Institutional Review Board approval for data collection. We gratefully acknowledge the helpful comments of Bruno Biais (the Editor), the anonymous Associate Editor, two anonymous reviewers, Peter Bossaerts, Martin Duwfenberg, Darren Duxbury, Catherine Eckel, Ernan Haruvy, Chad Kendall, Roman Kräussl, Ulf von Lilienfeld-Toal, Raj Mehra, Luba Petersen, Jianying Qiu, Kalle Rinne, Jean-Charles Rochet, Jason Shachat, and Julian Williams for helpful comments. We also thank seminar participants at Durham Business School, University of Luxembourg, and Radboud University of Nijmegen, and at the conferences Experimental Finance in Nijmegen, the Netherlands, Experimental Finance in Tucson, Arizona, the International Meeting on Experimental and Behavioral Sciences in Toulouse, France, Conference on Behavioural Aspects of Macroeconomics & Finance at House of Finance, Frankfurt, Germany, Barcelona GSE Summer Forum, Spain, and CESifo in Munich, Germany. The scientific research presented in this publication received financial support from the National Research Fund of Luxembourg (INTER/MOBILITY/12/5685107), and University of Luxembourg provided funding of the experiments through an internal research project (F2R-LSF-PUL-10IDIA).

DOI: 10.1111/jofi.12736

of asset returns, no fees on the use of leverage, etc.), the MM theorem has not been satisfactorily tested on real-world market data. Its empirical significance has thus been unclear.¹

Such a test is feasible in the laboratory, however. Providing an empirical test of the MM theorem is a primary purpose of this study. Since perfect return correlation is rare in naturally occurring equities, we also check how limits to arbitrage affect the empirical validity of the MM theorem with regards to cross-asset pricing. In particular, we examine whether a perfect positive correlation between asset returns is necessary for the empirical validity of value-invariance or whether the same expected (rather than identical) future return is sufficient, as suggested by the capital asset pricing model for our setting. Our data indeed suggest that perfect correlation is indeed necessary for the law of one price to prevail.

Our main design adapts the experimental asset market research of Smith, Suchanek, and Williams (1988), which features multiperiod cash flows, zero interest rates, and a repetition of markets with experienced subjects.² However, in contrast to the standard, single-asset market approach of Smith, Suchanek, and Williams (1988), and in line with MM, we allow for simultaneous trading in two shares of the same “equivalence class.” These “twin shares,” which we refer to as the L-shares and U-shares, are claims on the same underlying uncertain future cash flows. In one treatment, the returns of the L-share and the U-share are perfectly correlated and thus any price discrepancies that arise can be arbitrated away at no risk. In a second treatment, where the returns of the L-share and the U-share are uncorrelated, we study the impact of limits to arbitrage. In both cases, the expected stream to shareholders of L-shares and U-shares differs by a constant amount, which we refer to as the “synthetic” value of debt, as we discuss below, so the L-share and U-share represent “leveraged” and “unleveraged” equity streams.

¹ The assumption of perfect capital markets requires, among other things, that no taxes and transaction fees be levied and that the same interest rate applies to everyone. Lamont and Thaler (2003) present several real-world examples where the law of one price is violated. They argue that these violations result from limits to arbitrage. An early objection concerned the applicability of value-invariance in relation to the variation of payout policy. Modigliani and Miller (1959) replied to this objection by stating that a firm’s dividend policy is irrelevant for the value of the company. However, it is now widely accepted that dividends impact empirical valuations (for a recent discussion of the dividend puzzle, see DeAngelo and DeAngelo (2006)). With the dividend-irrelevance theorem thus empirically rejected, it is considered as of theoretical interest only. The value-invariance theorem and its proof, however, have remained widely accepted in the profession even without empirical evidence to support it.

² See Palan (2013) for a recent literature survey. The literature is mainly concerned with measuring mispricing in the single-asset market. The conclusion is that confusion of subjects and speculation are the main sources of price deviations from fundamentals in the laboratory (e.g., Smith, Suchanek, and Williams (1988), Lei, Noussair and Plott (2001), Kirchler, Huber, and Stoeckl (2012)). Mispricing occurs in the single-asset market, and also when assets are simultaneously traded in two markets (Ackert et al. 2009, Chan, Lei and Vesely 2013). Smith and Porter (1995), Noussair and Tucker (2006), and Noussair, Tucker, and Yu (2016) report reduced mispricing when a futures market enables subjects to arbitrage price discrepancies of underlying asset and futures contracts.

By comparing the market prices of shares, we present a very simple test of the MM theorem.³ If at any point in time the price deviates from parity, in other words if the difference between the L-share and the U-share is not the same as the synthetic debt value for the investor, then a market participant can exploit the price discrepancy. Since short-selling and borrowing is costless, a trader can make a riskless arbitrage gain by going short the expensive share and long the inexpensive share. Exploited pricing discrepancies thereby undo the divergence of market values.

Our data provide support for the MM theorem since average prices are close to parity, even though some price discrepancies and deviations from the risk-neutral value persist throughout the experiment. In our perfect-correlation treatment, we observe that perfect correlation is essential for value indifference as we control for variations in correlation. In our control no-correlation treatment, we consider independent draws of dividends of the two simultaneously traded shares. Here, L-shares and U-shares have the same expected dividend and idiosyncratic risk as in the perfect-correlation treatment, but an asset swap has risk.

We find a clear treatment effect: we observe a higher level of price discrepancies in the no-correlation treatment. With perfect correlation our measures of cross-asset price discrepancy, relative frequency of discrepant limit orders, and deviation from fundamental dividend value indicate smaller deviations from the theoretical benchmarks than in the no-correlation treatment.⁴ The market corrects relative mispricing with perfectly correlated returns but not as well with independent asset returns.

Thus, although potential price discrepancies never disappear in absolute quantitative terms for the perfect-correlation treatment, our data provide strong qualitative support for the equilibrium through the comparison of our treatments. That said, as with evidence observed with experienced subjects in single-asset market studies (e.g., Haruvy, Lahav, and Noussair (2007), Dufwenberg, Lindqvist, and Moore (2005)), the price deviation from fundamental dividend values declines in consecutive markets in both treatments. The movement towards the theoretical benchmarks, however, seems to be more rapid in the perfect-correlation treatment than in the no-correlation treatment, both in terms of the decrease in price discrepancies and the deviation from fundamental dividend value. Nevertheless, some potential price discrepancies persist in both treatments, even with experienced subjects.

We next consider the impact of traders' acuity, as measured by the cognitive reflection task (CRT; Frederick (2005)), on the level of price discrepancy. The literature suggests that smart traders search for and eliminate price discrepancies.⁵ Our measure correlates with the reduction in price discrepancy on the

³ In Section II, we show how our design maps into the MM theorem.

⁴ In line with studies that apply a zero discount rate in the Smith, Suchanek, and Williams (1988) experimental framework, we define the fundamental dividend value as the sum of discounted expected future dividends.

⁵ See the discussions in Shleifer (2000) and Lamont and Thaler (2003).

overall sample. Although there are some price discrepancies in markets with higher trader aptitude, these are substantially less common and smaller. However, we find no evidence of subjects specializing in arbitrage, so we conclude that existing discrepancies are dissolved by the equilibration dynamics of the market rather than by skilled individual traders. Even so, low trader aptitude leads to pricing discrepancies in the no-correlation treatment.

Overall, we provide evidence on how limits to arbitrage (such as our no-correlation treatment) impact value-invariance. One main contribution is that we are able to empirically validate value-invariance under perfect correlation. A second main contribution is the observation of a treatment effect, since price discrepancies in the market increase substantially when one moves from perfect positive correlation to zero correlation. The deviations appear to be driven by relative overvaluation of the U-share. In the perfect-correlation treatment these are evened out, but in the no-correlation treatment they prevail in markets with overall low trader aptitude. Finally, we provide evidence that high trader acuity significantly reduces the price discrepancy in the market and shares trade closer to fundamental dividend value.

We conduct additional treatments to better understand the patterns that we observe in the initial perfect-correlation and no-correlation treatments. In one treatment, we investigate 10-period markets with only a single asset present. We find no systematic mispricing in this treatment. In an additional treatment, we investigate a single-period market with two assets. We again find no significant differences from parity pricing. Earlier experimental results show that, in simple settings, double-auction market prices easily average around the competitive-equilibrium prediction.⁶ We propose a more extreme test of MM in the context of long-lived assets and declining dividend value, which is known for being cognitively demanding. We find differential behavior in the perfect-correlation treatment versus the no-correlation treatment.

The remainder of the paper is organized as follows. In Section I, we briefly review related literature and we discuss our experimental design. In Section II, we discuss the MM theorem in light of our design. In Section III, we introduce our measures of price discrepancies and develop our hypotheses. We present our experimental results in Section IV. Results of robustness tests are presented in Section V and we conclude in Section VI.

I. Literature Review and Experimental Design

A. Literature Review

Our study contributes to the modest experimental literature on the market's ability to reduce or eliminate arbitrage opportunities.⁷ The documented persistence in price discrepancies confirms earlier empirical results. O'Brien and

⁶ In the classic study of Vernon Smith (1962), market equilibrium dynamics were strong in small markets. Gode and Sunder (1993) found strong equilibrium dynamics even in simulations with randomized algorithms ("zero intelligence traders").

⁷ See the surveys in Cadsby and Maynes (1998) and Sunder (1995).

Srivastava (1991) replicate portfolios of options, stocks, and cash in a multiple-asset experimental market with two stages and information asymmetry. The authors show that if the information asymmetry cannot be resolved, price discrepancies frequently persist.

Oliven and Rietz (2004) investigate data from the 1992 Iowa presidential election market (IEM), a large-scale experiment conducted for several months on the Internet. Arbitrage opportunities in this market were quite easy to spot—if the value of the market portfolio deviated from the issue price, any trader could make an arbitrage gain by selling or buying at said issue price. Oliven and Rietz (2004) find a substantial number of price discrepancies, but show that these were quickly driven out. Rietz (2005) conducts a laboratory prediction-market experiment with state-contingent claims. Arbitrage opportunities were easily spotted (as in the IEM), but trading was over 100 minutes rather than 100 days. Rietz (2005) concludes that this market is prone to violate the no-arbitrage requirement. However, if (as in one treatment) the experimenter automatically eliminates each price discrepancy, this automatic arbitrageur was involved in most trades in the experiment. Abbink and Rockenbach (2006) report that, even after hours of experience, both students and professional traders left arbitrage opportunities unexploited in an individual task of allocating cash to options, bonds, and stock.

To the best of our knowledge, Levati, Qiu, and Mahagaonkar (2012) conduct the only other experimental study to test the MM theorem. Their design precludes the possibility of arbitrage or (homemade) leveraging and unleveraging.⁸ Levati, Qiu, and Mahagaonkar (2012) examine evaluations of eight independent lotteries with varying degree of risk in a sequence of experimental single-asset call auction markets, where the risk represents the level of company leverage. In contrast to our perfect-correlation treatment, but similar to our no-correlation treatment, the market data in Levati, Qiu, and Mahagaonkar (2012) show no support for value-invariance. The authors acknowledge that precluding any possibility of arbitrage is a potential reason for this result.

B. Experimental Design

In our experiment, subjects could buy and sell multiperiod-lived assets in continuous double-auction markets. The assets involved claims to a stochastic dividend stream over a lifespan of 10 periods, $T = 10$, after which the assets had no further value. The instructions can be found in the Internet Appendix.⁹

Trading occurred in two asset classes, which we refer to as L-shares and U-shares.¹⁰ We follow design 4 from Smith, Suchanek, and Williams (1988). Specifically, the dividend paid on an L-share was independently drawn {0, 8, 28 or 60 cents} with equal probability at the end of each period, so that the expected

⁸ Stiglitz (1969) proves MM value-invariance within a general equilibrium setting, without explicit arbitrage assumptions.

⁹ The Internet Appendix may be found in the online version of this article.

¹⁰ In the experiment we said “A-share” and “B-share” instead of L-share and U-share.

Table I
Initial Individual Endowments and Moments

The first column reports the individual initial unit endowments with shares and cash. The second and third columns report first and second moments of dividend value distributions; ^aexpected dividend value and total variance are per unit and period, $t \leq T = 10$. The initial total variance is 5,360 for each share; the initial expected payoffs of the L-share and the U-share are 240 and 480, respectively. Variances and expected dividend values decrease linearly over time.

	Initial unit endowment	Expected dividend value/unit ^a	Initial total variance/unit ^a
L-shares	2	$24 (T - t + 1)$	$536 (T - t + 1)$
U-shares	2	$48 (T - t + 1)$	$536 (T - t + 1)$
Cash units	1,200	1	–

dividends per period were 24 cents on L-shares. The possible dividends paid on the U-share were 24 cents higher, and so the expected dividends per share were 48 cents on U-shares. The interest rate was zero and thus the discounted sum of expected dividends, conventionally referred to as fundamental dividend value, of L-shares and U-shares were initially 240 and 480 cents and decreased by 24 and 48 cents per period (see Table I). After the last dividend payment, all shares were worthless and the final cash balance was recorded.

