ISSN (Print) 0473-453X
Discussion Paper No. 1257
ISSN (Online) 2435-0982

AN EXPERIMENTAL ANALYSIS ON CROSS-ASSET ARBITRAGE OPPORTUNITY AND THE LAW OF ONE PRICE

Jieyi Duan
Nobuyuki Hanaki

Revised February 2026
September 2024

The Institute of Social and Economic Research
The University of Osaka
6-1 Mihogaoka, Ibaraki, Osaka 567-0047, Japan

An experimental analysis on cross-asset arbitrage opportunity and the law of one price*

Jieyi Duan[†] Nobuyuki Hanaki[‡]

February 8, 2026

Abstract

This study experimentally investigates the impact of the lack of arbitrage opportunities across different assets on the realization of the law of one price. Our experiment is based on the framework established by Charness and Neugebauer (2019) where participants, acting as traders, are involved in transactions with two different types of assets. An increase in the magnitude of price discrepancies and fundamental mispricing are observed when traders are unable to engage in arbitrage between different assets.

JEL Code: C90, D84

Keywords: The law of one price, Arbitrage opportunities across different assets, Price discrepancy, Asset Pricing

*We thank comments and suggestions from participants at International Workshop on Experimental Economics, Osaka University, John Duffy in particular. We gratefully acknowledge financial support from the Joint Usage/Research Center at ISER, Osaka University, and Japan Society for the Promotion of Science KAKENHI Grant Numbers 18K19954, 20H05631, 23H00055, and 25H00388. The experiment reported in this paper was approved by the Research Ethics Committee at the Institute of Social and Economic Research, Osaka University (No. 20220604). This work was also supported by JSPS Core-to-Core Program (grant number:JPJSCCA20200001). The experiment is pre-registered at https://aspredicted.org/jh76w.pdf on 10 October, 2022.

†Otemon Gakuin University. E-mail: j-duan@haruka.otemon.ac.jp

‡Corresponding author. Institute of Social and Economic Research, Osaka University, Japan, and University of Limassol, Cyprus. E-mail: nobuyuki.hanaki@iser.osaka-u.ac.jp

1 Introduction

In a competitive market where traders engage in transactions with multiple distinct assets, the law of one price stipulates that the returns on these assets should be equalized through arbitrage (Childs and Mestelman, 2006).

Nonetheless, a growing body of research suggests that the law of one price is not consistently upheld, as documented in both empirical and experimental studies. Such violations manifest as price discrepancies, defined as the relative deviation of the price ratio between two distinct but fundamentally related assets from the parity relationship implied by their fundamental values (the precise definition of price discrepancy is provided in Subsection 3.1). An empirical analysis by Owen and Thaler (2003) found price discrepancies in situations involving closed-end country funds, twin shares, dual-class shares, and even corporate spin-offs, attributing these anomalies to the restricted capacity of rational arbitrageurs to act. Experimental studies by Childs and Mestelman (2006) and Chan et al. (2013), which introduced twin-market setups, revealed that the magnitudes of price discrepancies are linked to differences in asset properties. While Childs and Mestelman (2006) noted increased price discrepancies with divergent expected dividend values, Chan et al. (2013) observed reduced discrepancies under similar conditions, arguing that differences in asset characteristics promote cross-asset arbitrage, thus significantly curtailing mispricing.

Further experimental scrutiny of the Modigliani-Miller invariance theorem (MM theorem), such as Levati et al. (2012) and Charness and Neugebauer (2019), has also highlighted breaches of the law of one price, with both studies attributing these deviations to the absence of arbitrage opportunities across different assets. In particular, Charness and Neugebauer (2019) observed significant price discrepancies when trading a leveraged and an unleveraged asset simultaneously, especially when their future dividend processes were uncorrelated. They posited that the presence of arbitrage risk in future uncorrelated dividend processes deterred traders from engaging in arbitrage, leading to pronounced price discrepancies.

The aforementioned interpretations, however, appear to be at odds with conventional economic theories (e.g., Hirshleifer, 1966; Stiglitz, 1969) which posit that the lack of cross-assets arbitrage opportunities should not affect the level of price discrepancies or mispricing, as arbitrage within each asset is expected to correct mispricing relative to fundamentals, thereby reducing price discrepancies. Therefore, this perspective from traditional economic theories stands in opposition to the explanations provided by the aforementioned experimental studies.

Moreover, a substantial body of experimental research, building on the seminal asset market design of Smith et al. (1988), has examined how arbitrage and speculative trading influence asset market outcomes. For example, Noussair and Tucker (2006) finds that introducing futures markets can mitigate speculative bubbles by enabling arbitrage, while Duffy et al. (2022) shows that institutional trading via exchange-traded funds (ETFs) reduces mispricing and stabilizes prices. By contrast, Ben-David et al. (2018) provide empirical evidence that arbitrage activity can increase volatility when arbitrageurs interact with noise traders.

Notwithstanding these insights, prior experimental studies have not directly manipulated the feasibility of arbitrage between distinct asset markets, leaving an important gap in understanding its causal effect on price discrepancies and fundamental mispricing. This gap in the literature creates ambiguity regarding the connection between arbitrage opportunities across distinct assets and the achievement of the law of one price.

The present study addresses this gap by employing a controlled laboratory design that explicitly varies the feasibility of cross-asset arbitrage. This approach extends earlier experimental methodologies (Charness and Neugebauer, 2019; Childs and Mestelman, 2006; Chan et al., 2013) by directly controlling arbitrage opportunities rather than altering asset characteristics or dividend correlations. Should the lack of arbitrage between different assets indeed lead to price discrepancies, as some literature suggests, then limiting arbitrage opportunities across different assets might exacerbate these discrepancies. Furthermore, to distinguish arbitrage effects from general bubble phenomena, we employ a relative price discrepancy measure between two linked assets, thereby isolating arbitrage-induced mispricing from broader bubble dynamics.

To ensure comparability with existing experimental literature, our experimental design is based on Charness and Neugebauer (2019), utilizing the framework established by Smith et al. (1988)[1]. Participants, in the role of traders, engage in transactions with two distinct types of financial assets. Despite having identical risk profiles, these assets differ in their fundamentals. The future dividend processes of these assets are perfectly positively correlated, and transactions are conducted using experimental currency. The experiment features two treatments: in one, traders can transact both asset types simultaneously, allowing for cross-asset arbitrage; in the other, traders are limited to transactions with just one asset type, precluding cross-asset arbitrage. In both treatments, traders have access to the latest transaction and average prices of all asset types, enabling them to gather trading information about the other asset type regardless of their treatment.

Our findings suggest that the absence of arbitrage opportunities leads to a marginally significant increase in both the magnitude of price discrepancies and the degree of

[1] A substantial number of experimental studies have utilized this framework to explore the underlying causes of asset bubbles. Comprehensive surveys of these studies are provided by Palan (2013), Powell and Shestakova (2016), Nuzzo and Morone (2017), Duffy et al. (2022). and Angerer et al. (2023)

mispricing relative to fundamentals, consistent with the suggestions of Charness and Neugebauer (2019), Chan et al. (2013), and Levati et al. (2012).

Moreover, our data indicate that the treatment effect on price discrepancy is positively related to the amount of unrealized cross-asset arbitrage profits in outstanding orders, and that a substantial amount of cross-asset arbitrage behavior is observed when such arbitrage opportunities exist.

The rest of the paper is organized as follows. Section 2 describes the experimental design. Section 3 outlines the measures and hypotheses. Section 4 provides an analysis of the data. Section 5 reports robustness checks based on an additional treatment. Finally, Section 6 offers some conclusions.

2 Experimental Design

2.1 Basic experimental design

Our experimental design extends that of Charness and Neugebauer (2019). The experiment comprises three finite horizon economies, with each economy, termed a "sequence", consisting of $T = 10$ discrete trading periods. Within each trading period, participants can engage in buying and selling of two distinct types of multi-period-lived assets in continuous double-auction markets, using the experimental currency denominated as "Cents".

For two distinct types of multi-period-lived assets, one is denoted as "A-share." This asset pays its holders dividends drawn from {0, 8, 28, or 60 cents} with equal probability at the end of each period. In contrast, the other asset denoted as "B-share," consistently pays dividends 24 cents higher than the A-share at the end of each period.[2] Upon the conclusion of each sequence, all held shares are forfeited without

[2]Note that in this experiment, the random realization of dividends for A-shares in each period

compensation. Consequently, the risk neutral fundamental value of each share type, during each period, is equal to the expected dividend value of the remaining period. Within this framework, the fundamental value of B-shares, at the beginning of each sequence, is 24×10 cents higher than that of A-shares, despite their equivalent risk levels. In each sequence, there are 14 traders equally distributed into two types: Type A and Type B. The primary distinction between these two trader types lies in their endowments.

Specifically, at the outset of the sequence (t=1), each Type A trader receives an endowment of 6 units of A-share and 1,200 cents of experimental currency. Conversely, each Type B trader receives an endowment of 3 units of B-share and the same 1,200 cents of experimental currency. Table 1 shows the initial individual endowments of each type of trader, expected dividend value, and initial total variance for each type of share.

Moreover, all traders were able to borrow up to 2,400 cents for the purchase of assets and could short-sell up to four units of A-share and up to four units of B-share without any margin requirements. The trading flow was unaffected (i.e., there was no message indicating a short sale rather than a long sale) by short sales and borrowings, which were displayed as negative numbers. The shorted share pays a negative dividend to the holders.

To reduce confusion and, in turn, pricing discrepancies, participants were reminded on screen about 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.

---

was not determined in advance. As a result, the dividend processes varied across sessions.

Table 1: Initial Individual Endowments, Expected Dividend Values and Variances

	Trader	Initial unit endowment	Expected dividend value/unit	Initial total Variance/unit
A-shares	Type A trader	6	$24 \times (T - t + 1)$	$536 \times (T - t + 1)$
	Type B trader	0		
B-shares	Type A trader	0	$48 \times (T - t + 1)$	$536 \times (T - t + 1)$
	Type B trader	3		
Cents	All trader	1200	1	-

Note: The third column delineates the individual unit endowments in shares and cents for each type of trader. The fourth and fifth columns enumerate the expected values and variances of each share type's dividend, respectively. Given that $t \leq T = 10$, the expected payoffs for the A-share and B-share are 240 and 480 cents, respectively, with an initial total variance for each share being 5360. Over time, both the variances and expected dividend values exhibit a linear decline.

2.2 Trading mechanism

The experiment adopts an open-book continuous double-auction mechanism. Each trading period lasts 180 seconds in the first sequence and lasts 90 seconds in the remaining sequence. There are two markets, one for trading A-shares and the other for trading B-shares. Once trading begins, traders can submit a bid (a buy order) and/or an ask (a sell order) for a unit of shares in these markets simultaneously in continuous time. Traders are allowed to transact as many units of shares as they desire within their budget constraint until the countdown timer expires for each period. A transaction occurs when the best bid and the best ask cross at the price determined by whichever is submitted first. Once a transaction takes place, cash and share holdings are immediately updated, and all the outstanding bids of the buyer and all the outstanding offers of the seller are canceled in both markets.

2.3 Treatments

The experiment incorporates two treatments, wherein the presence of arbitrage opportunities between A-shares and B-shares is varied. Specifically, the treatment denoted below as W allows traders to engage in arbitrage across A-shares and B-shares, while adhering to the basic experimental design outlined above. Within this treatment, all traders can transact both types of shares simultaneously. Conversely, in the WO treatment, traders are restricted to trading only shares of their type. Hence, traders categorized as Type A can only transact A-shares, while those categorized as Type B can only transact B-shares. Additionally, Type A traders can short up to 12 units of A-shares but are prohibited from shorting B-shares, whereas Type B traders have the capability to short up to 6 units of B-shares but are restricted from shorting A-shares. As a result, traders within this treatment are precluded from engaging in arbitrage across the two share types.

2.4 Payment

Upon the conclusion of each sequence, the cents held are converted to Japanese Yen (JPY) at an exchange rate of 2 cents = 1 JPY, serving as the earnings of the participants for that sequence. Upon conclusion of the experiment, a computer randomly selects one sequence to calculate the participants' earnings. Participants receive cash payments based on their earnings in the selected sequence, plus a 1,500 JPY participation fee. If a participant incurs negative earnings in the selected sequence, the amount will be deducted from the participation fee; however, the total payment from the experiment did not fall below 1000 JPY.

2.5 Data collection

The experiment was executed utilizing z-Tree (Fischbacher, 2007) and conducted at the experimental laboratory of the Institute of Social and Economic Research at Osaka University from October 2022 to January 2023. All participants were students enrolled at Osaka University registered at ORSEE (Greiner, 2015) database.

Each treatment comprised 8 groups, with each group consisting of 14 participants, culminating in a total of 224 student participants across 16 sessions.

In every session, an instructional video[3] was presented to the participants, supplemented by printed handouts for reference. A quiz was administered to ascertain participants' comprehension of the experimental rules, including the computation of their payoffs. To ensure participants understood the rules, the experiment began only after all participants had answered all the questions correctly. Preceding the main experiment, participants' risk attitudes were assessed utilizing the Certainty Equivalent Method following the guidelines in Healy (2020).

The average payoff amounted to 3464 JPY ($\approx 30.48\, USD$, based on the prevailing exchange rate during the experiment period). The duration of each session ranged from 2 to 2.5 hours. No significant differences were observed in gender distribution (Percentage of females: 33% in WO and 34% in W; Fisher's Exact Test: $p = 1$) or in risk attitudes (the average scores were 46.9 in WO and 48.6 in W; Mann-Whitney U Test: $p = 0.338$) between the two treatments.

3 Measures and Hypotheses

3.1 Measures

In this paper, we define "price discrepancy" as the relative deviation of the price ratio between two distinct but fundamentally related assets from the parity relationship implied by their fundamental values. According to the law of one price, assets with identical cash-flow profiles (except for a constant difference) should exhibit an identical relative price ratio. Therefore, to accurately compare the extent of violations

[3] The English translation of the instruction can be downloaded from https://zenodo.org/records/18521815.

of the law of one price across different treatments, we employ two measures, denoted by Δ and PD, adapted from the work of Charness and Neugebauer (2019). These measures specifically capture the extent to which the observed price ratio deviates from the theoretical parity relationship, rather than merely measuring deviations from absolute fundamental values. For a given group g in sequence r,

$$\Delta_t^{g,r} = \frac{P_{B,t}^{r,g}}{P_{A,t}^{r,g} + (FV_{B,t} - FV_{A,t})} - 1 \tag{1}$$

$$PD^{g,r} = \frac{1}{T} \sum_{t=1}^{T} |\Delta_t^{g,r}| \tag{2}$$

where $P_{k,t}^{r,g}$ and $FV_{k,t}$ represent the average transaction price and the corresponding fundamental value of k-shares ($k \in \{A, B\}$) in group g at period t in sequence r.

However, it should be noted that both Δ and PD tend to underestimate the magnitude of discrepancies when A-shares are, on average, priced higher than B-shares. Specifically, while PD can exceed 1 if B-shares are overvalued relative to A-shares, it cannot exceed 1 when A-shares are similarly overvalued relative to B-shares. To address this bias, we introduce a new measure termed the Log-Price Discrepancy (LPD). For group r in sequence r,

$$LPD^{r,g} = \frac{1}{T} \sum_{t=1}^{T} |\log\left(\frac{P_{B,t}^{r,g}}{P_{A,t}^{r,g} + (FV_{B,t} - FV_{A,t})}\right)| \tag{3}$$

The measure Δ captures the direction and magnitude by which the price of one share type deviates from parity pricing relative to the other. Specifically, $\Delta > 0$ indicates that B-shares are overpriced relative to A-shares, whereas $\Delta < 0$ indicates the opposite. The measures PD and LPD represent the overall magnitude of price discrepancies, reflecting potential gains from selling the relatively overpriced asset and purchasing the relatively underpriced asset. Notably, PD and LPD equal zero when the prices of A-shares and B-shares align exactly with their respective fundamental values or deviate by the same magnitude.[4]

Given that asset bubbles commonly emerge in the experimental environment of Smith et al. (1988), absolute deviations from fundamental values may occur frequently. By explicitly focusing on relative prices between two assets rather than their absolute prices, PD and LPD measures seek to isolate the specific impact of cross-asset arbitrage opportunities from general bubble-driven deviations. Thus, despite the potential presence of asset bubbles, our defined measures of price discrepancy allow for clearer evaluation of arbitrage-driven deviations—an approach consistent with that of Charness and Neugebauer (2019).

Even when Δ, PD, LPD are zero, asset prices may still diverge from their fundamental values. To quantify deviations from fundamental values, we employ two established metrics from the single-asset market literature proposed by Stöckl et al. (2010): the Relative Deviation (RD) and the Relative Absolute Deviation (RAD). For type k share in group g in sequence r,

$$RD_k^{g,r} = \frac{1}{T}\sum_{t=1}^{T} \frac{P_{k,t}^{r,g} - FV_{k,t}}{\overline{FV}_k} \tag{4}$$

[4]We note that the analyses presented in this paper deviate from the original pre-analysis plan outlined in our pre-registration. Specifically, the deviations from our pre-registered analyses are as follows: First, we introduce two additional measures, LPD and RAD, which were not specified in the original pre-registration. Second, compared to our pre-registration, the alternative hypotheses employed in this paper are directional, allowing the use of one-sided rather than two-sided tests (see Subsection 3.2). The rationale behind these deviations is threefold. First, we introduced the LPD measure because the previously employed PD measure (as used by Charness and Neugebauer (2019)) tends to underestimate cases in which A-shares are, on average, priced higher than B-shares—a point further discussed in this subsection. Second, the RAD measure was included to enhance comparability with recent studies, such as Angerer et al. (2023), which used it to quantify asset mispricing relative to fundamental values. Lastly, we adopted directional alternative hypotheses since our hypotheses were originally formulated based on findings from Charness and Neugebauer (2019), making one-sided tests more appropriate. Nonetheless, to ensure transparency, detailed results based on the original pre-registered measures and hypotheses are provided in Online Appendix A.

$$RAD_k^{r,g} = \frac{1}{T}\sum_{t=1}^{T}\left|\frac{P_{k,t}^{r,g} - FV_{k,t}}{\overline{FV}_k}\right| \tag{5}$$

where $\overline{FV}_k = \frac{\sum_t FV_{k,t}}{T}$ denote the average fundamental value of type k share.[5]

The measure RD_k captures both the direction and magnitude by which the average transaction price of type-k deviates from their fundamental values. Specifically, $RD_k > 0$ indicates overpricing relative to fundamentals, while $RD_k < 0$ indicates underpricing. In contrast, the measure RAD_k quantifies the magnitude of deviation (ignoring direction), with RAD_k equaling zero if and only if the shares' average transaction prices equal their fundamental values in every period.

Lastly, to accurately compare the transaction price deviation from fundamentals for the entire market, we apply the Relative Absolute Deviation from Fundamentals of the whole market, denoted as DF, defined by Charness and Neugebauer (2019). For a group g in sequence r,

$$DF^{r,g} = \frac{1}{2T}\sum_{k\in\{A,B\}}\sum_{t=1}^{T}\left|\frac{P_{k,t}^{r,g} - FV_{k,t}}{FV_{k,t}}\right| \tag{6}$$

Notice the difference in the denominator between DF and RD or RAD, while the deviation of the average price from the fundamental value in period t is normalized by the fundamental value in the same period for DF, it is normalized by the average fundamental value across all period for RD and RAD.

3.2 Hypotheses

According to established economic theory (e.g. Hirshleifer, 1966; Stiglitz, 1969), the presence of arbitrage opportunities between two types of shares is not a necessary

[5]The pre-registration contained a typo in defining RD. Namely, $\overline{FV}_k$ in the denominator of Equation 4 was shown as $FV_{k,t}$.

condition for the law of one price to be valid. Furthermore, the Rational Expectations Equilibrium (REE), as posited by standard economic theory, suggests that the degree of mispricing for each share type remains unaffected by the existence of the arbitrage opportunities between the A-shares and B-shares. We thus state the following null hypotheses.