Our treatment variation between subjects is the correlation ρ_{LU} between the dividends on L-shares and U-shares. In the perfect-correlation treatment, where $\rho_{LU} = 1$, the U-share paid exactly 24 cents more in each period than does the L-share. In the no-correlation treatment, where $\rho_{LU} = 0$, the dividend on the U-share was independently drawn {24, 32, 52, 84 cents} with equal probability.

Each market participant could electronically submit an unlimited number of bids and asks in the two simultaneous markets. Submitted bids and asks were publicly visible and could not be cancelled. The best outstanding bid and asks were available for immediate sale and purchase through confirmation by the other market participants. Upon a transaction, all bids of the buyer and all offers of the seller were cancelled in both markets and the price was publicly recorded. However, the buyer may have also placed offers to sell and the seller may have also placed bids to buy; any such offers and bids remained in the market. So, any arbitrageur who wanted to exploit a price discrepancy could immediately buy low and sell high by pressing the buttons.

Market participants were initially endowed with two L-shares, two U-shares, and 1,200 cents cash (see Table I). Traders were able to borrow up to 2,400 cents for the purchase of assets and could short-sell up to four L-shares and up to four U-shares without any margin requirements.¹¹ The trading flow was unaffected (i.e., there was no message indicating a short sale rather than a long sale) by

¹¹ Subjects would not exhaust their borrowing capacity for strategic reasons. Only 0.1% of subjects' end-of-period records indicate a cash balance below –2,200. These results correspond to individual bankruptcies. Bankruptcy rarely occurred and never occurred in the last two rounds of the experiment.

short sales and borrowings, which were displayed as negative numbers.¹² Each subject thus had a wide scope for financial decision-making.

To reduce confusion and in turn pricing discrepancies, subjects were reminded on screen about the expected future dividends and the sum of expected dividends for the remaining periods. Dividends, prices (open, low, high, closing, and average), number of transactions, and portfolio compositions in each past period were reported in tables. Since we are interested in the effects of experience, an experimental session involved four consecutive *markets*, each designed for nine subjects. The period length was 180 seconds in the first market and 90 seconds in subsequent markets.¹³ One of the four markets was chosen (with equal probability) for payment at the end of the session. A subject determined the payoff-decisive market by a die roll. Subjects received their final cash balance in the payoff-decisive market plus the payoff from two pre-market tasks (detailed below) and a show-up fee of \$15 in an envelope at their cabin. In the case of a negative cash balance, the subject's show-up fee would be correspondingly reduced, but not by more than \$5.

Subjects were recruited from a pool of economics and science students at the University of California, Santa Barbara, via ORSEE (Greiner (2015)). Each person drew a number from a tray upon arriving at the laboratory. This number indicated one's cabin number. Before the instructions were read for the trading markets, subjects engaged in two pre-market tasks to reveal to us their relevant personal traits. Information on payoffs from these tasks was communicated only at the end of the session. The first task was an investment game to assess the subjects' degree of risk aversion (Charness and Gneezy (2010)). This task required that subjects choose an amount $0 \leq X \leq \$10$ to allocate to each a risky asset that paid either 0 or $2.5X$ with equal probability and to a safe asset $\$10 - X \geq 0$ that paid out with certainty. One of the nine participants in the asset market would receive the payoff from the first experiment. The second task was the CRT (Frederick 2005), where the three questions were asked in a random order.¹⁴ Subjects had 90 seconds to answer the questions and earned \$1 per correct answer.

As part of the instructions, subjects had to successfully complete four practice exercises: a dividend questionnaire, a trading round, a forecasting reward

¹² Subjects would typically not exhaust their short-sale capacities for strategic reasons. Overall, 8% of subjects' end-of-period results indicate negative shareholdings. The short-sale limit of -4 shares was reached in 2% of observations, of which more than 50% correspond to individual bankruptcies.

¹³ We allowed more time in the first market for people to get accustomed to the decisions. There is no evidence (including in questionnaire reports) that subjects were short of time in the shorter intervals.

¹⁴ The questions were: (1) A hat and a suit cost \$110. The suit costs \$100 more than the hat. How much does the hat cost? (2) If it takes five machines five minutes to make five widgets, how long would it take 100 machines to make 100 widgets? (3) In a lake, there is a patch of lily pads. Every day the patch doubles in size. If it takes 48 days for the patch to cover the entire lake, how long would it take for the patch to cover half of it?

questionnaire,¹⁵ and a second trading round. After the four markets, subjects were debriefed in a questionnaire on personal details. The experiment was computerized with the software z-Tree (Fischbacher (2007)). Written instructions were distributed, and verbal instructions were played from a recording.

After conducting the two initial treatments, we conducted two additional treatments to further investigate patterns found in the first treatments. These consisted of 10-period asset markets with a single asset and one-period asset markets with two assets. The design and results can be found in Section V.

II. Modigliani-Miller Invariance Theorem

In this section we discuss the invariance theorem of MM (1958) in light of our experimental design (in particular, the perfect-correlation treatment).¹⁶ We begin by restating the MM [Invariance theorem]... (without taxes) and sketching the proof (MM (1958) p. 268f).

“[Invariance theorem]: Consider any company j and let \bar{X}_j stand for the ... expected profit before deducting interest. Denote by D_j the ... value of the debts of the company; by S_j the market value of ... common shares; and by $V_j \equiv S_j + D_j$ the market value of all securities.... Then ... we must have in equilibrium:

$$V_j \equiv S_j + D_j = \bar{X}_j / r_k, \quad \text{for any firm } j \text{ in class } k.” \quad (1)$$

The expected cash flows are discounted by the market-required return on assets, r_k , which is determined by the equivalence class k of the company's assets and the risk attitude of the market. The MM (1958) invariance theorem states that the total market value of the firm is invariant to its capital structure. It implies that identical cash flows are priced the same in equilibrium.¹⁷

¹⁵ Prior to each period, subjects made predictions about the average prices at which the assets would trade during the period. By asking subjects to reflect and predict market outcomes prior to the period, we aimed to deeply engage subjects in the experiment. Similar to the design in Haruvy, Lahav, and Noussair (2007), subjects initially submitted a forecast of the average price of each asset for each future period. In our design, however, subjects only updated their price forecasts of the current period prior to subsequent market openings. Subjects received salient rewards for accurate forecasts—the mean percentage deviation of the forecast from the realized average price was subtracted from one and the remainder was multiplied by \$6. The deviation in any period was capped at 100%. Periods without transactions did not count in the determination of the payoff, in either the numerator or the denominator. Subjects were rewarded for either the accuracy of initial forecasts or the accuracy of updated forecasts, with equal probability. The decision was made by computerized random draw after the last market period. The reward from forecasting was included in the subject's final cash balance.

¹⁶ Comments from Bruno Biais, Peter Bossaerts, and an Associate Editor helped us significantly improve this section.

¹⁷ Since in the MM world without taxes and riskless debt the assets are the same for the unleveraged firm (U) and the leveraged (L) firm, the required expected total return on assets must be the same independently of the company's debt ratio (see MM (1958) equation (4), p. 268);

$$\frac{\bar{X}_j}{V_U} = \frac{\bar{X}_j}{S_U} = \frac{\bar{X}_j}{S_L + D_L} = \frac{\bar{X}_j}{V_L} = r_k.$$

To prove the implication of the invariance theorem, MM (1958) compare two companies with the same total cash-flow X over the infinite horizon. The first company U is financed entirely by equity, so that at the end of the period equity holders get the entire return X . The other company L is leveraged, with debt of face value D , so that at the end of the period debt holders get rD and equity holders get $X - rD$. MM then show by arbitrage that one must have $V_U = S_U = S_L + D = V_L$, where S_U and S_L are the market values of equity, and V_U and V_L are the total values of the unleveraged and the leveraged company. Starting on page 269, MM analyze the arbitrage trade when the total market value of company L is larger than that of company U , that is, $V_L = S_L + D > V_U$. To take advantage of this situation, the arbitrageur sells shares in the leveraged company, borrows, and buys shares in the unleveraged company. He is therefore long a portfolio whose value is that of the unleveraged company minus debt service, and he is short the equity of the leveraged company. The former is a synthetic version of the latter. Since cash flows on both portfolios are the same, the investor is not exposed to risk. But at the inception of the position, the portfolio purchased is cheaper than the portfolio sold. Hence, there is an arbitrage profit. Since arbitrage cannot exist in equilibrium, the proof concludes.

How does our experimental setting map into this? Consider two stocks L-share and U-share. The dividend on the U-share is 24, 32, 52, or 84 with equal probability. The dividend on the L-share is 24 lower than that of the U-share (in each state). To map this into MM, assume the two companies hold the same real asset. The payoff generated by the real asset is 24, 32, 52, or 84 with equal probability. The U-share can be thought of as the equity share of the unleveraged firm, which gets the entire payoff. The L-share can be thought of as the equity share of the leveraged company, which holds the same asset but that issued debt with value 24 cents (D in MM), and thus its shares get the cash flow minus 24. The MM arbitrage argument states that the market value of the unleveraged company, S_U (the price of the U-share), must be equal to the value of the leveraged company, which is the market value of equity (S_L , the price of the L-share) plus the value of the debt, D :

$$V_U = S_U = S_L + D = V_L. \quad (2)$$

An arbitrage opportunity similar to that analyzed on page 269 of MM (1958) is one in which, in contrast to equation (2), the total market value of the leveraged firm is larger than that of the unleveraged firm, as $S_L + D > S_U$. The immediate, riskless gain from going long the unleveraged U-share and short the leveraged L-share equals the difference between the two, $S_L + D - S_U > 0$. This arbitrage gain results from the fact that there are no margin or repurchase requirements in our experiment, all shares are cancelled at maturity T , and debt D is a synthetic debt stream without real impact.¹⁸ In equilibrium, arbitrage cannot

¹⁸ Debt is a pure accounting stream not noted by experimental subjects. Trading costs and interest rates are zero and the shares need not be repurchased but are simply cash at the end of

exist. The prices of U-share and L-share must therefore differ exactly by the value of debt.

MM (1958) discuss expected future streams and suggest one discount rate for the same risk, but the proof of the theorem requires identical future streams. One key research question that we raise in our experimental control treatment (with zero correlation) is whether empirical validation of the invariance theorem actually requires identical asset returns.¹⁹ Our research question is two-fold:

- (i) Are the prices of two portfolios equal when they have perfectly correlated payoffs (modulo a shift) and when short sales and borrowing are unconstrained?
- (ii) Are the prices of two portfolios equal when their payoffs are uncorrelated but have the same distribution (modulo a shift) and short sales and borrowing are unconstrained?

Question i) is a direct test of the MM (1958) theorem. For question ii), the relevant theoretical result is that (in perfect markets) two portfolios with the same payoffs should have the same equilibrium price (and this holds irrespective of utility functions). This argument can be motivated by the capital asset pricing model as described in the experimental finance literature (Asparouhova, Bossaerts, Plott (2003), Bossaerts, Plott, Zame (2007),

$$\begin{bmatrix} S_{Ut} \\ S_{Lt} \end{bmatrix}^* = \begin{bmatrix} F_{Ut} \\ F_{Lt} \end{bmatrix} - K \begin{bmatrix} \sigma_{Ut}^2 & \rho_{LU} \sigma_{Lt} \sigma_{Ut} \\ \rho_{LU} \sigma_{Lt} \sigma_{Ut} & \sigma_{Lt}^2 \end{bmatrix} \begin{bmatrix} \bar{Z}_U \\ \bar{Z}_L \end{bmatrix}. \quad (3)$$

This formula, adapted here to two risky assets, takes into account market supply and demand. The equilibrium price is determined by the mean holdings of the market portfolio \bar{Z} of risky assets (as indicated in Table I); the expected dividend payoffs of risky assets F_{jt} ($j = U, L$); the (harmonic mean) risk aversion of the market, K ; and the covariance matrix. With equal supply, that is, $\bar{Z}_U = \bar{Z}_L$, equal payoff variance of L-shares and U-shares, that is, $\sigma_L^2 = \sigma_U^2$ and zero correlation, $\rho_{LU} = 0$, the equilibrium price difference equals the difference

the market sequence. The accounting exercise is easy because cash financing and debt financing have the same consequences in the experiment.