Hypothesis 1 *There are no statistically significant differences in Δ, PD, and LPD between the W treatment and the WO treatment.*

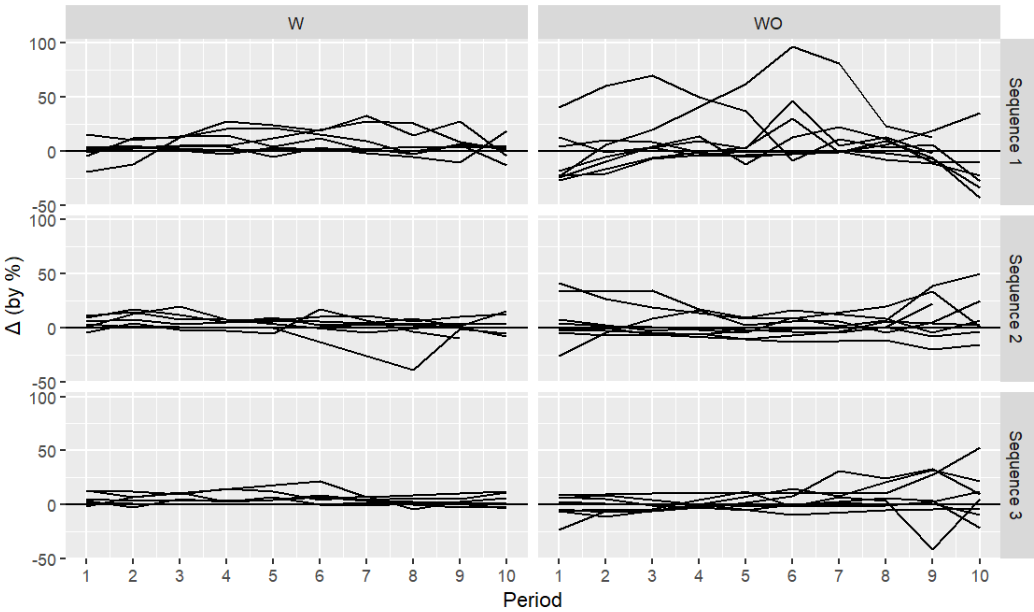
Hypothesis 2 *There are no statistically significant differences in RD_k, RAD_k, and DF between the W and WO treatments.*

In contrast, Charness and Neugebauer (2019) posit that the absence of cross-asset arbitrage opportunities can induce participants to overprice B-shares relative to A-shares, thereby amplifying deviations from parity pricing. If their assertion holds true, deviations from parity pricing and the magnitude of mispricing should be more pronounced when cross-asset arbitrage opportunities are limited. Consequently, our alternative hypotheses are that Δ, PD, LPD, RAD_k, or DF will be higher, and that RD_k will deviate more substantially from zero, in the WO treatment than the W treatment.

Finally, based on evidence of mispricing decreasing as participants repeat the sequence (Smith et al., 1988; Dufwenberg et al., 2005), we anticipate a reduction in mispricing in later sequences as traders gain experience. Thus we hypothesize:

Hypothesis 3 *The values of Δ, PD, LPD, RD_A, RD_B, RAD_A, RAD_B and DF will converge towards 0 over sequence.*

Figure 1: The dynamic of Δ of each sequence in each treatment



4 Results and Discussions

4.1 Price Discrepancy between twin shares

Figure 1 presents the dynamics of Δ for each treatment across sequences. Specifically, the three panels on the left display the evolution of Δ in the W treatment for each sequence, while the three panels on the right do the same for the WO treatment. Each solid line represents the trajectory of Δ in an individual market.

The figure shows the greater dispersion of Δ in the WO treatment than in the W treatment. However, it is not visually apparent which treatment exhibits a higher average value of Δ. To obtain a more accurate assessment, we conduct statistical comparisons using the mean values of Δ, as well as the PD and LPD.

Table 2: The results of the Mann–Whitney tests conducted between treatments across Δ, PD, LPD.

Data	Our study			Charness and Neugebauer(2019)		
Treatments	W (n=7)	WO (n=8)	W = WO^a	PC (n=12)	NC (n=8)	PC = NC^b
Panal A. The average value of Δ						
Sequence 1	0.062**	0.060	p = 0.768	0.070	0.277**	p = 0.031
Sequence 2	0.035	0.045	p = 0.748	0.017	0.299**	p = 0.001
Sequence 3	0.044**	0.033	p = 0.567	0.035	0.179**	p = 0.165
Sequence 4	-	-	-	0.026	0.145**	p = 0.064
Total	0.047**	0.046	p = 0.694	0.037	0.225**	p = 0.006
Panal B. The average value of PD						
Sequence 1	0.087	0.173	p = 0.047	0.229	0.398	p = 0.124
Sequence 2	0.065	0.102	p = 0.116	0.137	0.301	p = 0.014
Sequence 3	0.049	0.091	p = 0.060	0.138	0.233	p = 0.198
Sequence 4	-	-	-	0.151	0.204	p = 0.082
Total	0.067	0.122	p = 0.076	0.164	0.291	p = 0.082
Panal C. The average value of LPD						
Sequence 1	0.082	0.159	p = 0.036	-	-	-
Sequence 2	0.064	0.096	p = 0.116	-	-	-
Sequence 3	0.046	0.088	p = 0.060	-	-	-
Total	0.064	0.114	p = 0.027	-	-	-

a: p-values are based on the one-tailed, two-sample Mann-Whitney tests. H_1: WO > W.
b: p-values are based on the two-tailed, two-sample Mann-Whitney tests
*, **, and *** indicate a significant difference from 0 at the 10, 5, and 1% significance levels, respectively, based on the two-tailed, one-sample Mann-Whitney tests.

Observation 1 *There is no statistically significant difference between the WO and W treatments in Δ. However, PD and LPD are greater in the WO treatment than in the W treatment.*

Support: The left side of Table 2 presents the comparative results for Δ, PD, and LPD. The second and third columns report the mean values of each measure for W and WO treatments, respectively. The p-values shown in the fourth column are derived from one-tailed Mann-Whitney tests. Additionally, superscripts * , ** and *** next to the mean values indicate whether they are significantly different from zero at the 10%, 5%, and 1% significance levels, repectively. The fifth through seventh columns show the results of Charness and Neugebauer (2019) for comparison.

Panel A compares the average value of Δ. We find no statistically significant differences in the average value of Δ between treatments, neither within individual sequences nor across all sequences combined. However, consistent with the visual findings depicted in Figure 1, the data indicate some statistically significant differences between treatments in the magnitude of deviations from parity pricing.[6]

Specifically, Panel B demonstrates that the PD values in the WO treatment are significantly higher, at 5% level, than those in the W treatment in Sequence 1 (0.087 in W vs. 0.173 in WO, $p = 0.047$).[7] Similarly, Panel C reveals that LPD values in the WO treatment are significantly higher than those in the W treatment in Sequence 1 (0.082 in W vs. 0.159 in WO, $p = 0.036$) and overall (0.064 in W vs. 0.114 in WO, $p = 0.027$).[8]

To provide robustness checks for these findings, Table 3 also reports linear regression results controlling for participants' risk attitudes and sequence numbers. The regression results are consistent with the two-sample tests: no statistically significant difference emerges for the average value of Δ, whereas some statistically significant differences appear for PD and LPD between treatments. Specifically, in Models (1) and (2), with the average value of Δ as the dependent variable, coefficients on the

[6]When analyzing the PD and $LPDs$, the data from one group in the W treatment were flagged as an outlier by the Smirnov-Grubbs test. As a result, only the data from 15 groups were included in all subsequent analyses. Note that this exclusion criterion was not pre-registered. During the analysis, however, we observed significant deviation in the pricing within a specific group from the W treatment, where, for example, the average LPD for this group across three sequences was 0.241. This value is 2.8 times the average LPD of 0.086 for the W treatment, and 2.2 times higher than the second highest observation (= 0.108) within the treatment (Smirnoff-Grubbs test: t=2.21, p=0.011). Therefore, we excluded this outlier from the analysis presented in the main body of the paper to avoid biasing the results. We report results without dropping this outlier in Online Appendix B.

[7]Below, statistical significance refer to $p < 0.05$. It is also marginally significantly so in Sequence 3 (0.049 in W vs. 0.091 in WO, $p = 0.060$) and across all sequences combined (0.067 in W vs. 0.122 in WO, $p = 0.076$).

[8]It is also marginally significantly so in Sequence 3 (0.046 in W vs. 0.088 in WO, $p = 0.060$). If we employ a two-sided Mann–Whitney test, the differences in PD between treatments are marginally significantly different from zero in Sequence 1 ($p = 0.094$), and the differences in LPD between treatments are marginally significantly different from zero in Sequence 1 ($p = 0.072$) and overall ($p = 0.054$).

Table 3: The regression on Δ, PD, and LPD

Dependent variable	Average value of Δ		PD		LPD	
Models	(1)	(2)	(3)	(4)	(5)	(6)
Intercept	0.078	0.070	0.014	-0.009	-0.011	-0.030
	(0.542)	(0.548)	(0.366)	(0.366)	(0.293)	(0.294)
WO	0.003	0.018	0.059**	0.103*	0.054**	0.090**
(=1: if WO)	(0.045)	(0.072)	(0.028)	(0.056)	(0.024)	(0.044)
The average score	-0.0001	-0.0001	0.002	0.002	0.003	0.003
on risk attitude	(0.011)	(0.011)	(0.008)	(0.008)	(0.006)	(0.006)
Sequence	-0.012	-0.009	-0.030***	-0.019***	-0.027***	-0.018***
	(0.009)	(0.010)	(0.009)	(0.006)	(0.007)	(0.005)
WO×	-	-0.007	-	-0.022	-	-0.018
Sequence	-	(0.018)	-	(0.016)	-	(0.013)
Adjusted R^2	-0.057	-0.083	0.200	0.196	0.245	0.241
Observations			45			

Note: The independent variable, "The average score on risk attitude" quantifies the average risk aversion level of each group, with a lower score indicating a higher degree of risk aversion. Robust standard errors, enclosed in parentheses, are clustered by group. In both panels, *, **, and *** indicate a significant difference from 0 at the 10, 5 and 1% significance levels.

WO dummy variable are not statistically significantly different from zero. In contrast, Model (3), using PD as the dependent variable, indicates a significant positive coefficient for WO, although in Model (4) it is only marginally significant. Similarly, Models (5) and (6), employing LPD as the dependent variable, yield statistically significant positive coefficients on the WO dummy variable. These findings provide evidence of a moderate treatment effect, suggesting that restricting arbitrage opportunities across different assets may increase the magnitude of deviations from price parity.

Observation 2 Δ *does not significantly change across repeated sequences, whereas* PD *and* LPD *declines across sequences in both treatments.*

Support: In Table 3, Models (1) and (2)—which use the average value of Δ as the dependent variable—indicate that the coefficients for "Sequence" are not statistically significantly different from zero. These results suggest no evidence that Δ converges toward zero as participants gain more trading experience.