¹⁹ MM (1958, p. 266) assume identical risk classes “such that the return on the shares . . . in any given class is proportional to (and hence perfectly correlated with) the return on the shares . . . in the same class . . .” These conditions are satisfied in our perfect-correlation treatment. Conditions are similar in our no-correlation treatment, since the expected returns of assets are the same as in the perfect-correlation treatment. However, returns are independent in the no-correlation treatment so that, in the strict sense of the definition, L-shares and U-shares are not in the same risk class. Our experimental test checks whether the strict implementation of the definition is critical for achieving empirical support for value-invariance. In the laboratory it is not clear whether notable differences can be detected. On the one hand, earlier research has suggested that human behavior may disregard correlations between asset returns (e.g., Kroll, Levy, and Rapoport (1988)). On the other hand, our setting favors pricing at expected dividend value, since subjects are informed about the fundamental dividend values in the instructions, which could reduce confusion of subjects and mispricing of assets (e.g., Kirchler, Huber, and Stöckl (2012)).

in expected dividend payoffs, that is, $S_{U_t}^* - S_{L_t}^* = F_{U_t} - F_{L_t}$, $t = 1, \dots, T$.²⁰ Hence, for both treatments, theoretical arguments are available to expect parity pricing in the market. We consider a perfect (no-fee) capital market setting and ask whether actual human beings conform to the predictions of theory.

Riskless exploitation of price discrepancies is impossible where payoffs are only identically distributed rather than perfectly correlated. So in this case smart traders may require a risk premium to keep relative prices in balance, thereby permitting a larger deviation from parity pricing.²¹ Evidence from real-world markets shows that the relative mispricing of twin shares can occur over an extended horizon when there are limits to arbitrage (e.g., Froot and Dabora (1999)). Shleifer (2000) suggested that in the pricing of twin shares, “noise trader risk” can result in limits to arbitrage. The risk exists that discrepancies may widen instead of narrowing when traders follow price trends that move away from equilibrium.

III. Measures and Hypotheses

A. Measures

We now formulate our measures and testable hypotheses. As above, let F_{U_t} and F_{L_t} denote the fundamental dividend values, and let S_{U_t} and S_{L_t} denote the share prices of the U-share and the L-share in period t , respectively. We measure the relative difference from parity pricing of the U-share and the L-share as follows:

$$\Delta_t = \frac{S_{U_t}}{S_{L_t} + D_t} - 1 = \frac{(S_{U_t} - S_{L_t}) - (F_{U_t} - F_{L_t})}{S_{L_t} + (F_{U_t} - F_{L_t})}. \quad (4)$$

The first ratio in (4) relates the value of the “unleveraged” company U to the value of the “leveraged” company L , where the difference between fundamental dividend values in the second denominator stands for the synthetic value of debt, $D_t = F_{U_t} - F_{L_t}$. The right-hand side in (4) is obtained by replacing the synthetic value of debt. The invariance theorem requires pricing at parity, $\Delta_t = 0$ (see Hypothesis 1). Thus, market values differ by as much as but not more than fundamentals. In our data, market value is the average market price of the period.

²⁰ Note that, in equation (2), the same is true for perfect correlation, $D = F_U - F_L$. The formula generally holds if the agent buys for keeps (not speculatively), but it does not typically extend to other utility functions (unless payoffs are Gaussian); variance is not the right measure of risk in general. However, the CAPM formula provides a valid first-order approximation to prices if total risk is small (see the proof in Judd and Guu (2001)).

²¹ We note that markets are incomplete in our no-correlation treatment, since a self-financing replicating portfolio cannot be formed. Therefore, in incomplete markets, (riskless) arbitrage is generally impossible. However, if market prices confirmed fundamentals or if deviations from fundamentals were the same in both of our markets, no cross-asset price discrepancy would arise and the law of one price would still obtain.

We define the measure cross-asset price discrepancy, PD , as follows, where $T = 10$ is the number of market periods:

$$PD = \frac{1}{T} \sum_t |\Delta_t|. \quad (5)$$

The PD is the average absolute percentage deviation from parity pricing during the course of a market. A price discrepancy indicates potential gains by selling high and buying low based on average prices. This value is usually positive even if the average difference from parity pricing is zero. The PD is zero if L-share and U-share prices are equal to fundamental dividend values, or if prices deviate by the same amount from fundamental dividend values.

Even with zero PD , prices can deviate from fundamental value. To measure price deviations from fundamentals, we define the relative absolute deviation from fundamentals, DF . The DF -value here measures the expected excess return of buying and selling at prices off the known fundamental dividend value ($j = 1, 2$ indicates L-shares and U-shares, $J = 2$):

$$DF_{jt} = \left| \frac{S_{jt}}{F_{jt}} - 1 \right|; \quad DF = \frac{1}{JT} \sum_{j=1}^J \sum_{t=1}^T DF_{jt}. \quad (6)$$

The DF represents expected gains by purchasing one share if the price is below or selling one share if the price is above the fundamental dividend value. It compares the price to the predicted value under the equilibrium hypothesis, whereas the PD compares the price to the predicted value under the no-discrepancy hypothesis. By comparing the two, we can examine whether mispricing is more severe with regard to dividend value or with regard to the other asset. Note that DF and PD can be compared, which allows us to assess whether the dividend value or the arbitrage opportunity drives the equilibration dynamics in our experiment. Similar measures for mispricing have been proposed in the literature for the single-asset market (e.g, Stöckl, Huber, and Kirchler (2010)). In particular, the relative deviation RD of asset $j = L, U$, where F_j is the average cumulative dividend value, has been used to measure the magnitudes of bubbles in experiments:

$$\text{Relative deviation} = RD_j = \frac{1}{TF_j} \sum_{t=1}^T S_{jt} - F_{jt}. \quad (7)$$

Potential discrepancies can arise in real time in submitted orders. If the adjusted bid exceeds the best outstanding offer of the twin share, an arbitrageur can sell high and buy low to realize an expected gain. Under perfect correlation, the gain from eliminating a price discrepancy is riskless (arbitrage); under no-correlation, the expected gain is risky, since differences in future dividends can lead to losses or gains. We refer to as discrepant each limit order that (upon submission) leads to a cross-asset discrepancy in outstanding orders,

that is, a potential expected gain for another trader.²² We measure the relative frequency of discrepant limit order submission ($RFDLO_i$) for each subject i and then aggregate as follows (where $\#LO_i$ denotes the number of limit orders submitted by the subject, and $\#DLO_i$ denotes the number of discrepant limit orders):

$$RFDLO_i = \frac{\#DLO_i}{\#LO_i}; \quad RFDLO = \frac{\sum_i \#DLO_i}{\sum_i \#LO_i}. \quad (8)$$

We measure a trader's aptitude as the reverse of the frequency of submitting a discrepant limit order, $1 - RFDLO_i$. The trader's aptitude is therefore high if $RFDLO_i$ is small and vice versa.

Our empirical measure of acuity is the individual's *CRT* score, that is, the number of correct answers in the pre-market cognitive reflection task (Frederick (2005)). Recent results in single-asset market experiments point to systematic effects of *CRT* scores (Corgnet et al. (2015), Akiyama, Hanaki, and Ishikawa (2017), Breaban and Noussair (2015), Noussair, Tucker, and Yu (2016)). Below, we also measure the effect of gender and risk aversion (as elicited in the investment game) on our mispricing measures (4) to (8). Recent research from experimental single-asset markets suggests that price levels may be lower if the level of risk aversion increases (Breaban and Noussair (2015)) and also if the share of female traders increases (Eckel and Füllbrunn (2015)).

B. Testable Hypotheses

With measures (4) to (8) at hand, we formulate our testable hypotheses. Our most efficient theoretical benchmark (which is standard in experimental studies that apply the Smith, Suchanek, and Williams (1988) design, but is unlikely to obtain with inexperienced subjects in the laboratory) requires that prices be equal to risk-neutral fundamental dividend values.

HYPOTHESIS 0 (Risk-neutral pricing): *There are no deviations of prices from fundamentals in either share class, that is, $DF = 0$.*

In line with the literature, we also look at a weak form of Hypothesis 0, that is, we investigate whether the *RD* measure is equal to zero on average. Risk-neutral pricing would require that the price equal the fundamental dividend value in each period. In fact, risk-neutral pricing is a sufficient (but not necessary) condition for obtaining MM invariance (MM (1958)). Indeed, previous experimental evidence (Palan (2013)) shows that pricing deviations from fundamental dividend value can be expected to occur in experimental asset markets both from above and below. This evidence does not invalidate the MM law of arbitrage-free pricing. For arbitrage-free pricing and the MM theorem,

²² In a different trading environment, Biais and Pouget (1999, p. 15) refer to similar trades as “a noise trading [cf. Shleifer and Summers 1990] component in the order flow ... as they [tend] ... to lose money.”

we require that our L-share and U-share be priced at parity, which is indeed a much weaker requirement than risk-neutral equilibrium pricing.

HYPOTHESIS 1A (MM invariance theorem): *The adjusted market values of L-shares and U-shares will be the same, i.e., $\Delta_t = 0$.*

HYPOTHESIS 1B (No cross-asset price discrepancy): $PD = 0$.

HYPOTHESIS 1C (No discrepant orders): $RFDLO = 0$.

Hypothesis 1A requires that an L-share and a U-share (adjusted for fundamentals) be priced the same on average. Hypothesis 1B (which implies 1A) requires that eventual deviations from fundamentals occur simultaneously and equally in both shares. Hence, we investigate three levels of market efficiency: the first is based on the absence of differences from parity pricing on average (Hypothesis 1A), the second is based on the absence of price discrepancies (Hypothesis 1B) and discrepant orders (Hypothesis 1C), and the third requires that prices coincide with risk-neutral dividend values (Hypothesis 0).

These levels of market efficiency are tested in two treatment dimensions. First, we consider perfect correlation, where elimination of pricing discrepancies is riskless because dividend streams differ by a constant and can be arbitrated. Second, we consider the case of no correlation, where elimination of pricing discrepancies is risky because dividends are independent of each other. Since the L-share and the U-share have identical idiosyncratic risks in both treatments, and based on earlier evidence that suggests insensitivity of behavior to changes in return correlations (e.g., Kroll, Levy, and Rapoport (1988)), one might expect no treatment effect in the laboratory. However, the MM arbitrage result concerns only the relative pricing of two assets, not the absolute level of their prices. Thus, it could be the case that both assets are mispriced in absolute terms but their relative pricing is aligned, so there is no arbitrage opportunity. Given that mispricing is frequently observed in the experimental design we consider, we expect to observe some mispricing in our experiments, which leads to the following testable hypothesis.

HYPOTHESIS 2 (Hedging effect): *The cross-asset price discrepancy (PD) is larger in the no-correlation treatment than in the perfect-correlation treatment.*

The above hypothesis suggests that markets can eliminate some price discrepancies in the absence of risk even without a dedicated (automatic) arbitrageur, and so decrease the magnitude of PD in the presence of perfect correlation. Hypothesis 2 points to the relative validity of MM invariance under perfect correlation, relative to no correlation.

Based on experimental evidence of mispricing in early markets (e.g., Smith, Suchanek, and Williams (1988) and Dufwenberg, Lindqvist, and Moore (2005)), we expect a decrease in mispricing in our consecutive markets.

HYPOTHESIS 3 (Experience effect): *The cross-asset price discrepancy, PD , deviation from fundamental dividend value, DF , and relative frequency of discrepant orders, $RFDLO$, decrease over time (in consecutive markets).*

An oft-cited conjecture holds that smart investors eliminate mispricing in the market (see, e.g., Shleifer (2000)). Although our design has no dedicated arbitrageur, we can nevertheless address the predictive power of this argument in our data. We compare the magnitude of price discrepancies vis-à-vis characteristics of market participants that suggest a relationship between the acuity and aptitude of traders.

HYPOTHESIS 4 (Smart trader effect): *The cross-asset price discrepancy (PD), deviation from fundamental dividend value (DF), and the relative frequency of discrepant orders (RFDLO) will decrease with the market participants' degree of measured acuity.*

We are also interested in the equilibration dynamics of the market. When a price discrepancy arises, that is, when the L-share and the U-share are mispriced relative to one another, at least one share (if not both) are mispriced vis-à-vis the fundamental dividend value. Exploitation of this mispricing leads to expected gains. We expect that the more profitable trade will happen first in both treatments. The more profitable trade is the one whose expected return is the larger one within the discrepancy pair. It can be either a long or a short position. Moreover, we also expect, especially in the perfect-correlation treatment (but also in the no-correlation treatment), that subjects exploit price discrepancies by simultaneously taking long and short positions, thus eliminating the discrepancy and booking an arbitrage (expected) gain. The latter behavior is what MM suggest in their arbitrage proof.