In contrast, the coefficients for "Sequence" in Models (3) through (6)—which use PD and LPD as dependent variables—are significantly negative at the 1% significance level. Additionally, the interaction terms between the WO dummy and "Sequence" exhibit negative but statistically insignificant coefficients in Model (4) and Model (6). These findings indicate that the magnitude of absolute deviations from parity pricing, as measured by PD and LPD tend to decrease and move toward zero as participants accumulate trading experience. This implies that price discrepancies diminish to some extent as subjects become more experienced in trading, across both treatments.

It should be noted that in the comparative results for PD and LPD, our data indicate that the differences between treatments become statistically insignificant in Sequence 2 but regain marginal significance in Sequence 3. Such a pattern appears counterintuitive if one expects treatment effects to gradually weaken with participants' increased trading experience. It is possible that shortening of the trading duration between Sequences 1 and 2, i.e., from 180 seconds in Sequence 1 to 90 seconds in Sequence 2, have interacted with participants learning.

It is pertinent to highlight that the configuration of the W treatment in our study closely mirrors the *Perfect Correlation* (PC) treatment described by Charness and Neugebauer (2019). The primary distinction between these treatments lies in the initial allocation of shares: in our W treatment, half of the traders receive 6 units of A-shares, while the other half receive 3 units of B-shares at the beginning of each sequence. In contrast, in the PC, each trader is initially endowed with 2 units of A-shares and 2 units of B-shares. Additionally, the sole difference between the *No Correlation* (NC) treatment and the PC treatment is that, in the former, the future dividend processes of A-shares and B-shares are independent.

Charness and Neugebauer (2019) report a higher degree of price discrepancy in the NC treatment relative to the PC treatment. They interpret this treatment effect as arising from the lack of risk-free arbitrage opportunities between the two types of shares in the NC treatment, which may impede participants' cross-asset arbitrage activities, subsequently exacerbating price discrepancies. Given that our experimental setup and interpretation of treatment effects closely align with theirs, a comparison of our results is meaningful. The results regarding the measures of price discrepancy between the PC and NC treatments reported in Charness and Neugebauer (2019) are presented on the right side of Table 2.

Firstly, Charness and Neugebauer (2019) report a different pattern in Δ across treatments compared to our findings. In their experiment, only the NC treatment exhibits a high average value of Δ (0.277 in Sequence 1, 0.299 in Sequence 2, 0.145 in Sequence 3, 0.145 in Sequence 4, and 0.255 overall), which is significantly higher than zero and than the corresponding values in the PC treatment (0.070 in Sequence 1, 0.017 in Sequence 2, 0.035 in Sequence 3, 0.026 in Sequence 4, and 0.037 overall).

In contrast, in our experiment, the average value of Δ in the W treatment (0.062 in Sequence 1, 0.035 in Sequence 2, 0.044 in Sequence 3, and 0.047 overall) is significantly higher than zero in Sequences 1, 3, and overall, whereas the corresponding values in the WO treatment (0.060 in Sequence 1, 0.045 in Sequence 2, 0.033 in Sequence 3, and 0.046 overall) are not significantly different from zero or from those in the W treatment. Furthermore, the Δ values in the W treatment are not significantly different from those in the PC treatment reported by Charness and Neugebauer (2019). We consider that the differences in the results regarding Δ between Charness and Neugebauer (2019) and our study may stem from variations in the experimental settings between the WO and NC treatments. Specifically, the correlation of the future dividend processes between the two shares may influence the average value of Δ, independent of the effect of cross-asset arbitrage.

Secondly, our data reveal a similar pattern in PD to that observed by Charness and Neugebauer (2019). Specifically, PD is significantly higher in the WO treatment than in the W treatment in Sequences 1 and and overall, and marginally so in Sequence 3. Similarly, in Charness and Neugebauer (2019), PD is significantly higher in the NC treatment than in the PC treatment in Sequences 2, 4, and overall, and marginally so for Sequence 1.[9]

However, compared to the findings of Charness and Neugebauer (2019), our data exhibit lower average values of PD. Particularly noteworthy is that the average PD values for the W treatment are less than half of those for the PC treatment, despite the near-identical settings of these treatments. This difference can be attributed to the lower variance in the average Δ values observed in our experiment, indicating a narrower fluctuation range around zero compared to the results reported by Charness and Neugebauer (2019).

In addition, Angerer et al. (2023), who investigate the relationship between algorithmic arbitrage and the law of one price, employ a baseline setup that is similar to our W treatment and the PC treatment of Charness and Neugebauer (2019).[10] They report that the average PD value across all three sequences is 0.320, which is nearly five times higher than that observed in our experiment.

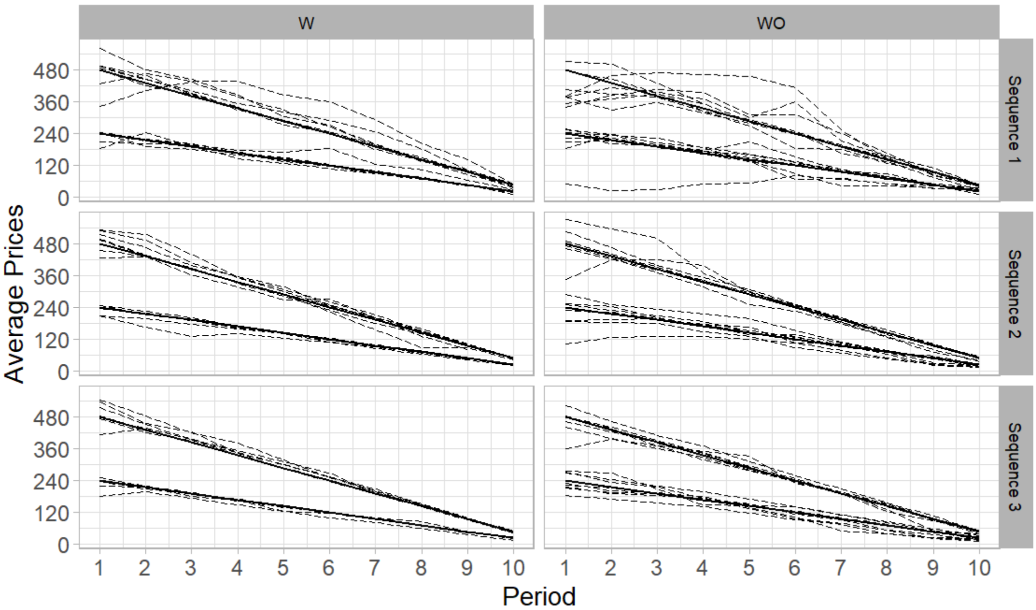
4.2 Bubble Magnitude

Figure 2 illustrates the price dynamics of each type of share across periods. In each panel, the lower and upper solid lines represent the fundamental values of A-shares

[9]Note that we employ a one-tailed test to compare PD between treatments, whereas Charness and Neugebauer (2019) use a two-tailed test. Consequently, the observed treatment effect in PD appears larger in Charness and Neugebauer (2019) than in our study.

[10]Compared to the setting of the W treatment, the only difference in the baseline of Angerer et al. (2023) lies in the participants' endowments: each participant is endowed with 4 units of A-shares, 4 units of B-shares, and 1,300 cents of experimental currency.

Figure 2: Price dynamic of each sequence in each treatment



and B-shares, respectively. The dashed lines depict the price dynamics of each share type in each market.

As shown in the figure, the trading prices of both share types in both treatments tend to deviate more from their fundamental values in Sequence 1 than in Sequences 2 and 3. In particular, as indicated by the dashed lines in the upper portions of the panels, B-shares appear to be consistently overpriced in both treatments, though these deviations tend to diminish and converge toward the fundamentals as the sequence progresses.

For A-shares, represented by the dashed lines in the lower parts of the panels, although some deviations from the fundamental values are observed, there is no clear indication of systematic overpricing or underpricing. Similar to B-shares, the deviations of A-share prices also tend to converge toward the fundamentals over time in both treatments. However, the magnitude of these deviations is generally larger in the WO treatment compared to the W treatment. To statistically assess these observed trends, we conduct analyses using the measures of RD, RAD, and DF in the subsequent sections.

Observation 3 *There is no statistically significant difference in RD_A and RD_B across treatments.*

Support: As reported on the left side of Panels A and B in Table 4, no statistically significant differences in RD_A and RD_B are observed between the W and WO treatments in any sequence. In addition, there is no evidence that RD_A or RD_B significantly differ from zero in either the W or WO treatment, except for RD_B in the W treatment in Sequence 1, which is marginally significantly greater than zero. These results suggest that neither type of share is systematically overpriced or underpriced in the W or WO treatments.

These observations differ from the findings of Charness and Neugebauer (2019)—as shown on the right side of Panel B—who find that B-shares tend to be significantly overpriced by participants when the risk-free cross-asset arbitrage opportunity does not exist (i.e., in the NC treatment).

Observation 4 *RAD_A and DF is elevated in the WO treatment compared to the W treatment.*

Support: Left-side of Panel C of Table 4 reveals that the median RAD_A is statistically significantly higher in the WO treatment than in the W treatment in all the sequences.[11] These results suggest that restricting arbitrage opportunities increases

[11] 0.060 in W vs. 0.171 in WO, $p = 0.047$, in Sequence 1, 0.055 in W vs. 0.127 in WO, $p = 0.036$,

Table 4: The results of the Mann–Whitney tests conducted between treatments across RD, RAD, DF

Data	Our study			Charness and Neugebauer(2019)		
Treatments	W (n=7)	WO (n=8)	W = WO[a]	PC (n=12)	NC (n=8)	PC = NC[b]
Panal A. The average value of RD_A						
Sequence 1	0.002	-0.070	p =0.694	0.028	0.002	-
Sequence 2	-0.045	-0.036	p =0.955	-0.055	-0.050	-
Sequence 3	-0.028	-0.021	p =0.955	-0.110**	-0.059	-
Sequence 4	-	-	-	0.037	-0.055	-
Total	-0.024	-0.042	p =0.955	-0.025	-0.040	-
Panal B. The average value of RD_B						
Sequence 1	0.068*	0.007	p =0.336	0.095	0.221**	-
Sequence 2	0.024	0.014	p =1.000	-0.022	0.225***	-
Sequence 3	0.029	-0.003	p =0.281	0.024	0.126**	-
Sequence 4	-	-	-	0.023	0.114**	-
Total	0.040	0.006	p =0.281	0.023	0.172**	-
Panal C. The average value of RAD_A						
Sequence 1	0.060	0.170	p =0.047	-	-	-
Sequence 2	0.055	0.127	p =0.036	-	-	-
Sequence 3	0.039	0.122	p =0.005	-	-	-
Total	0.051	0.140	p =0.020	-	-	-
Panal D. The average value of RAD_B						
Sequence 1	0.104	0.131	p =0.268	-	-	-
Sequence 2	0.061	0.058	p =0.693	-	-	-
Sequence 3	0.049	0.091	p =0.567	-	-	-
Total	0.072	0.078	p =0.478	-	-	-
Panal E. The average value of DF						
Sequence 1	0.089	0.151	p =0.027	0.246	0.395	p =0.320
Sequence 2	0.059	0.098	p =0.095	0.153	0.304	p =0.045
Sequence 3	0.043	0.107	p =0.007	0.158	0.223	p =0.320
Sequence 4	-	-	-	0.217	0.174	p =0.758
Total	0.064	0.118	p =0.027	0.193	0.274	p =0.563

a: p-values are based on the two-tailed, two-sample Mann-Whitney tests. H_1: WO > W, for RAD and DF. For RD, they are based on two-tailed, two-sample Mann-Whitney tests.
b: p-values are based on the two-tailed, two-sample Mann-Whitney tests
*, **, and *** indicate a significant difference from 0 at the 10, 5, and 1% significance levels, respectively. p-values are based on the two-tailed, one-sample Mann-Whitney tests.

the magnitude of mispricing of A-shares, which are characterized by relatively lower fundamentals. In contrast, Panel D shows no significant differences in the RAD_B, indicating that restricting arbitrage does not affect the pricing of B-shares, which are characterized by relatively higher fundamentals.

Additionally, the left side of Panel E indicates that the DF is statistically significantly higher in the WO treatment compared to the W treatment in Sequences 1 and 3, as well as Overall. For Sequence 2, the difference is marginally significant.[12] These results suggest that constraining arbitrage opportunities increase the price deviations from fundamentals for the entire market.

Table 5 reports the linear regression results for RD, RAD, and DF, controlling for participants' risk attitudes and sequence numbers, and provides robustness checks for these findings. Specifically, the dependent variable is RD_A in Models (1) and (2), RD_B in Models (3) and (4), RAD_A in Models (5) and (6), RAD_B in Models (7) and (8), and DF in Models (9) and (10), respectively.

Consistent with the results of the two-sample Mann–Whitney tests, the coefficients of the WO dummy are not significantly different from zero in Models (1) through (4), (6) to (8), and (10), suggesting that there is no significant treatment effect on RD_A, RD_B, or RAD_B. In contrast, the coefficients of the WO dummy in Model (5) and Model (9) are positive and significantly greater than zero, suggesting that restricting cross-asset arbitrage may increase the magnitude of absolute deviations of asset prices from fundamentals.

in Sequence 2, 0.039 in W vs. 0.122 in WO, $p = 0.005$, in Sequnce 3, and 0.051 in W vs. 0.140 in WO, $p = 0.020$, over all. The p-values are based on one-sided Mann-Whitney test. If we employ two-sided Mann-Whitney tests, these p-values become, $p = 0.094$, $[= 0.072$, $p = 0.009$, and $p = 0.040$, for Sequences 1, 2, 3, and Overall, respectively

[12] 0.089 in W vs. 0.151 in WO, $p = 0.027$, for Sequence 1, 0.059 in W vs. 0.098 in WO, $p = 0.095$, for Sequence 2, 0.043 in W vs. 0.107 in WO, $p = 0.007$, for Sequence 3, and 0.064 in W vs. 0.118 in WO, $p = 0.027$, overall. If we employ a two-sided Mann-Whitney test, these p-values become $p = 0.054$, $p = 0.189$, $p = 0.014$, and $p = 0.054$ for Sequences 1, 2, 3, and overall, respectively.

Table 5: The regression on RD, RAD, and DF

Dependent variable	RD^A		RD^B		RAD^A		RAD^B		DF	
Models	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Intercept	0.005	0.048	-0.067	-0.052	-0.051	-0.065	0.058	0.042	0.0001	0.002
	(0.647)	(0.683)	(0.501)	(0.506)	(0.467)	(0.496)	(0.389)	(0.382)	(0.333)	(0.333)
WO	-0.020	-0.099	-0.030	-0.058	0.093**	0.120	0.009	0.039	0.059**	0.056
(=1: if WO)	(0.069)	(0.132)	(0.032)	(0.062)	(0.054)	(0.116)	(0.020)	(0.048)	(0.025)	(0.053)
The average score	-0.001	-0.001	0.003	0.003	0.003	0.003	0.002	0.002	0.002	0.002
on risk attitude	(0.140)	(0.014)	(0.011)	(0.011)	(0.010)	(0.010)	(0.008)	(0.008)	(0.007)	(0.007)
Sequence	0.006	-0.015*	-0.012	-0.019**	-0.018	-0.011	-0.036***	-0.028***	-0.023**	-0.023*
	(0.021)	(0.008)	(0.010)	(0.008)	(0.019)	(0.012)	(0.009)	(0.010)	(0.009)	(0.014)
WO×	-	0.040	-	0.014	-	-0.013	-	-0.015	-	0.001
Sequence	-	(0.039)	-	(0.019)	-	(0.036)	-	(0.017)	-	(0.019)
Adjusted R^2	-0.066	-0.076	0.022	0.006	0.129	0.110	0.152	0.141	0.212	0.192
Observations	45									

Note: The independent variable, "The average score on risk attitude" quantifies the average risk aversion level of each group, with a lower score indicating a higher degree of risk aversion. Robust standard errors, enclosed in parentheses, are clustered by group. In both panels, *, **, and *** indicate a significant difference from 0 at the 10, 5 and 1% significance levels.

Our results regarding DF, on the one hand, differ somewhat from those of Charness and Neugebauer (2019), as reported on the right side of Panel E in Table 4. They find that the lack of risk-free cross-asset arbitrage opportunities does not strongly increase DF, except for Sequence 2 in which the difference between PC and NC treatments is significant. Therefore, they suggest that constrained risk-free arbitrage opportunities across shares are not linked to mispricing from fundamentals for the entire market.

On the other hand, Angerer et al. (2023) provide results regarding RAD that share some similarities with ours. They find that in certain specific situations, an increase in cross-asset arbitrage activity may reduce RADs. For example, when algorithmic arbitrageurs, who always submit a delayed matching limit order following a subject-initiated order in the other market, participate in trading, the RAD decreases compared to when they are absent.

Observation 5 *RAD_B and DF decline across sequences in both treatments, whereas RD_A, RD_B, and RAD_A remain relatively stable.*

Support: As shown in Table 5, the coefficients of "Sequence" in Models (7) through (9) (-0.036 in Model (7), -0.028 in Model (8), and -0.023 in Model (9)) are significantly negative. And that for Model (10), -0.023, is also marginally significantly negative. These results imply that the RAD_B and DF tend to decrease and converge toward zero as participants accumulate trading experience.

In contrast, the coefficients "Sequence" in Models (5) and (6) do not differ significantly from zero, suggesting that RAD_A may not be affected by participants' increasing trading experience. Additionally, Models (1) and (3) show similar results for "Sequence," suggesting no significant correlation between the degree of RD and participants' trading experience.[13]

Table 6 summarizes the predictions and the main outcomes reported in Subsections 4.1 and 4.2.

4.3 Further Analysis of Treatment Effects

In this subsection, we delve deeper into examining why limiting arbitrage opportunities results in a higher level of price discrepancy. Note that the analyses presented in this subsection are exploratory.

We consider that if the treatment effect on the degree of price discrepancy is driven by the lack of cross-asset arbitrage, then the data should exhibit two features. First, the higher the degree of price discrepancy in the market, the greater the amount of unrealized cross-asset arbitrage profits present in the outstanding orders. Second, in each sequence of each session under the W treatment, cross-asset arbitrage behavior should be observed.

We first examine the relationship between the treatment effect and the ratio of unrealized cross-asset arbitrage profits (*UCAP-ratio*), computed from outstanding orders. For group g in sequence r,

$$UCAP\text{-}ratio^{r,g} = \frac{\sum_{t=1}^{T} \left(profit_{AB,t}^{r,g} + profit_{BA,t}^{r,g} \right)}{Expected\ capital\ per\ participant}, \tag{7}$$

where $profit_{AB,t}^{r,g}$ ($profit_{BA,t}^{r,g}$) denotes the arbitrage profit obtainable from outstanding orders in period t of sequence r in group g by selling (buying) A-shares and buying (selling) B-shares. *Expected capital per participant* is defined as the sum of expected

[13]Note that although Model (4) reports significantly negative coefficient, and Model (2) marginally significantly negative coefficient, for "Sequence," we do not rely on these results to interpret the relationship between the degree of RD and participants' trading experience, since the adjusted R^2 values are quite low (-0.076 in Model (2) and 0.006 in Model (4)), indicating very weak explanatory power.

Table 6: The summary of predictions and outcomes

Predictions	**Outcomes**
Hypohesis 1 There are no statistically significant differences in Δ, PD, or LPD between the W treatment and the WO treatment. **The alternative Hypothesis** Δ, PD, and LPD will be higher in the WO treatment compared to the W treatment.	No statistically significant difference in Δ is observed between the W and WO treatments. PD and LPD are slightly but significantly higher in the WO treatment than in the W treatment in Sequence 1 and overall (**Hypothesis 1** is rejected, and **the alternative hypothesis** is partially supported.)
Hypohesis 2 There are no statistically significant differences in RD_k, RAD_k, and DF between the W and WO treatments. **The alternative Hypothesis** RAD_A, RAD_B or DF will be higher in the WO treatment compared to the W treatment, and that RD_A, or RD_B will deviate more substantially from zero in the WO treatment than in the W treatment.	No statistically significant differences in RD_A, RD_B and RAD_B are observed between the W and WO treatments. RAD_A is significantly higher in the WO treatment than in the W treatment across all sequences, and DF is slightly but significantly higher in the WO treatment than in the W treatment in Sequences 1 and 3, and over all. (**Hypothesis 2** is rejected, and **the alternative hypothesis** is partially supported.)
Hypohesis 3 The values of Δ, PD, RD_A, RD_B, RAD_A, RAD_B, and DF will converge towards zero over sequence.	The PD, LPD, RAD_B, and DF converge toward zero over the sequences. (**Hypothesis 3** is partially supported.)