HYPOTHESIS 5A (Equilibration dynamics): *Where cross-asset price discrepancy arises, the higher expected return twin share trades first.*

HYPOTHESIS 5B (Asset swap/MM arbitrage): *A subject exploits the price discrepancy through (almost) simultaneous trade in the twin shares.*

Finally, we investigate the effect of return correlation on individual portfolio diversification. Portfolio diversification reduces fluctuations in portfolio returns when correlations are smaller. Financial economics suggests that people are averse to return volatility. We therefore conjecture that the difference between the number of U-shares and L-shares is smaller in the no-correlation treatment than in the perfect-correlation treatment.²³ Let Z_{Uit} and Z_{Lit} be the number of U-shares and L-shares in the portfolio of subject i at the end of period t , and $dZ_{it} = |Z_{Uit} - Z_{Lit}|$ be the absolute difference between the two numbers. We compare the average individual portfolio diversification over subjects and periods, which we denote by $dZ(\rho_{LU})$, across markets and treatments.

HYPOTHESIS 6 (Individual portfolio diversification): *Subjects are more diversified in the no-correlation treatment (0) than in the perfect-correlation treatment (1), $dZ(0) \leq dZ(1)$.*

²³ Our measure of diversification is very simple. We look at individual deviations from the market portfolio. Investor subjects are perfectly diversified in our design if they hold an equal number of L-shares and U-shares. Different from the no-correlation treatment, return volatility does not decrease with increased diversification in the perfect-correlation treatment.

IV. Results

In total, 174 students participated, earning an average of \$50 in three hours. Each subject participated in only one session. We have 12 sessions in the perfect-correlation treatment and eight sessions in the no-correlation treatment.²⁴ Each session comprises four consecutive markets.²⁵

A. MM Invariance Theorem

OBSERVATION 1: *L-share and U-share prices are close to parity and in line with the MM invariance theorem in the perfect-correlation treatment. U-share prices are above parity in the no-correlation treatment.*

Support. In Figure 1, we see the difference in L-share and U-share prices relative to parity for each market, period, and treatment, aggregated separately over the 12 sessions of the perfect-correlation treatment and the eight sessions of the no-correlation treatment. In the case of a missing L-share or U-share price, the period in the session is treated as a missing observation.²⁶ The prices in the perfect-correlation treatment are close to parity, whereas the prices in the no-correlation treatment are above parity. In view of our theoretical discussion, we note that no direction of deviation could be predicted. In contrast, the data suggest that the deviation of relative market valuation has a direction, indicating a relatively higher price for the U-share than the L-share.²⁷ Below we investigate this pattern further.

Table II reports the average differences from parity for each market in both treatments together with the z -statistics of the two-tailed Wilcoxon one-sample test, which indicates significant deviations from zero. The average results for the perfect-correlation treatment are not significantly different from parity in any market, $\Delta_t = 0$; the overall average price difference from parity per period is 3.67%. By period, we find that only 3 out of 40 ($= 4 \text{ markets} \times 10 \text{ periods}$)

²⁴ We do not have the same number of sessions across treatments. Unfortunately, we discovered a problem in the software after conducting the first four sessions. Owing to this problem, the L-share holders received no cash dividend; instead, both L-share and U-share dividends were distributed to U-share holders. Due to logistical and timing problems, we were unable to book the lab to re-run these four sessions and so have only 8 cohorts in the no-correlation treatment rather than the planned 12 cohorts.

²⁵ In the fourth market of one session of the perfect-correlation treatment, one subject submitted an obviously erroneous bid (instead of an asking price) for \$20 on the L-share worth \$0.24 that led to a transaction. We eliminated this period from the data since the transaction impacts our average estimates. In the second market of the two other perfect-correlation sessions, the data for period 10 are missing due to a server crash. One session of the perfect-correlation treatment and two sessions of the no-correlation treatments had only seven participants.

²⁶ On average we have 11% and 5% missing observations in the perfect-correlation and no-correlation treatments, respectively.

²⁷ In this context the dividend puzzle comes to mind (DeAngelo and DeAngelo (2006)), that is, the empirical observation that the market value of dividend-paying stock is higher than of zero dividend stock. Whereas the U-share always pays a positive dividend, the L-share indeed pays a zero dividend 25% of the time.

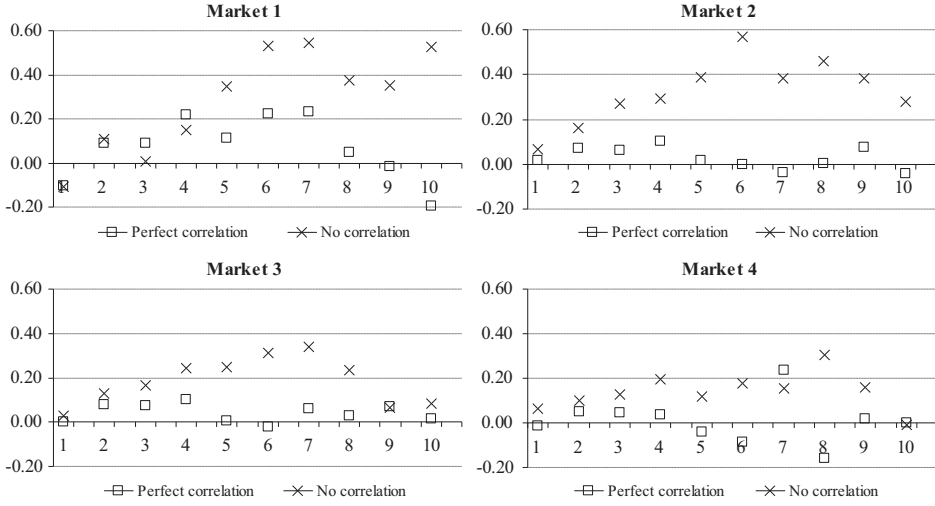


Figure 1. Evolution of the relative difference from price parity of L-shares and U-shares Δ_t from Market 1 to Market 4. The abscissa shows periods 1 to 10 and represents price parity between L-share and U-share prices. The ordinate indicates the relative difference from parity from -0.2 to 0.6 . Each chart represents the relative difference from price parity for one market of 10 periods.

Table II
Average Difference from Parity Δ in Percent

The table reports the difference from parity (equation (4)) averaged over 10 periods and all sessions for each market and both treatments. The first column reports the averages for the perfect-correlation treatment and the second column for the no-correlation treatment. The bottom line reports the average difference from parity over all periods. Asterisks indicate significant differences from parity measured by two-tailed, one-sample Wilcoxon signed-ranks tests. The last column reports p -values resulting from two-tailed, two-sample Mann-Whitney tests; z -statistics are in brackets. Significant differences are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$.

Treatment	Perfect correlation ($n = 12$)	No correlation ($n = 8$)	Mann-Whitney test (between treatments), p -value, [z -statistic]
Market 1	6.98 [0.39]	27.72** [2.24]	0.031** [2.16]
Market 2	1.72 [0.04]	29.92** [2.52]	0.001*** [3.28]
Market 3	3.49 [1.49]	17.94** [2.10]	0.165 [1.39]
Market 4	2.62 [1.26]	14.49** [2.52]	0.064* [1.85]
Average	3.67 [1.53]	22.52** [2.52]	0.006*** [2.78]

differences from parity in the perfect-correlation treatment are significant at the 5% level; the pricier share is once the L-share and twice the U-share.²⁸ In the no-correlation treatment, parity is rejected for each market since the U-share is always pricier than the L-share (see Table II); the overall average price difference from parity per period is 22.52% in the no-correlation treatment, but only 3.67% in the perfect-correlation treatment.²⁹

So, the differences from price parity are much smaller in the perfect-correlation treatment than in the no-correlation treatment. These differences are 4 to 15 times larger in the no-correlation treatment in the four markets. Overall, the average difference from parity per period across the treatments is highly significant, with the p -value of the two-tailed Mann-Whitney test equal to 0.006. We conclude that our data are in line with the MM theorem (and thus with Hypothesis 1A) in the perfect-correlation treatment, but reject Hypothesis 1A for the no-correlation treatment.

The results indicate that riskless exploitation of discrepancies, as in the perfect-correlation treatment, is required to approximate value-invariance. In the no-correlation treatment the discrepancy thus seems to have two complementary potential sources. First, the exploitation of discrepancies is risky, since future dividend streams do not allow riskless arbitrage. The risky exploitation of a discrepancy requires a risk premium, which implies larger discrepancies. Second, the market discounts the U-share less than the L-share. In principle, this could indicate the existence of a certainty effect in the data (Tversky and Kahneman (1986)) concerning non-zero payoffs: whereas the U-share dividend is strictly positive by definition, the L-share dividend can be zero in every period. However, our follow-up treatments below do not find supportive evidence, and so the reason for this behavior is not completely clear to us.

B. Pricing Discrepancies and Deviation from Fundamentals

OBSERVATION 2: *The absolute difference from parity, PD, deviation from fundamentals, DF, and relative frequency of discrepant orders, RFDLO, are reduced in the perfect-correlation treatment relative to the no-correlation treatment.*

Support. The data in Table III, Panels A, B, and C, record the averages of PD, DF, and RFDLO by treatment, respectively.³⁰ From Table III, Panel A we

²⁸ The L-share is significantly pricier than the U-share in the first period of the first market ($p = 0.041$), and the U-share is pricier in period 10 of the first market ($p = 0.041$) and in period 7 of the third market ($p = 0.039$).

²⁹ Since outliers could be affecting the averages, we redid the analyses with median prices instead of simple average prices. The results are very similar and are reported in Tables IA.I, IA.IIA, and IA.IIB in the Online Appendix.

³⁰ These values have been computed according to equations (5), (6), and (8). A missing average price of L-share or U-share in a certain period is treated as a missing observation in the corresponding session. The minimum and the first quartile PD overall is 5.0% and 9.6% in the perfect-correlation treatment and 8% and 12.8% in the no-correlation treatment, respectively.

Table III
Average Absolute Deviation from Parity, *PD*; Average Absolute Deviation from Fundamentals, *DF*; and Average Relative Frequency of Discrepant Limit Orders *RFDLO* in Percent

Panel A reports the price discrepancy (equation (5)) for the 10 market periods averaged across all sessions for each market and treatment. Panel B reports the deviation from fundamentals (equation (6)) for the 10 market periods averaged across all sessions for each market and treatment. Panel C reports the frequency of discrepant limit orders relative to submitted limit orders (equation (8)) by session for each market and treatment. The bottom lines show the average price discrepancy across all periods, and the one-tailed Page trend test, which checks for a significant decline in repeated markets. The last column reports *p*-values resulting from two-tailed, two-sample Mann-Whitney tests; *z*-statistics are in brackets. Significant differences are indicated as follows: ****p* < 0.01, ***p* < 0.05, **p* < 0.10. ^aOne-tailed tests.

Panel A. <i>PD</i> in Percent			
	Perfect correlation ($n_1 = 12$)	No correlation ($n_2 = 8$)	Mann-Whitney test (between treatments), <i>p</i> -value, [<i>z</i> -statistic] ^a
Market 1	22.92	39.76	0.124 [1.16]
Market 2	13.69	30.09	0.014** [2.20]
Market 3	13.82	23.31	0.198 [0.85]
Market 4	15.14	20.37	0.284 [0.58]
Total	16.39	29.13	0.082* [1.39]
<i>p</i> -value (one-tailed Page trend test) ^a	0.159 [−1.00]	0.011** [−2.30]	
Panel B. <i>DF</i> in Percent			
	Perfect correlation ($n_1 = 12$)	No correlation ($n_2 = 8$)	Mann-Whitney test (between treatments), <i>p</i> -value [<i>z</i> -statistic]
Market 1	24.59	39.52	0.320 [1.00]
Market 2	15.25	30.36	0.045** [2.01]
Market 3	15.87	22.34	0.320 [1.00]
Market 4	21.65	17.35	0.758 [−0.31]
Total	19.34	27.39	0.563 [0.58]
<i>p</i> -value (one-tailed Page trend test) ^a	0.159 [−1.00]	0.001*** [−3.30]	

(Continued)

Table III—Continued

Panel C. <i>RFDLO</i> in Percent			
	Perfect correlation ($n_1 = 12$)	No correlation ($n_2 = 8$)	Mann-Whitney test (between treatments), p -value, [z -statistic]
Market 1	6.76	10.90	0.190 [1.31]
Market 2	4.96	10.58	0.123 [1.54]
Market 3	4.81	9.61	0.105 [1.62]
Market 4	5.51	9.81	0.037** [2.08]
Total	5.51	10.23	0.045** [2.01]
p -value (one-tailed Page trend test) ^a	0.242 [−0.70]	0.460 [−0.10]	

see that, by buying the lower-priced share and selling the higher-priced twin share (both at the average price in a randomly determined period), the average riskless return would be 22.92% of the trading value in the first market of the perfect-correlation treatment. This number is relatively large if compared to the second market, where the average *PD* drops to one-half of the first market.³¹ However, the amount of the decrease is rather small relative to the first market in the no-correlation treatment. As indicated by the one-tailed Mann-Whitney test reported in the table, the *PD* is larger overall (at the 10% significance level) than in the no-correlation treatment.