Table 7: Comparison of unrealized cross-asset arbitrage profit ratios (UCAP-ratios) between Treatments

	W ($n = 8$)	WO ($n = 8$)	The p-values (H_0: W = WO)
Sequence 1	0.025	0.027	0.347
Sequence 2	0.013	0.024	0.076*
Sequence 3	0.024	0.025	0.567

Note: p-values are based on the one-tailed, two-sample Mann-Whitney tests. H_1: WO > W.

dividends and initial cash endowments per participant, which equals 2,640 cents. Specifically,

$$profit^{r,g}_{AB,t} = \sum_m \Gamma_{profit}\left[\left(ask^{r,g,m}_{out,B,t} - bid^{r,g,m}_{out,A,t}\right) - (FV_{B,t} - FV_{A,t})\right], \tag{8}$$

and

$$profit^{r,g}_{BA,t} = \sum_n \Gamma_{profit}\left[(FV_{B,t} - FV_{A,t}) - \left(bid^{r,g,n}_{out,B,t} - ask^{r,g,n}_{out,A,t}\right)\right], \tag{9}$$

where $ask^{r,g,m}_{out,B,t}$ ($bid^{r,g,m}_{out,A,t}$) is the m-th lowest (highest) outstanding ask (bid) for B-shares (A-shares), and $bid^{r,g,n}_{out,B,t}$ ($ask^{r,g,n}_{out,A,t}$) is the n-th highest (lowest) outstanding bid (ask) for B-shares (A-shares), in period t of sequence r in group g. The indicator Γ_{profit} equals 1 if the expression in brackets is positive and 0 otherwise.

Observation 6 *Markets with a high UCAP-ratio exhibit high levels of PD and LPD. When UCAP-ratio is included as an independent variable in regressions on PD and LPD, the coefficients on the WO dummy become insignificant.*

Support: Table 7 compares UCAP-ratios between treatments, and Table 8 presents the corresponding regression results. From Table 7, we do not observe statistically significant difference in UCAP-ratios between treatments: the UCAP-ratio is only marginally significantly higher in the WO treatment than in the W treatment in Sequence 2.

Table 8: Regression results: UCAP-ratio and mispricing measures.

Dependent variable	PD		LPD		RAD_A		RAD_B		DF	
Models	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)	(9)	(10)
Intercept	0.171	0.004	0.124	-0.006	0.010	-0.182	0.229	0.136	0.125	-0.006
	(0.364)	(0.332)	(0.290)	(0.270)	(0.452)	(0.536)	(0.313)	(0.309)	(0.301)	(0.288)
WO	0.044	0.022	0.042*	0.025	0.088*	0.068*	-0.006	0.002	0.047*	0.051**
(=1: if WO)	(0.029)	(0.029)	(0.024)	(0.023)	(0.052)	(0.036)	(0.021)	(0.027)	(0.024)	(0.021)
UCAP-ratio	1.617**	4.267***	1.389***	3.476***	0.632	4.004**	1.763**	4.328***	1.290**	4.490***
	(0.622)	(1.365)	(0.484)	(1.049)	(0.637)	(1.691)	(0.653)	(1.138)	(0.570)	(1.151)
The average score	-0.002	0.0003	-0.001	0.001	0.001	0.003	-0.003	-0.002	-0.001	0.0001
on risk attitude	(0.007)	(0.007)	(0.006)	(0.006)	(0.010)	(0.011)	(0.006)	(0.006)	(0.006)	(0.006)
Sequence	-0.029***	0.013	-0.026***	0.007	-0.017	0.033**	-0.034***	-0.003	-0.022	0.019*
	(0.006)	(0.016)	(0.005)	(0.012)	(0.018)	(0.016)	(0.008)	(0.011)	(0.008)	(0.011)
UCAP-ratio×WO	-	1.091	-	0.818	-	0.954	-	-0.365	-	-0.137
	-	(1.221)	-	(0.927)	-	(2.169)	-	(1.245)	-	(1.031)
UCAP-ratio×Sequence	-	-1.664**	-	-1.298**	-	-1.991*	-	-1.199**	-	-1.587***
	-	(0.756)	-	(0.563)	-	(1.166)	-	(0.458)	-	(0.466)
Adjusted R^2	0.323	0.396	0.378	0.440	0.118	0.145	0.361	0.425	0.313	0.429
Observations					45					

Note: The independent variable, "The average score on risk attitude" quantifies the average risk aversion level of each group, with a lower score indicating a higher degree of risk aversion. Robust standard errors, enclosed in parentheses, are clustered by group. In both panels, *, **, and *** indicate a significant difference from 0 at the 10, 5 and 1% significance levels.

However, Table 8 shows UCAP-ratios are significantly positively correlated with both PD and LPD. Specifically, in Models (1) and (2), where the dependent variable is PD, the coefficients are 1.617 and 4.267, respectively—both significantly greater than zero. Similarly, in Models (3) and (4), where the dependent variable is LPD, the coefficients are 1.389 and 3.476, respectively, and are also significantly greater than zero.

Furthermore, when comparing these results with those in Table 3, which reports regressions on PD and LPD without including UCAP-ratio as an independent variable, we find that WO coefficients loses significance in Models (1), (2), and (4). In Model (3), it becomes only marginally significant.

These results suggest that the treatment effect on PD and LPD is mediated by the lack of cross-asset arbitrage.

Observation 7 *Although markets with a high UCAP-ratio tend to exhibit elevated levels of* RAD_A, RAD_B, *and* DF, *the* WO *dummy remains significantly positive even after controlling for UCAP-ratio for* RAD_A *and* DF *in the regression models.*

Support: Table 8 shows that the coefficients of the UCAP-ratio in Models (6) to (10) are positive and significant. Nevertheless, the treatment dummy still yields at least marginally significantly positive coefficients in Models (6), (9), and (10). These results suggest that although a high UCAP-ratio is associated with a greater RAD_A, RAD_B and DF, it does not fully account for the treatment effect on RAD_A and DF.

Observation 8 *A fair amount of cross-assets arbitrage behavior was observed in the* W *treatment.*

Support: We subsequently examine whether and how frequently cross-asset arbitrage behavior occurs in the W treatment. Our results indicate that, on average, approximately 17.8% of all successful transactions are identified as hedging trades—such as participants securing risk-free profits by selling an A-share and purchasing a B-share during the same period, or vice versa—corresponding to an average of 9.5 instances, or 19 successful transactions, per sequence.

The average frequency of cross-asset arbitrage is highest in Sequence 1 at 24.3% (19.4 instances), but decreases to 10.0% (4 instances) in Sequence 2 and 12.0% (4.8 instances) in Sequence 3. Notably, substantial cross-market heterogeneity is observed, with the lowest frequency being 0% (only in one market) and the highest reaching 41% (57 instances). **Observations 7** and **8** imply that the degree of price discrepancy may increase due to the lack of cross-asset arbitrage.

These results contrast with the findings of Charness and Neugebauer (2019), who observed that such hedging trades occurred only about 1% of the time when a discrepancy arose in the PC treatment, and that 99% of discrepancies were eliminated by independent trades in the market rather than by individuals engaging in arbitrage. They suggest that the existence of a perfect positive correlation in the dividend processes between twin shares may enhance traders' acuity, thereby helping to reduce the degree of price discrepancy and mispricing relative to fundamentals.

However, our data do not contradict the explanation proposed by Charness and Neugebauer (2019). For example, as shown by the regression results reported in Table 8, the UCAP-ratio does not fully account for the treatment effects on RAD_A and DF. Thus. similar mechanisms suggested by Charness and Neugebauer (2019) may still operate.

5 Additional experiment

In our experimental design, participants in the *WO* treatment can trade only one of the two share types, whereas participants in the *W* treatment can trade both. Consequently, as the Active Participation Hypothesis (Lei et al., 2001) suggests, participants in the former may engage in speculative trading to a greater extent than those in the latter. Such excess trading could, in turn, exacerbate price bubbles and reduce market efficiency.[14]

Accordingly, the observed treatment effect may reflect not the presence of arbitrage opportunities across assets, but rather overtrading driven by speculative motives. To assess this possibility, we follow Lei et al. (2001)'s approach and extend the *WO* treatment so that, in addition to an asset, participants can also trade a "good" in a third market. We refer to this additional treatment as *WOT* (i.e., *WO* with a Third market).

5.1 Experimental setting and data collection

In the *WOT* treatment, the 14 participants in each group are first assigned equally to two trader types: A-type traders, who can trade only A-shares, and B-type traders, who can trade only B-shares. In addition, each participant is randomly assigned with equal probability to be either a seller or a buyer in an additional market for a "good." Consequently, four trader categories exist in the market: (i) sellers of the good who can trade only A-shares, (ii) buyers of the good who can trade only A-shares, (iii) sellers of the good who can trade only B-shares, and (iv) buyers of the good who can trade only B-shares.

Unlike the shares, the goods do not generate dividends. Instead, participants can

[14] We thank an anonymous reviewer for pointing this out.

earn experimental currency (cents) by trading the good. Specifically, buyers obtain utility from purchasing the good in cents: for each unit purchased, a buyer's payoff equals the unit's value minus the transaction price. Sellers, in contrast, purchase (procure) the good by paying a unit cost in cents and then sell it in the market; for each unit sold, a seller's payoff equals the transaction price minus the unit cost.

For each participant, the cents used in the good market are held in the same account as the cents used in the share markets. Thus, dividend income from shares can be used to trade goods, and earnings from trading goods can be used to trade shares.

The good is traded via a continuous double auction, as in the share markets, and trading occurs simultaneously with share trading. In each period, each buyer and each seller can purchase or sell up to seven units of the good. At the end of each period, buyers' purchased quantities and sellers' sold quantities are reset. Buyers' marginal values are decreasing, whereas sellers' marginal costs are increasing. Concretely, buyers' marginal values for units 1 through 7 are 780, 730, 690, 670, 630, 600, and 570 cents, respectively, and sellers' marginal costs for units 1 through 7 are 560, 620, 670, 680, 720, 750, and 780 cents, respectively. Under these parameters, the competitive equilibrium price lies between 670 and 690 cents, and the equilibrium per-capita trading quantity is three units.[15]

The experiment was conducted between October 3 and November 7, 2025, using zTree. Participants are students of Osaka University who have not participated in our original set of experiments.[16] As in the W and WO treatments, the WOT treatment comprises eight groups with 112 participants in total.