We see in Panel B of Table III that the results for the *DF* point in the same direction. The overall average absolute deviation from the fundamental dividend value is 19.34% in the perfect-correlation treatment and 27.39% in the no-correlation treatment. Nevertheless, the differences between the two treatments are significant (at the 5% level) only for the second market. The results suggest that pricing relative to fundamentals is (relatively) independent of whether there are perfect arbitrage conditions. In line with the literature, both treatments trend towards the fundamental dividend value with repetition. We conclude that the data support rather than reject Hypothesis 2 and Observation 2 vis-à-vis the *PD* and *DF*, particularly for Market 2, although the evidence is not statistically significant for all markets.

Panel C of Table III reports the average number of discrepant orders relative to the submitted orders in a market. The overall average number of discrepant limit orders is 5.51% in the perfect-correlation treatment and 10.23% in the no-correlation treatment. The differences are significant across treatments.

³¹ The decrease from the first to the second market is significant at the 1% level ($Z = 2.353$, $p = 0.009$, one-tailed Mann-Whitney test, $n = 12$).

In terms of absolute numbers of arbitrage and mispricing, Hypothesis 0 and Hypotheses 1B and 1C must be rejected given that no session and market involves zero DF , zero PD , or zero $RFDLO$.

OBSERVATION 3: *The consecutive-markets analysis suggests reduced price discrepancies with experienced subjects, which occurs earlier in the perfect-correlation treatment than in the no-correlation treatment. We observe no change in the relative frequency of discrepant limit orders.*

Support. The data in Panels A and B of Table III (see also Panels A and B of Table IA.II in the Internet Appendix) indicate a decrease in price discrepancies in the first three consecutive markets, but not always from the third to the fourth market. The nonparametric Page trend test ($n = 12$, $k = 4$) as reported in the bottom row of Panel A of Table III suggests no significant decrease of the PD over four markets in the perfect-correlation treatment.³² In contrast, the results reported in Panel B of Table III indicate a decrease for the deviation from fundamentals that is significant at the 10% level. The results of the one-tailed Page test are significant at the 5% level for the no-correlation treatment, indicating a reduction over time in the values of PD and DF .

We find no significant evidence for continuous convergence to the theoretical predictions over all four markets in the perfect-correlation treatment, since convergence seems to occur all at once from the first to the second market, after which deviations remain at the same level until the end. Thus, the perfect-correlation treatment exhibits a significant decline at the 5% level in the values of PD and DF between the first and the second markets, but not between any other consecutive markets.³³ In contrast, the main decline in the no-correlation treatment occurs from Market 2 to Market 3.³⁴ Thus, the data suggest a more rapid adjustment to the theoretical benchmarks in the perfect-correlation treatment, that is, when exploiting price discrepancies is riskless.

Panel C of Table III suggests no decline in $RFDLO$ by the one-tailed Page test.

OBSERVATION 4: *The differences between PD and DF suggest that pricing discrepancies are a more focal driver of behavior than the deviation from fundamentals in the perfect-correlation treatment when compared to the no-correlation treatment.*

Support. Comparing Panels A and B of Table III, the values of PD are smaller than the values of DF for each market of the perfect-correlation treatment,

³² In Market 4 in some sessions, subjects pushed prices to irrationally high levels in late periods. Even if these data points are ignored, however, the price discrepancies would not be reduced from the level of Market 3. That said, the reported p -value of the Page test suggests that a declining PD in the perfect-correlation treatment would then be significant at the 5% level.

³³ The one-tailed Wilcoxon signed ranks tests that suggest a decline between Markets 1 and 2 are as follows: PD ($p = 0.010$, $Z = 2.316$), and DF ($p = 0.009$, $Z = 2.353$).

³⁴ The PD ($p = 0.034$, $Z = 1.820$) and the DF ($p = 0.034$, $Z = 1.820$) decrease significantly at the 5% significance level between Market 2 and Market 3, but not between Market 1 and Market 2 ($p = 0.116$, $Z = 1.193$), ($p = 0.104$, $Z = 1.260$).

whereas they are larger for each market of the no-correlation treatment.³⁵ The average difference between PD and DF over all markets is -0.029 in the perfect-correlation treatment and 0.017 in the no-correlation treatment. The difference between treatments is significant—according to the one-tailed Mann-Whitney two-sample test, the p -value is 0.071 . Hence, the data suggest that pricing vis-à-vis the other asset is focal in the perfect-correlation treatment compared with the no-correlation treatment. The results reported in part E of Section IV, where we examine the equilibrating dynamics, underline this effect.

C. Bubble Magnitude

Following the standard line of experimental asset market research, we report the standard measure of bubble magnitude RD (relative deviation) for our treatments by asset. Given that the literature on the single-asset market shows that asset mispricing occurs frequently (Palan (2013)), we have a surprising result.

OBSERVATION 5: *a) Asset prices average close to the dividend value in the perfect-correlation treatment. The RD measures of the L-share and the U-share are near zero. b) The price of the U-share in the no-correlation treatment exceeds the dividend value, that is, RD is significantly positive, while the average price of the L-share is close to the expected dividend value.*

Support. Figures 2 and 3 show the average price trajectories for all sessions and overall. Table IV reports for each market (by treatment) the standard measure of mispricing—the relative deviation RD measure—for the L-share and the U-share (and their differences) as described in Section III and as commonly applied in the literature (Stöckl, Huber and Kirchler (2010)). The figures illustrate observations 5a and 5b. The overall average RD measures are -0.041 and 0.171 for the L-share and the U-share in the no-correlation treatment and -0.025 and 0.023 in the perfect-correlation treatment, respectively. The deviations from expected cumulative dividend value are statistically significant for the U-share in the no-correlation treatment, where RD is significantly different from (i.e., larger than) zero both for each market and overall. In the perfect-correlation treatment, the overall RD measure is not significantly different from zero; a significant (negative) RD is only reported for the L-share in Market 3.

Comparing across assets, we find that the RD values of the L-share and the U-share are significantly different in the no-correlation treatment, while in the perfect-correlation treatment there are no significant differences between RD values of the L-share and the U-share.³⁶ If we compare RD val-

³⁵ On average across all markets, the difference between PD and DF is significantly different from zero in the perfect-correlation treatment. Per the two-tailed Wilcoxon signed ranks test, the p -value is 0.0995 . In the no-correlation treatment we have no significant differences between PD and DF .

³⁶ Based on the two-tailed Wilcoxon signed ranks test, we reject the null hypothesis of equal RD measures between L-share and U-share for the no-correlation treatment; the p -value is 0.012 . In

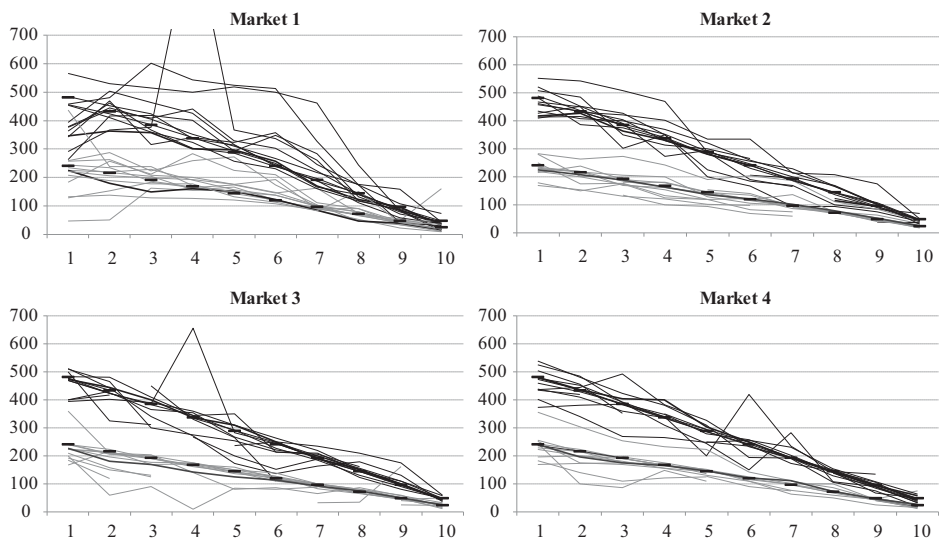


Figure 2. Average prices in Markets 1 to 4 of the perfect-correlation treatment. The trajectories of all 12 cohorts are displayed for both L-share and U-share assets, and the average for each asset is indicated in bold. The dividend values are represented by cascading short horizontal lines, but here they are almost covered by the lines that indicate the overall averages. The ordinate indicates the average prices from 0 to 700, and the abscissa the periods 1 to 10.

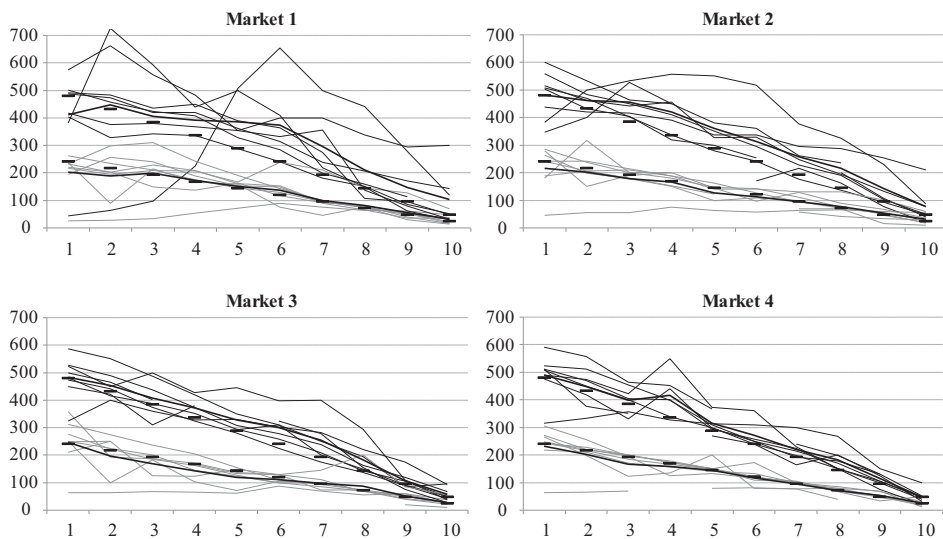


Figure 3. Average prices in Markets 1 to 4 of the no-correlation treatment. The trajectories of all eight cohorts are displayed for both L-share and U-share assets, and the average for each asset is indicated in bold. The dividend values are represented by cascading short horizontal lines, which are well visible for the U-share, but are almost covered for the L-share by the bold line that indicates the overall average. The ordinate indicates the average prices from 0 to 700, and the abscissa the periods 1 to 10.

Table IV
Average *RD* Measure and Difference between Shares in Percent

The table reports the average relative deviation (equation (7)) for each market, assets L-share and U-share, and treatment. The bottom row reports the average across all markets. The third and sixth columns report the differences between *RD* values of assets L-share and U-share. The *RD* values and their differences between assets L-share and U-share are tested, *z*-statistics are in brackets. Significant differences from zero resulting from the two-tailed, one-sample Wilcoxon signed ranks test are indicated as follows: ****p* < 0.01, ***p* < 0.05, **p* < 0.10.

	Perfect correlation ($n_I = 12$)			No correlation ($n_O = 8$)		
	(1) RD_L	(2) RD_U	(3) $RD_U - RD_L$	(4) RD_L	(5) RD_U	(6) $RD_U - RD_L$
Market 1	2.80 [0.31]	9.48 [1.02]	6.68 [0.16]	0.24 [0.56]	22.13** [2.38]	21.90* [1.68]
Market 2	-5.53 [-1.10]	-0.61 [-0.39]	4.92 [1.02]	-5.03 [0.14]	22.49*** [2.52]	27.52*** [2.52]
Market 3	-11.00** [-2.28]	-2.19 [-0.83]	8.82** [2.28]	-5.91 [-0.21]	12.57** [2.24]	18.48* [1.40]
Market 4	3.68 [-0.24]	2.44 [0.39]	-1.24 [0.24]	-5.54 [0.14]	11.40** [2.38]	16.94** [2.38]
Total avg.	-2.51 [-0.78]	2.28 [1.26]	4.79 [0.94]	-4.06 [0.14]	17.15** [2.52]	21.21*** [2.52]

ues between treatments and shares, we find that only the *RD* of the U-share in the no-correlation treatment is significantly different from (i.e., larger than) the others.³⁷ This misbehavior of the U-share is puzzling and seems to drive the reported violation of parity pricing in the no-correlation treatment. The Spearman rank correlation between RD_U and Δ is 0.168 and 0.810 in the perfect-correlation and the no-correlation treatments, respectively; the latter indicates a significant correlation ($p = 0.015$) whereas the former does not ($p = 0.602$).³⁸ The question here is why the deviations occur in the U-share of the no-correlation treatment but not in the L-share. Below we investigate this finding further by looking at both the market behavior and the valuations of the L-share and the U-share in the single-asset market.

D. Smart Traders

OBSERVATION 6: *Individual performance (i.e., payoff) correlates with measured acuity.*

the perfect-correlation treatment, the *p*-value is 0.347 and so we conclude that the *RD* measures do not significantly differ across assets.

³⁷ The *p*-values of the two-tailed Mann-Whitney two-sample test (on the total average) are as follows: between the U-shares ($p = 0.000$), between the L-shares ($p = 0.227$), and between the L-shares and U-shares (0.007).

³⁸ Note also that the differences in *RD* between L-share and U-share and deviations from parity are highly correlated. The Spearman rank correlation coefficient is 0.862 in the overall sample, which is significant at any commonly accepted significance level (*p*-value is 0.000).

Table V
Regression Results and Spearman Rank Correlation Coefficient with CRT Score

In Panel A, the first two columns report the outcomes of the cluster regression with robust standard errors of the within-group individual payoff and aptitude (the reciprocal of the relative frequency of a limit order that leads to an arbitrage opportunity) on the individual *CRT* score (equation (8)), risk aversion, and female gender dummy. The other columns record the OLS regression results of measures (4) to (8) on the group average of the individual measures. Results of two-tailed *t*-tests are indicated [z-statistics in brackets] as follows: *** $p < 0.01$, ** $p < 0.05$, * $p \leq 0.10$. In panel B, correlation coefficients are reported for *CRT* effects of measures (4) to (8) for each treatment separately. Results of two-tailed Spearman rank correlation tests are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p \leq 0.10$.

Panel A. Regression results							
Variable	$Payoff_i$	$1 - RFDLO_i$	Δ	PD	DF	$RFDLO$	RD_U
Intercept	2717.9*** [22.8]	0.976*** [138.9]	0.925** [0.02]	0.949* [2.04]	0.811** [2.73]	0.345*** [3.42]	0.514** [2.77]
<i>CRT</i> score	161.8*** [3.67]	0.005** [2.07]	-0.643*** [-3.84]	-0.69*** [-3.10]	-0.579*** [-4.06]	-0.142*** [-2.98]	-0.30*** [-3.34]
Female	-241.0*** [-2.90]	0.008 [1.31]	-0.198 [-1.20]	-0.164 [-0.75]	0.114 [0.81]	-0.086* [-1.82]	-0.058 [-0.66]
Risk aversion	-0.76 [-0.59]	-0.000 [-0.52]	0.001 [0.12]	0.002 [0.29]	-0.000 [-0.09]	-0.001 [-0.75]	-0.000 [-0.12]
Perfect-correlation treatment dummy			-0.824*** [-3.10]	-0.81** [-2.29]	-0.506** [-2.23]	-0.155* [-2.04]	-0.369** [-2.60]
<i>CRT</i> -score \times treatment dummy			0.646** [2.42]	0.697* [1.95]	0.427* [1.87]	0.102 [1.35]	0.218 [1.53]
#obs.	174	174	20	20	20	20	20
#clusters	20	20					
R^2	.125	.039	.653	.484	.640	.591	.728
Panel B. Correlation with <i>CRT</i>							
	Δ	PD	DF	$RFDLO$	RD_U		
Perfect-correlation treatment, $n = 12$	-0.236	-0.271	-0.236	-0.268	-0.168		
No-correlation treatment, $n = 8$	-0.732**	-0.634*	-0.634*	-0.683*	-0.927***		

Support. Panel A of Table V reports the regression results. We use the subject's average payoff over the four markets as the individual's performance measure. As stated above, we measure individual acuity by the *CRT* score, that is, the individual's number of correct answers. The average *CRT* score in our sample was 0.97.³⁹ As a risk-aversion measure, we use the individual

³⁹ This is a bit lower than average of 1.24 (see Frederick (2005), who finds an average in a very large sample). One might have expected a better result here for *CRT* because of incentivization. The

percentage allocation to the safe asset in the investment game (the average allocation was 60%). Finally, we consider gender, assigning a value of one for a female investor and zero otherwise. The share of females is 49% in our subject pool (see subject pool composition details in Table IA.III in the Internet Appendix).⁴⁰

The outcomes of the OLS regression of the individual performance measure on the *CRT* score, female, and risk aversion are reported in the first column of Panel A of Table V. The OLS regressions use clustered standard errors (clustered by independent session). As indicated, the *CRT* score is a significant determinant of an individual's performance in this regression, but gender and the risk-aversion measure, are not significant. We find no significant differences across treatments.⁴¹ The evidence clearly supports Observation 6.

Instead of measuring performance by payoff as in Observation 6, we can measure performance by the number of bad decisions in trading. As explained above, we compute the ratio of the number of discrepant limit orders to total limit orders, $RFDLO_i$. We then compare this ratio to the individual's *CRT* score.

OBSERVATION 7: *Traders' aptitudes are correlated with their measured acuities.*

Support. The trader's aptitude is inversely related to the relative frequency of submissions of discrepant limit orders to the trader's total number of submitted limit orders, $1 - RFDLO_i$. As characterized above, a discrepant limit order offers an arbitrageur an opportunity to realize an expected gain by swapping assets. The measured trader's aptitude is regressed on individual characteristics. The second column of Panel A of Table V reports OLS regressions with clustered standard errors (clustered by independent cohort). We find that individual acuity as measured by the *CRT* score explains much of the bidding data, in line with Observation 7.

OBSERVATION 8: *When there is high overall trader acuity, price discrepancies are smaller, fewer discrepant orders are submitted, and the price of assets is closer to parity.*

Support. In Panel A of Table V we also report four OLS regressions of our measures (4), (5), (6), and (8) on the group aggregate personal measures. In these regressions the average *CRT* score always shows up significantly. The difference of price from parity, deviation from fundamentals, and price discrepancy as well as the relative frequency of discrepant orders decrease in markets with overall higher trader acuity. In the regression, we also control for

lower number could have resulted from the fact that our subjects were time-constrained whereas Frederick's were not.

⁴⁰ As in many previous studies (see Charness and Gneezy (2012)), we find that females are significantly more financially risk-averse than males ($Z = 2.366$, $p = 0.018$, two-tailed test). In line with the results of Frederick (2005), we also find that males in our sample have significantly higher *CRT* scores ($Z = 3.838$, $p < 0.001$, two-tailed test).

⁴¹ The two-tailed Mann-Whitney test between treatments returns the following p -values on the individual-level data: 0.559, risk aversion, 0.809, female, and 0.987, *CRT* score, and similar p -values obtain in the session-level data, 0.165, risk aversion, 0.754, female, and 0.529, *CRT* score.

treatment effects of *CRT*. As the interaction with the treatment dummy indicates, the *CRT* effect is strong for the no-correlation treatment. In Panel B of Table V we report Spearman rank-correlation coefficients of the *CRT* score with our measures of market efficiency. The coefficients also hint at a treatment effect. The correlation coefficients usually have the same sign in both treatments, but the magnitude of the coefficients is larger for the no-correlation treatment. The results indicate that the reported *CRT* effects could be driven by the no-correlation treatment. We therefore conclude that, in line with Hypothesis 4 and Observation 8, acuity could actually affect the reduction of price discrepancy, but potential discrepancies do not disappear completely.

OBSERVATION 9: *When there is low overall trader acuity, or low trader aptitude, U-share prices exceed fundamentals in the no-correlation treatment. There is no significant impact of trader acuity, risk aversion, or gender composition with the L-share in the no-correlation treatment or either share in the perfect-correlation treatment.*

Support. In the last column of Panel A of Table V, we report results of a simple OLS regression of our acuity measure on the group aggregate personal measures for the no-correlation treatment. The results show that *RD* is inversely related to the *CRT* score. This effect is present in the U-share of the no-correlation treatment only, as the effects in the L-share and in the perfect-correlation treatment are not significantly different from zero (see also Panel B of Table V).⁴²

E. Equilibrating Dynamics

OBSERVATION 10: *The relative frequency of discrepancies in the perfect-correlation treatment is lower than in the no-correlation treatment. Where a pricing discrepancy arises, subjects trade the high-expected-return twin share first, rather than the less profitable one. Individual subjects rarely exploit pricing discrepancies by asset swaps.*

Support. As reported in Panel C of Table III, on average 5.5% and 10.2% of all submitted orders lead to discrepancies in the order book for the perfect-correlation treatment and the no-correlation treatment, respectively.⁴³ We observe equilibrating dynamics. When a discrepancy arises, a trade in one

⁴² A significant correlation between RD_U and the *CRT* score is only notable in the no-correlation treatment. Conducting an OLS regression (with *CRT* as the only explanatory variable) separately for each treatment results in *p*-values of 0.318 (perfect correlation) and 0.026 (no correlation) for the slope of the *CRT* score. The other explanatory variables, including our measure of risk aversion and the gender share, have no significant impact on any *RD* in the two treatments.

⁴³ The difference between treatments is in the expected direction and is significant as reported above. The share of discrepancies and the *PD* in a session are highly correlated. The OLS regression (on the level of the independent session) of the *RFDLO* on the *PD* indicates a significantly positive slope (*z*-statistic of 2.43, that is, $p = 0.026$) and a coefficient of determination of 0.25. When a discrepancy arises, the median response time is four seconds.

Table VI

Equilibrating Dynamics: Relative Frequency of Market Orders in the High-Expected-Return Share of Given Discrepancy

This table reports the relative frequency in percent with which the market trades the more profitable rather than the less profitable share when a discrepancy arises. Significant differences per one-tailed Wilcoxon signed-ranks tests are indicated as follows: *** $p < 0.01$, ** $p < 0.05$, * $p < 0.10$. The last column reports p -values resulting from two-tailed, two-sample Mann-Whitney tests; z -statistics are in brackets.

	Perfect correlation ($n_1 = 12$)	No correlation ($n_2 = 8$)	Mann-Whitney test (between treatments), p -value [z -statistic]
Market 1	85.0*** [2.99]	95.8*** [2.52]	0.054** [2.47]
Market 2	88.4*** [3.07]	98.1*** [2.55]	0.086* [1.71]
Market 3	93.8*** [3.10]	81.8** [2.25]	0.332 [−0.97]
Market 4	87.0*** [3.07]	90.9*** [2.52]	0.586 [0.56]
Total	88.6*** [3.06]	91.7*** [2.52]	0.758 [0.31]

twin share typically leads to a higher expected return than a trade in the other.

As reported in Table VI, the relative frequencies at which the high-return trade is made first are 0.886 and 0.916 in the perfect-correlation treatment and the no-correlation treatment, respectively. The results of the Wilcoxon signed-ranks test show that the equilibration dynamics work as suggested above, in line with Hypothesis 5A. It is significantly more likely that a transaction occurs first in the high-expected-return twin share, thus eliminating the price discrepancy without necessarily booking an arbitrage gain. So prices are pushed in the direction of fundamental dividend value.⁴⁴

In fact, subjects rarely exploit the price discrepancy (as arbitrageurs would do) by selling the relatively expensive twin share and simultaneously buying the cheap one. Such hedging trades happen about 1% of the time when a discrepancy arises. Thus, our Hypothesis 5B is not supported by the data—99% of the discrepancies were eliminated by independent trades in the market rather than by individuals engaging in arbitrage, with no arbitrageur subjects dominating the market. We conclude that the market equilibration dynamics, rather than high-skilled individual traders, imply a narrowing of the gap between prices of twin shares in our treatments.

⁴⁴ The effect is a bit stronger in the no-correlation treatment than in the perfect-correlation treatment. However, equilibration dynamics are strong in both treatments. The difference is possibly related to the above-described fact that relative pricing is a more important driver of price behavior in the perfect-correlation treatment than in the no-correlation treatment.

F. Individual Portfolio Diversification

Despite the fact that diversification can reduce return volatility in the no-correlation treatment but not in the perfect-correlation treatment, the average difference in individual subjects' portfolios of L-shares and U-shares is 2 ($= dZ$) in each treatment. The difference in individual portfolio diversification between treatments is close to zero.⁴⁵ Hence, we do not find evidence in support of Hypothesis 6.

G. Summary of Experimental Results

We do not observe the pricing of assets at fundamental dividend value, but we have support for Hypothesis 1A (MM invariance theorem) in our perfect-correlation treatment, where exploiting price discrepancies is riskless. This differs in relative terms when we compare the data to the no-correlation treatment, where exploiting price discrepancies across assets has risk. Hypothesis 1B (no price discrepancy) is not supported in either of our treatments as price discrepancies seem to persist. However, under perfect correlation, adjustment to a level where discrepancies stabilize is fast (achieved by the second market), whereas under no correlation adjustment takes more time. Hypothesis 2 (hedging effect) is thus confirmed: the price discrepancy results of the perfect-correlation treatment are closer to the theoretical prediction than those of the no-correlation treatment.