[15] The English translation of the instruction can be downloaded from https://zenodo.org/records/18521815.

[16] This experiment was pre-registered on AsPredicted on September 27, 2025 (AsPredicted $249,213$).

5.2 Hypotheses

Under the above design, trading the additional "good"—unlike trading the shares—can generate direct profits for participants. We therefore expect participants to trade in the good market, which should reduce speculative trading activity in the twin-share markets in the WOT treatment relative to the WO treatments. This yields the following hypothesis:

Hypothesis 4 (Trading volume) *(1) Per-asset and total trading volumes in the share markets are lower in WOT than in WO; (2) Trading volume in WOT is not higher than in W treatment.*

Consequently, if the main treatment effect is driven by variation in trading volume arising from speculative motives, as suggested by the Active Participation Hypothesis, it implies the following hypotheses:

Hypothesis 5a (Volume-driven explanation for cross-market price discrepancy) *If volume is the driver, the cross-market price discrepancy between the two original shares in WOT is lower than in WO treatment.*

Hypothesis 6a (Volume-driven explanation for efficiency) *If treatment effects are driven by differences in trading volume, then mispricing relative to fundamentals in WOT is lower than in WO treatment.*

By contrast, if the main treatment effect in our study is driven by the presence or absence of arbitrage opportunities across assets, we derive the following hypotheses:

Hypothesis 5b (Arbitrage-driven explanation for cross-market price discrepancy) *If arbitrage is the driver, the discrepancy in WOT exceeds that in W.*

Table 9: Comparison of trading volume between the W and WOT treatments, and between the WO and WOT treatments.

	A-shares		B-shares		Total	
	W-WOT	WO-WOT	W-WOT	WO-WOT	W-WOT	WO-WOT
Sequence 1	79-37	85-37	62-25	43-25	141-62	129-62
	=42***	=48***	=37***	=18***	=79***	=67***
Sequence 2	38-31	43-31	33-16	21-16	72-47	64-47
	=7	=12*	=17***	=5	=25**	=17**
Sequence 3	46-33	42-33	37-18	26-18	83-51	68-51
	=13*	=9	=19***	=8**	=32**	=17*
Total	164-101	170-101	132-59	91-59	296-160	261-160
	=63**	=69**	=73***	=32***	=136***	=101***

Each cell reports the mean trading volume in each treatment, along with the difference between W and WOT or between WO and WOT. *, **, and *** denote differences that are statistically significant at the 10%, 5%, and 1% levels, respectively, based on one-tailed Mann–Whitney tests. H1: W–$WOT > 0$ or H1: WO–$WOT > 0$

Hypothesis 6b (Arbitrage-driven explanation for efficiency) *If treatment effects are driven by the presence of arbitrage opportunities, then mispricing in WOT exceeds that in W.*

5.3 Results

Observation 9 *In WOT, both the per-share trading volume and the total trading volume are lower than in W and WO.*

Support: Table 9 compares trading volume in WOT with those in W and WO treatments. Each cell reports the mean trading volume in W and WOT (or in WO and WOT), together with the corresponding difference and its statistical significance. Specifically, Columns 2, 4, and 6 report the mean trading volumes for W and WOT, the difference between them, and the associated significance levels for A-shares, B-shares, and the total across both shares, respectively. With the exception of A-share trading volume in Sequences 2 and 3, trading volumes in W are significantly higher than those in WOT for A-shares, B-shares, and in total—across all sequences. It is also marginally so for A-shares in Sequence 3.

Table 10: Comparison of PD and LPD between the W and WOT treatments, and between the WO and WOT treatments.

	PD		LPD	
	WOT-W	WO-WOT	WOT-W	WO-WOT
Sequence 1	0.127-0.087=0.030	0.173-0.127=0.046	0.127-0.082=0.045	0.159-0.127=0.032
Sequence 2	0.107-0.065=0.042	0.096-0.107=-0.011	0.099-0.064=0.035	0.096-0.099=-0.003
Sequence 3	0.076-0.049=0.027	0.091-0.076=0.015	0.073-0.046=0.027	0.088-0.073=0.014
Overall	0.103-0.067=0.036	0.122-0.103=0.019	0.100-0.064=0.036	0.114-0.100=0.014

Each cell reports the mean PD (or LPD) in each treatment, along with the difference between WOT and W or between WO and WOT. *, **, and *** denote differences that are statistically significant at the 10%, 5%, and 1% levels, respectively, based on one-tailed Mann–Whitney tests. H1: $WOT - W > 0$ or H1: $WO - WOT > 0$

Similarly, as reported in Columns 3, 5, and 7, trading volumes in WO are also significantly higher than those in WOT for A-shares, B-shares, and in total, with the exception of A-shares in Sequences 2 and 3, and B-shares in Sequence 2. For A-shares in Sequence 2, it is marginally significantly so.

Overall, these results confirm that the WOT design successfully reduces trading volume by introducing an additional market in which participants have direct incentives to trade, thereby supporting H4.

Observation 10 *The average PD and LPD in the WOT treatment are lower than in the WO treatment and higher than in the W treatment; however, these differences are not statistically significant.*

Support. Table 10 reports comparisons between WOT and W and between WO and WOT. As shown in Columns 2 and 4 of Table 10, average PD and LPD are higher in WOT than in W across all sequences, but none of these differences is statistically significant. Likewise, as shown in Columns 3 and 5, average PD and LPD are lower in WOT than in WO except in Sequence 2 but these differences are not statistically significant.

Table 12 presents linear regression results for PD and LPD (Models (1) and (2), respectively). For both outcome variables, the estimated coefficients on the WO and

WOT dummies are not statistically significantly different from each other (p-values from the tests of equality are 0.550 for PD and 0.685 for LPD). Likewise, the estimated coefficients on the interaction terms between the treatment dummies and Sequence are not significantly different from one another.

Overall, these results provide no evidence that variation in trading volume due to the existence of the third market affects price discrepancy, nor do they strengthen our main-experiment finding that the presence (versus absence) of arbitrage opportunities across assets affects price discrepancy; thus, they fail to support either H5a or H5b.

Observation 11. *The degree of RD in WOT is not statistically higher than in W, and the degree of RD in WO is not statistically higher than in WOT, across all sequences and for both share types.*

Support. Table 11 reports comparisons of RD between WOT and W, as well as between WO and WOT. Specifically, Columns 2 and 3 present the comparisons for RD_A between W and WOT and between WO and WOT, respectively, while Columns 4 and 5 present the corresponding comparisons for RD_B between W and WOT and between WO and WOT. Overall, the data provide no evidence that RD is higher in WOT than in W, nor that RD is higher in WO than in WOT, for either share type.

Table 12, which reports linear regression results for RD_A and RD_B (Models (3) and (4), respectively), yields findings consistent with the nonparametric comparisons in Table 11. There is no significant difference between the estimated coefficients for WO and WOT dummies (p-values are 0.298 and 0.561 for RD_A and RD_B based on the test of the equality of the estimated coefficients).

Observation 12. *After controlling for participants' risk preferences, there are no statistically significant differences in RAD_A, RAD_B, or DF between WOT and WO,*

Table 11: Comparison of RD, RAD and DF between the WOT and W treatments, and between the WO and WOT treatments.

	RD_A		RD_B		RAD_A		RAD_B		DF	
	WOT-W	WO-WOT	WOT-W	WO-WOT	WOT-W	WO-WOT	WOT-W	WO-WOT	WOT-W	WO-WOT
Sequence 1	0.109-0.002	-0.070-0.109	0.059-0.068	0.007-0.059	0.213-0.060	0.170-0.213	0.077-0.104	0.131-0.077	0.129-0.089	0.151-0.129
	=0.107	=-0.179	=-0.009	=-0.052	=-0.153*	=-0.043	=-0.027	=0.054*	=0.040	=0.022
Sequence 2	-0.028-(-0.045)	-0.036-(-0.028)	0.050-0.024	0.014-0.050	0.085-0.055	0.127-0.085	0.063-0.061	0.058-0.063	0.086-0.059	0.107-0.086
	=0.017	=0.064	=0.026	=-0.036	=0.030	=0.042**	=0.002	=-0.005	=0.027	=0.021
Sequence 3	-0.007-(-0.028)	-0.021-(-0.007)	0.047-0.029	-0.003-0.047	0.059-0.035	0.122-0.059	0.049-0.049	0.091-0.049	0.068-0.043	0.098-0.068
	=0.021	=-0.014	=0.018	=-0.044	=0.024	=0.063**	=0	=0.042	=0.025*	=0.030*
Overall	0.024-(-0.024)	-0.042-0.024	0.052-0.040	0.006-0.052	0.119-0.051	0.140-0.119	0.063-0.072	0.078-0.063	0.094-0.064	0.118-0.094
	=0.048	=-0.066	=0.012	=-0.046	=0.068**	=0.021**	=-0.009	=0.015	=0.030	=0.024

Each cell reports the mean RAD_A, RAD_B or DF in each treatment, along with the difference between WOT and W or between WO and WOT. *, **, and *** denote differences that are statistically significant at the 10%, 5%, and 1% levels, respectively, based on one-tailed Mann–Whitney tests. H1s are WOT–$W > 0$ for comparison between WOT and W and WO–$WOT > 0$ for comparison between WO and W.

Table 12: Regression results in PD, LPD, RD, RAD and DF after adding the WOT treatment.