Support for Hypothesis 3 (experience effect) is no surprise: experienced subjects trade closer to the theoretical predictions than inexperienced subjects. Nonetheless, unlike the results of the single-asset market (Smith, Suchanek, and Williams (1988)), our data do not suggest a complete convergence on fundamental dividend values or vanishing price discrepancies in repeated markets. Moreover, discrepant order submissions do not cease with repetition, as indicated by the stable relative frequency of discrepant limit orders. We find some support for Hypothesis 4 (smart trader effect), as our measure of aptitude increases with our measure of acuity and whole market mispricing decreases with acuity. To the extent that experience may substitute for initial acuity, our results confirm other results in the experimental asset markets literature (Palan (2013)).

The results support Hypothesis 5A. When a discrepancy arises, the more profitable transaction occurs first, which pushes price dynamics towards fundamentals. These dynamics occur in both treatments. In the perfect-correlation treatment, market equilibration vis-à-vis both fundamentals and the balancing of relative prices is strong. Both effects seem to lead to the confirmation of MM invariance in the perfect-correlation treatment. We note that response times to discrepancies are relatively low, and discrepancies are fewer than in the no-correlation treatment. In contrast, we do not find support for Hypothesis 5B as individual arbitraging through asset swaps is rare when discrepancies arise.

⁴⁵ The p -value of the one-tailed Mann-Whitney between-sample test is 0.47.

There are two surprising effects here. First, we observe no significant deviations from dividend value in the perfect-correlation treatment or for the L-share in the no-correlation treatment. This result runs counter to reports of mispricing in the single-asset market (see the survey by Palan (2013)). Second, we observe a positive significant deviation from dividend value only for the U-share in the no-correlation treatment, particularly when trader aptitude is low.⁴⁶ This effect seems to be driving the deviation from price parity in the no-correlation treatment in the direction as reported in Observation 1. Regarding the first effect, we must address the question of whether the co-existence of two assets limits the occurrence of bubbles. It has been suggested in the literature that trading in two assets can lead to smaller bubbles (Lee, Noussair, and Plott (2001)). Other papers (e.g., Kirchler and Huber (2012) and Palan (2013)) suggest that a careful explanation of the environment can also lead to smaller bubbles. If not spurious, the second effect could result from a behavioral bias such as the certainty effect (Tversky and Kahneman (1986)).⁴⁷ L-shares involve a possible zero dividend, since with a probability of 0.25 a zero dividend can occur in each period. In this context it seems noteworthy that the cumulative prospective utility of the lotteries corresponding to the L-share and the U-share differs by more than the difference in expected value when the parameter estimates of Tversky and Kahneman (1992) are applied. There is a small risk of not receiving any dividend pay at all. Given the extant literature and our no-correlation treatment results, we test the following hypothesis.

HYPOTHESIS 7 (certainty effect): *Share prices are higher when possible dividends are strictly positive than when dividends can be zero.*

We test this hypothesis in the following section using a between-subjects single-asset market experimental design. In one treatment subjects trade the L-share and in the other treatment they trade the U-share. We expect the *RD* measure to be higher for the U-share than for the L-share. If we do not observe this effect in the single-asset market setting, we must conclude that the price of the U-share results from relative market valuation in the no-correlation treatment.

V. Further Experiments

We conducted two additional treatments to further investigate patterns found in our two initial treatments. We describe these additional treatments below.

⁴⁶ The correlation of the U-share bubble with low trader aptitude could be a spurious effect.

⁴⁷ Tversky and Kahneman (1986) report evidence on the common ratio effect. In binary-choice tasks with a safe play and a risky lottery, they find that subjects had a preference for the safe play if the risky lottery involved the possibility of a zero payoff.

A. Single-Asset Market Experiments

We conducted a third treatment with only one asset to try to cleanly address whether the nonexistence of bubbles reported above results from the two-asset environment or from our general settings.⁴⁸ This third treatment also helps us address Observation 5b to see whether the higher relative market valuation of the U-share in the no-correlation treatment is a behavioral artifact similar to the dividend puzzle, according to which shares with nonzero dividends fetch a higher price (DeAngelo and DeAngelo (2006)).

The experimental procedures in the single-asset market experiments were the same as those described above (see the instructions in the Internet Appendix). However, instead of simultaneous trading in two assets, subjects traded in a single-asset market involving 10 end-of-period dividend draws. Applying between-subjects variation, subjects traded the shares with the equally likely dividends {0, 8, 28, 60} in the L-share treatment and with the equally likely dividends {24, 32, 56, 84} in the U-share treatment. Different from the original sessions, subjects participated in only two (not four) consecutive repetitions of the asset market.⁴⁹ The expected payoff for participation in the experiment was the same across all treatments. In the L-share treatment, subjects were endowed with six shares, and in the U-share treatment subjects were endowed with three shares. Subjects were able to short-sell double the number of shares with which they were endowed.

The data include eight independent cohorts in each of the L-share and U-share treatments, with each trading for two consecutive market rounds in the continuous double auction. A total of 138 subjects participated in these sessions. As before, the market size was nine.⁵⁰

OBSERVATION 11: *L-share and U-share average prices are close to dividend value in the single-asset environment.*

Support. Table VII reports the *RD* measures for the two single-asset treatments. The results of the Wilcoxon signed-ranks test indicate that deviations of the *RD* measures are not significantly different from zero for any market of the single-asset treatments. Compared to the *RD* measure of the U-share in the no-correlation treatment, the *RD* measure in the single-asset market treatments

⁴⁸ Besides the two-asset setting, other features in our design could have contributed to the nonexistence of bubbles. Subjects had to pass several comprehension tests and the expected dividend value was exhibited in each period on subjects' screens. This transparency may have decreased the lack of common knowledge (Cheung, Hedgaard, and Palan (2014)). Short-selling is possible, which can increase downward pressures of prices (Haruvy and Noussair (2006)). Also, the number of periods is 15 in the standard setting whereas it is 10 in ours.

⁴⁹ In total, subjects traded three times in 10 periods in continuous double-auction markets. Subjects traded in the L-share or the U-share twice. The third market of trading involved another experiment with single-period markets, which we describe further below. In half of the sessions, we had the two single-asset markets first and then the single-period markets. In the other half the ordering was reversed. We find no evidence of order effects, so we pool the data.

⁵⁰ In each of the two treatments, we had one cohort with seven, another with eight, and six with nine participants.

Table VII
RD in Single-Asset Market Experiment in Percent

This table reports the RD values for the single-asset sessions. There are no significant differences from zero per the two-tailed Wilcoxon signed-ranks tests. The last column reports p -values resulting from two-tailed, two-sample Mann-Whitney tests, z -statistics are in brackets.

	L-share treatment ($n_1 = 8$)	U-share treatment ($n_2 = 8$)	Mann-Whitney test (between treatments), p -value [z -statistic]
Market 1	3.57 [0.42]	-2.00 [0.42]	1.000 [0.00]
Market 2	-6.33 [-0.42]	-2.29 [0.00]	0.529 [-0.63]
Total average	-1.38 [-0.14]	-2.15 [0.28]	0.753 [-0.32]

is significantly smaller. Compared to each other, compared to the L-share in the no-correlation treatment, and compared to both the L-share and the U-share in the perfect-correlation treatment, all differences are insignificant according to a two-tailed Mann-Whitney test of two independent samples.⁵¹

Observation 11 indicates that we have no systematic mispricing in the single-asset treatments. Prior literature has suggested that carefully introducing subjects to the environment could amend mispricing (Palan (2013)). In our experiment, subjects were informed in each period about the remaining dividend draws and the expected sum of remaining dividends. It is quite possible that this information influenced the decision-making of subjects in our markets.⁵² Given Observation 11, we find no support for Hypothesis 7 in the single-asset environment. The relatively high valuation of the U-share in the no-correlation treatment seems to be an artifact of relative valuation. To see whether this result and also the confirmation of the MM theorem is an artifact of the described environment (Smith, Suchanek, and Williams (1988)), in the following section we report another presumably simpler experimental treatment with single-period markets.

B. Single-Period Market Experiments

Our fourth treatment involves single-period asset markets with four repetitions and a 10-fold dividend.⁵³ The overall procedures were the same as

⁵¹ The p -values of the test between single-asset market treatments are reported in the table. The p -values between RDs for the U-share in the no-correlation treatment and the U-share treatment are 0.046 and 0.016 for the first and second market, respectively; these p -values for the L-share treatment are 0.074 and 0.002, respectively. All other between-treatment comparisons, as listed in the text, show p -values above 0.480.

⁵² We provided this information to the subjects in an effort to reduce the cognitive burden in this rather complex environment. This may have also served to reduce bubbles, but this was not the primary point of the paper.

⁵³ In total there were 10 periods. The described assets were traded in periods 1, 4, 7, and 10. In the other periods, assets with the same expected values were traded. For ease of presentation we

described above apart from the following differences. Subjects were endowed in every period with two L-shares and two U-shares that they traded in the double-auction market. These shares paid one dividend at the end of the period and then were cancelled. In the experiment, one period would be selected for payment. Possible dividends on an L-share were $\{0, 80, 280, 600\}$. In the perfect-correlation treatment, the possible dividends on the U-share were always 240 cash units higher than the dividend on the L-share. In the no-correlation treatment, the dividends of the L-share and the U-share were independent; the equally likely dividends per U-share were $\{240, 320, 520, 840\}$.

Including the pilot session, nine independent cohorts participated in the perfect-correlation treatment, and eight independent cohorts in the no-correlation treatment.

OBSERVATION 12: We find significant differences from parity pricing neither in any period nor in any single-period asset market treatment. Between treatments we observe no significant differences in any period.

Support. Table IA.IV in the Internet Appendix reports the average difference from parity for the repetitions. The results show that the difference from price parity between assets is not significant in any treatment and period.⁵⁴ There are no significant differences between the two treatments.

Given Observation 12, on average subjects trade at fundamentals and at parity. However, there are significant deviations from parity within cohorts. In some cohorts the L-share is valued relatively above the U-share, while in other cohorts the U-share is valued above the L-share. Hence, we find no support for Hypothesis 7 in the single-period market; the high relative valuation of the U-share of the no-correlation treatment is not reproduced in this comparatively simple market environment. In line with Observation 8, the shares in the single-period treatment trade closer to fundamentals with higher acuity in the market. In contrast to Observations 1 and 2, there are no significant differences between treatments in the single-period treatment. This result suggests that the differences between arbitrage dynamics and equilibrium dynamics cannot be easily detected in a simple market setting. To detect such differences, one may need a rather extreme test like the one we use in the context of bubbles and declining fundamental values (Smith, Suchanek, and Williams (1988)).

do not discuss these assets here. However, results based on those assets are completely in line with the reported data. Please see the instructions in the Online Appendix for further details. Subjects traded in a total of three consecutive markets of 10 periods, one involving a single-period treatment and the other two involving a single-asset treatment. We found no order effects in the data, so we pool the data by treatment.

⁵⁴ For completeness, the average absolute deviations from parity are 14% and 12% and the deviations from fundamentals are 16% and 13% for the perfect-correlation treatment and the no-correlation treatment, respectively. The absolute deviations and the deviations from fundamentals do not decline significantly between repetitions in either treatment.

VI. Conclusions

Our laboratory results support the MM invariance theorem in the following way. As proposed in the arbitrage proof of the original paper (MM (1958)), we observe the same average market values for shares of perfectly correlated returns. The results suggest that perfect correlation of future returns (where price discrepancies can be arbitrated away without risk) is required to obtain the law of one price in our setting. This necessary condition is recommended through our control treatment, where perfect correlation is removed. The deviations from parity pricing and the observed pricing discrepancies are significantly larger in the control treatment than in the perfect-correlation treatment. Despite the fact that earlier research reports that behavior is insensitive to changes in return correlations (e.g., Kroll, Levy, and Rapoport (1988)), we do find a significant effect. Under no correlation, that is, with independent multiperiod-lived assets, we do not obtain the law of one price for the same idiosyncratic risks of our single assets. With limits to arbitrage, the market manifests a risk premium. The (“leveraged”) L-share is less pricey than the (“all equity”) U-share. This could be a spurious effect—our research in this respect is not conclusive. We investigated whether this effect may be driven by a behavioral bias in relative valuation similar to the certainty effect (Tversky and Kahneman (1986)). The fact that the dividend in the U-share is strictly positive (whereas the L-share dividend can be zero) could possibly bias subjects to push U-share prices up when trader acuity is low. We observe this effect only in the two-asset setting of the no-correlation treatment. The higher market valuation of the U-share compared to the L-share treatment in the no-correlation treatment is not observed in our single-asset market treatments. For existing limits to arbitrage, our results (in the no-correlation treatment) suggest that capital structure may indeed affect the market value of the company. The finding that value-invariance holds under perfect correlation but not necessarily with imperfect correlation is a key contribution of the paper. This result is intriguing, as the CAPM suggests that the law of one price holds in our zero-correlation setting. This result, along with the result that single-asset markets behave differently from multiple-asset markets, indicates that more experimental research is needed to uncover the trading dynamics and interactions of multiple multiperiod-lived assets.

Our study does not provide a test of the unrestrained validity of MM invariance, since we consider a zero interest rate and nontraded debt. Nonetheless, we note that we obtain this result in the standard setting of Smith, Suchanek, and Williams (1988), known to produce mispricing among inexperienced subjects. We find that value-invariance can also be obtained in simpler settings, including those with constant fundamental dividend value. For the single-period setting, our results suggest that the law of one price can be obtained on average even if returns are not perfectly correlated but only the same in expected terms. In more complex settings, as suggested in our experiments with two multiperiod-lived assets, MM invariance seems to require perfect correlation of future return streams.

Our results also shed light on different levels of market efficiency. We find support for market efficiency in the case of riskless arbitrage. We also find that market efficiency increases with repetition. Price discrepancies decrease and prices move towards fundamental dividend value with repetition, and we document the existence of equilibrium dynamics when price discrepancies arise. We observe that high overall acuity, as measured by the CRT score, reduces price discrepancy in the market. Deviations from theoretical predictions become smaller when we look at markets with overall high trader acuity. Nonetheless, the deviations are never entirely removed. So for the results from our design, we have the sense that Rietz (2005) is correct when he states that without a dedicated arbitrageur who takes advantage of any mispricing, the laboratory market does not completely eliminate price discrepancies.

As is frequently observed with market studies, we conclude that the data suggest a movement of behavior in the *direction* of the predictions of financial economic theory, but do not provide full quantitative support for the theoretical benchmark. Hence, as price discrepancies do not cease during the experiment, our results hint at significant potential gains for arbitrageurs. In the real world, exploiting price discrepancies requires that the arbitrageur take risk (as has been shown in the twin-share LTCM cases; see Shleifer (2000)). So in the real world, arbitrageurs may require deep pockets and patience to lock in their arbitrage gains.

Initial submission: February 2, 2016; Accepted: September 13, 2017
 Editors: Bruno Biais, Michael R. Roberts, and Kenneth J. Singleton

REFERENCES

- Abbink, Klaus, and Bettina Rockenbach, 2006, Option pricing by students and professional traders: A behavioural investigation, *Managerial and Decision Economics* 27, 497–510.
- Ackert, Lucy F., Narat Charupat, Richard Deaves, and Brian D. Kluger, 2009, Probability judgment error and speculation in laboratory asset market bubbles, *Journal of Financial and Quantitative Analysis* 44, 719–744.
- Akiyama, Eizo, Nobuyuki Hanaki, and Ryuichiro Ishikawa, 2017, It is not just confusion! Strategic uncertainty in an experimental asset market, *The Economic Journal* 127, 563–580.
- Asparouhova, Elena, Peter Bossaerts, and Charles R. Plott, 2003, Excess demand and equilibration in multi-security financial markets: The empirical evidence, *Journal of Financial Markets* 6, 1–21.
- Biais, Bruno, and Sébastien Pouget, 1999, Microstructure, incentives and convergence to equilibrium in experimental financial markets, Working paper, Toulouse University, France.
- Bossaerts, Peter, Charles R. Plott, and William R. Zame, 2007, Prices and portfolio choices in financial markets: Theory, econometrics, experiments. *Econometrica* 75, 993–1038.
- Breaban, Adriana, and Charles N. Noussair, 2015, Trader characteristics and fundamental value trajectories in an asset market experiment, *Journal of Behavioral and Experimental Finance* 8, 1–17.
- Cadbury, Bram C., and Elisabeth Maynes, 1998, Laboratory experiments in corporate and investment finance: A survey, *Managerial and Decision Economics* 19, 277–298. [https://onlinelibrary.wiley.com/doi/abs/10.1002/\(SICI\)1099-1468\(199806/08\)19:4/5%3C277::AID-MD-E891%3E3.0.CO;2-3](https://onlinelibrary.wiley.com/doi/abs/10.1002/(SICI)1099-1468(199806/08)19:4/5%3C277::AID-MD-E891%3E3.0.CO;2-3)

- Chan, Kenneth S., Vivian Lei, and Filip Vesely, 2013, Differentiated assets: An experimental study on bubbles, *Economic Inquiry* 51, 1731–1749.
- Charness, Gary, and Uri Gneezy, 2010, Portfolio choice and risk attitudes: An experiment, *Economic Inquiry* 48, 133–146.
- Charness, Gary and Uri Gneezy, 2012, Strong evidence for gender differences in risk-taking, *Journal of Economic Behavior and Organization* 83, 50–58.
- Cheung, Stephen L., Morten Hedegaard, and Stefan Palan, 2014, To see is to believe: Common expectations in experimental asset markets, *European Economic Review* 66, 84–96.
- Corgnet, Brice, Roberto Hernán-González, Praveen Kujal, and David Porter, 2015, The effect of earned vs. house money on price bubble formation in experimental asset markets. *Review of Finance* 19, 1455–1488.
- DeAngelo, Harry, and Linda DeAngelo, 2006, The irrelevance of the MM dividend irrelevance theorem, *Journal of Financial Economics* 79, 293–315.
- Dufwenberg, Martin, Tobias Lindqvist, and Evan Moore, 2005, Bubbles and experience: An experiment, *American Economic Review*, 95, 1731–1737.
- Eckel, Catherine C., and Sascha Füllbrunn, 2015, Thar “she” blows? Gender, competition, and bubbles in experimental asset markets, *American Economic Review* 105, 906–920.
- Fischbacher, Urs, 2007, z-Tree: Zurich toolbox for ready-made economic experiments, *Experimental Economics* 10, 171–178.
- Frederick, Shane, 2005, Cognitive reflection and decision making, *Journal of Economic Perspectives* 19, 25–42.
- Froot, Kenneth A., and Emil M. Dabora, 1999, How are stock prices affected by the location of trade?, *Journal of Financial Economics* 53, 189–216.
- Gode, Dhananjay, and Shyam Sunder, 1993, Allocative efficiency of markets with zero-intelligence traders: Market as a partial substitute for individual rationality, *Journal of Political Economy* 101, 119–137.
- Greiner, Ben, 2015, Subject pool recruitment procedures: Organizing experiments with ORSEE, *Journal of the Economic Science Association* 1, 114–125.
- Haruvy, Ernan, Yaron Lahav, Charles Noussair, 2007, Traders’ expectations in asset markets: Experimental evidence, *American Economic Review* 97, 1901–1920.
- Haruvy, Ernan, and Charles N. Noussair, 2006, The effect of short selling on bubbles and crashes in experimental spot asset markets, *Journal of Finance* 61, 1119–1157.
- Judd, Kenneth L. and Sy-Ming Guu, 2001, Asymptotic methods for asset market equilibrium analysis, *Economic Theory* 18, 127–157.
- Kirchler, Michael, Jürgen Huber, and Thomas Stöckl, 2012, Thar she bursts—Reducing confusion reduces bubbles, *American Economic Review* 102, 865–883.
- Kroll, Yoram, Haim Levy, and Amnon Rapoport, 1988, Experimental test of the separation theorem and the capital asset pricing model, *American Economic Review* 78, 500–519.
- Lamont, Owen A., and Richard H. Thaler, 2003, Anomalies: The law of one price in financial markets, *Journal of Economic Perspectives* 17, 191–202.
- Lei, Vivian, Charles N. Noussair, and Charles R. Plott, 2001, Nonspeculative bubbles in experimental asset markets: Lack of common knowledge of rationality vs. actual irrationality, *Econometrica* 69, 831–859.
- Levati, M. Vittoria, Jianying Qiu, and Prashanth Mahagaonkar, 2012, Testing the Modigliani-Miller theorem directly in the lab. *Experimental Economics* 15, 693–716.
- Modigliani, Franco and Merton H. Miller, 1958, The cost of capital, corporate finance and the theory of investment, *American Economic Review* 48, 261–97. https://www.jstor.org/stable/1809766?casa_token=hAebTjMM-uMAAAAA:6fRqndfM6KOYx9IMmXIpAlO-2Acm09KkdGrBQFjLIR7-LRcJKJft5KdC-amuHIQW4cMeMqzr5MprzEZA2gCxL_E6v89LG-pNufcz2NgyqfStQVQQSKw&seq=1#metadata_info_tab_contents
- Modigliani, Franco, and Merton H. Miller, 1959, The cost of capital, corporate finance and the theory of investment: Reply, *American Economic Review* 49, 655–669. https://www.jstor.org/stable/1812919?casa_token=qGPOIgZWoroAAAAA:plQ2sl1EkvFnRT8aMy7G-f8-xxDf0kY7fZc4givS3z-tudY9Gca1bAiFAjSzcCaRj1rAiY1KQchg_O4DMkL9NdHjcg_RN4-x92zk9cl20RJoe5CPb_g&seq=1#metadata_info_tab_contents

- Modigliani, Franco, and Merton H. Miller, 1963, Corporate income taxes and the cost of capital: A correction, *American Economic Review* 53, 433–443. https://www.jstor.org/stable/pdf/1809167.pdf?casa_token=qv1mRs5mZ4kAAAAA:ncKy2dSHwEcVn-ZsKOIECD3mI_g-9FMymJRHer4p1VEZNVczKjK25wYj_iH8aGxtF4gtVPn_hgMH-O_oK8i42-aG-UI5veRm5N55Xa86RKsh93lSfCA
- Noussair, Charles N., and Steven Tucker, 2006, Futures markets and bubble formation in experimental asset markets, *Pacific Economic Review* 11, 167–184.
- Noussair, Charles N., Steven Tucker, and Yilong Xu, 2016, Futures markets, cognitive ability, and mispricing in experimental asset markets, *Journal of Economic Behavior & Organization* 130, 166–179.
- O'Brien, John, and Sanjay Srivastava, 1991, Dynamic stock markets with multiple assets: An experimental analysis, *The Journal of Finance* 46, 1811–1838.
- Oliven, Kenneth, and Thomas A. Rietz, 2004, Suckers are born but markets are made, *Management Science* 50, 336–351.
- Palan, Stefan, 2013, A review of bubbles and crashes in experimental asset markets, *Journal of Economic Surveys* 27, 570–588.
- Porter, David P., and Vernon L. Smith, 1995, Futures contracting and dividend uncertainty in experimental asset markets. *Journal of Business* 68, 509–541.
- Rietz, Thomas A., 2005, Behavioral mispricing and arbitrage in experimental asset markets. Working paper, University of Iowa.
- Shleifer, Andrei, 2000, *Inefficient Markets: An Introduction to Behavioral Finance*, Clarendon Lectures in Economics (Oxford University Press, Oxford).
- Shleifer, Andrei, and Lawrence H. Summers, 1990, The noise trader approach to finance, *Journal of Economic Perspectives* 4, 19–33.
- Smith, Vernon L., 1962, An experimental study of competitive market behavior, *Journal of Political Economy*, 70, 111–137.
- Smith, Vernon L., Gerry L. Suchanek, and Arlington W. Williams, 1988, Bubbles, crashes, and endogenous expectations in experimental spot asset markets, *Econometrica* 56, 1119–1151.
- Stiglitz, Joseph, 1969, A re-examination of the Modigliani-Miller theorem, *American Economic Review* 59, 784–793. https://www.jstor.org/stable/1810676?casa_token=7OneR-m0lmcAAAAA:ZuDBkh2rPeiae12tftN4q6gFeYsmoqXhLsGsbQVR-oEN2nVFHrV1e0-DvBNTIEIAKNIo-bk8jfxRdyQJ_76oOOCbbvaSBsnhvcnyGI8mNSaXLYZyAUrA&seq=1#meta-data_info_tab_contents
- Stöckl, Thomas, Jürgen Huber, and Michael Kirchler, 2010, Bubble measures in experimental asset markets, *Experimental Economics* 13, 284–298. <https://link.springer.com/article/10.1007/s10683-010-9241-9>
- Sunder, Shyam, 1995, Experimental asset markets, in John Kagel and Al Roth, eds.: *The Handbook of Experimental Economics* (Princeton University Press, Princeton, NJ).
- Tversky, Amos, and Daniel Kahneman, 1986, Rational choice and the framing of decisions, *Journal of Business* 59, S251–S278.
- Tversky, Amos, and Daniel Kahneman, 1992, Advances in prospect theory: Cumulative representation of uncertainty, *Journal of Risk and Uncertainty* 5, 297–323.

Supporting Information

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

Appendix S1: Internet Appendix. Replication Code.