Dependent variable	PD	LPD	RD_A	RD_B	RAD_A	RAD_B	DF
Models	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	0.175	0.164	-0.067	-0.042	0.248	0.039	0.143
	(0.197)	(0.178)	(0.240)	(0.204)	(0.359)	(0.160)	(0.148)
WO	0.097*	0.084*	-0.095	-0.058	0.109	0.039	0.051
(=1:if WO)	(0.057)	(0.044)	(0.121)	(0.062)	(0.109)	(0.052)	(0.054)
WOT	0.053	0.057	0.130	-0.022	0.210	-0.040	0.046
(=1:if WOT)	(0.052)	(0.055)	(0.164)	(0.054)	(0.187)	(0.052)	(0.078)
The average score	-0.001	-0.001	0.001	0.002	-0.004	0.002	-0.0001
on risk attitude	(0.004)	(0.004)	(0.005)	(0.004)	(0.007)	(0.004)	(0.003)
Sequence	-0.019***	-0.018***	-0.150*	-0.019**	-0.011	-0.028***	-0.023*
	(0.006)	(0.005)	(0.008)	(0.008)	(0.012)	(0.010)	(0.014)
$WO\times$ Sequence	-0.022	-0.018	0.040	0.014	-0.013	-0.015	0.001
	(0.016)	(0.013)	(0.038)	(0.019)	(0.035)	(0.017)	(0.019)
$WOT\times$ Sequence	-0.006	-0.009	-0.043	0.013	-0.066	0.013	-0.007
	(0.014)	(0.015)	(0.071)	(0.014)	(0.056)	(0.014)	(0.023)
p-value (WO=WOT)	0.550	0.685	0.298	0.561	0.627	0.108	0.951
p-value ($WO\times$Sequence= $WOT\times$ Sequence)	0.591	0.729	0.284	0.974	0.461	0.163	0.765
Adjusted R^2	0.134	0.150	-0.023	0.046	0.045	0.120	0.113
Observations				69			

Note: The independent variable, "The average score on risk attitude" quantifies the average risk aversion level of each group, with a lower score indicating a higher degree of risk aversion. Robust standard errors, enclosed in parentheses, are clustered by group. In both panels, *, **, and *** indicate a significant difference from 0 at the 10, 5 and 1% significance levels. The p-values for the coefficient comparison between the WO and WOT dummy variables and between $WO \times$ *Sequence* and $WOT \times$ *Sequence* are computed using a Wald test with robust standard errors.

or between WOT and W.

Support. Columns 6–9 of Table 11 report Mann–Whitney comparisons for RAD_A (Column 6–7), RAD_B (Column 8–9), and DF (Column 10–11) between W and WOT and between WO and WOT. The nonparametric tests indicate several significant differences. For RAD_A, Column 6 shows that RAD_A is significantly higher in WOT than in W across sequences, and marginally for Sequence 1. Column 7 shows that RAD_A is significantly higher in WO than in WOT in Sequences 2 and 3 and across sequences. Column 9 indicates that RAD_B is marginally significantly higher in WO than in WOT in Sequence 1. Finally, Columns 10 and 11 show that DF is marginally significantly higher in WOT than in W, and in WO than in WOT in Sequence 3.

However, these differences become statistically insignificant once we control for participants' risk preferences in regression analyses. Table 12 reports OLS results for RAD_A, RAD_B, and DF (Models (5), (6), (7), respectively). In none of the Models, the estimated coefficients on WO and WOT are significantly different from each other, nor the interactions between the treatment dummies and Sequence (p-values from the test of the equality of the estimated coefficients are (p-values from the test of the equality of the estimated coefficients are 0.627, 0.108, and 0.951 for RAD_A, RAD_B, and DF for the former, and 0.461, 0.163, and 0.765, for the latter, respectively).

Accordingly, Observations 11 and 12 provide no support for either H6a or H6b. Overall, the results provide no evidence that variation in trading volume reduces price bubbles or improves market efficiency, and they also do not strengthen the main-experiment finding that the presence of arbitrage opportunities reduces mispricing.

5.4 Discussion

This section reports the results from an additional treatment, WOT, designed to isolate the role of the Active Participation Hypothesis (Lei et al., 2001) from the main treatment effect in our experiment. The WOT treatment builds on the WO environment by introducing an additional market in which participants can earn direct profits from trading the "good." Under this design, we expect trading volume in the share markets to decline in WOT relative to both W and WO. Moreover, if the main treatment effects are driven (at least in part) by excessive trading induced by speculative motives, then both the cross-market price discrepancy and mispricing should be lower in WOT than in WO. In contrast, if the main treatment effects arise from arbitrage opportunities across the two shares, then cross-market price discrepancies and mispricing in WOT should exceed those in W.

The results are partly consistent with these predictions. Relative to W and WO, trading volume in WOT is significantly lower, consistent with the mechanism emphasized by Lei et al. (2001). Although the average price discrepancy in WOT is lower than in WO and higher than in W—patterns that are directionally consistent with our conjectures—these differences are not statistically significant.

Turning to market efficiency, the evidence is likewise nuanced. We find no statistically significant differences in RD, RAD, or DF between WOT and W, or between WO and WOT.

These results provide no support for the view that the treatment effect on price discrepancy and market outcomes are driven by overtrading induced by speculative motives, and they also do not strengthen the interpretation that the effects documented in the main experiment are solely attributable to arbitrage opportunities across the two shares.

6 Conclusions

In this paper, we conducted an experiment to explore how the absence of arbitrage opportunities across different assets affects asset pricing in a twin market. Our experimental design draws upon Charness and Neugebauer (2019), which in turn is grounded in the framework established by Smith et al. (1988). Prior experimental investigations into multi-asset markets have focused on the law of one price, analyzing price discrepancies by manipulating asset characteristics (Chan et al., 2013; Childs and Mestelman, 2006) and examining the interplay between future dividend processes (Charness and Neugebauer, 2019). Unlike these earlier studies, our research, to our knowledge, is the first experimental inquiry to elucidate the impact of arbitrage opportunities across assets on price discrepancies.

The experimental findings underscore the pivotal role of arbitrage opportunities across assets, not only in achieving the law of one price but also in asset pricing relative to their fundamentals. Specifically, the data indicate that the absence of such arbitrage opportunities tends to exacerbate price discrepancies and the degree of deviation from fundamental values across the market.

Moreover, our data indicate that the treatment effect on price discrepancies is positively associated with the amount of unrealized cross-asset arbitrage profits in outstanding orders, and that substantial cross-asset arbitrage behavior is observed when such arbitrage opportunities exist.

These findings contrast with those of Charness and Neugebauer (2019), who observed that such hedging trades occurred only about 1% of the time when a discrepancy arose in the *PC* treatment, and that 99% of discrepancies were eliminated by independent trades in the market rather than by individuals engaging in arbitrage.

Our experiment demonstrates that when arbitrage between two fundamentally linked assets is blocked, significant price discrepancies and mispricing emerge. This outcome contradicts the predictions of REE and the law of one price in frictionless markets, which would predict price convergence. Recent evidence from financial markets corroborates our finding: even assets with identical fundamentals can trade at divergent prices when arbitrage is impeded. For example, Rakowski et al. (2024) show that cross-listed stocks often exhibit cross-border price divergences due to institutional frictions (such as mutual-fund flow pressures), with mispricing especially pronounced for less liquid stocks and emerging markets. Similarly, in cryptocurrency markets, the "Kimchi premium" on Bitcoin—a gap of over 20% in prices on Korean vs. U.S. exchanges—persisted in 2021 because capital controls limited arbitrageurs' ability to equalize prices.

The above real-world deviations from the law of one price underscore that classical asset pricing models (which assume free arbitrage and fully rational pricing) may fail to describe actual market outcomes when trading is constrained or traders are biased. Our findings align with the limits-to-arbitrage literature, which argues that mispricings can persist if arbitrage is risky, costly, or prohibited. In line with this, Chu et al. (2020) exploit a natural experiment (the SEC's Regulation SHO pilot program) and find that relaxing short-sale constraints significantly dampens pricing anomalies, reducing their average long–short returns by roughly 8.6% per year. This implies that many asset pricing "anomalies" are indeed manifestations of mispricing that survive due to arbitrage frictions—consistent with what we observe in our experiment when arbitrage is curtailed. From an experimental economics perspective, our study reinforces the critical role of arbitrage in promoting price efficiency. Past market experiments have found that introducing arbitrage mechanisms (e.g. algorithmic "arbitrage bots") drives prices of related assets closer together and closer to fundamentals. Angerer et al. (2023), for instance, report that in twin asset markets, the presence of high-speed arbitrage traders enforces the law of one price and reduces mispricing between the assets.

Our paper contributes to the experimental literature by cleanly isolating how arbitrage links (or their absence) affect market efficiency. It also highlights that laboratory markets, like real markets, can fail to reach REE when institutional constraints or imperfect trader rationality are present. In practical terms, market design and regulatory policies should heed these insights. If regulators or exchanges impose rules that impede arbitrage (e.g. bans on short-selling, capital controls, or segmented trading platforms), they may unintentionally foster greater mispricing and volatility. Our experimental evidence confirms that well-intentioned interventions (like limiting cross-market trades to "stabilize markets") can backfire by impairing price discovery. Conversely, a policy that facilitates arbitrage across linked assets (for example, allowing short sales or interoperable exchanges) is likely to enhance pricing efficiency and keep values aligned, albeit potentially transferring profits to specialized arbitrageurs. In sum, our findings suggest that asset pricing theory and regulatory practice must account for real-world frictions: prices can deviate significantly from fundamentals when arbitrage is limited, so models and policies assuming frictionless markets may misjudge actual outcomes.

To align our study with the broader experimental literature on multi-asset markets, we adopted the framework of Smith et al. (1988). However, this setting does not inherently provide strong incentives to transact. As emphasized by Lei et al. (2001) and Crockett et al. (2019), weak transaction incentives may affect trading intensity and, in turn, pricing outcomes. In line with this concern, it has been pointed out that the result of our original experiment might be due to the Active Participation Hypothesis (Lei et al., 2001), i.e., those participants who can trade only one share type engaging in speculative trading to a greater extent than those who can trade both share types. Under this interpretation, the treatment effects documented above could reflect differences in excess trading volume rather than the presence (or absence) of cross-asset arbitrage opportunities. To assess this possibility, we implemented an additional treatment in which participants could trade only one share type but also had access to an additional market that provided direct incentives to trade. Comparisons between this additional treatment and the main treatments do not support this interpretation, although they also do not decisively strengthen our main-experiment interpretation.

An alternative experimental framework that explicitly incentivizes trade is proposed by Asparouhova et al. (2016) and Crockett et al. (2019). This framework has been utilized to test the robustness of Lucas model features in a three-period cyclical world (Carbone et al., 2021) and to explore the impact of quantitative monetary easing policies on financial asset pricing (Duan and Hanaki, 2023). To the best of our knowledge, however, the framework of Asparouhova et al. (2016) and Crockett et al. (2019) has not yet been applied to investigate the law of one price or the underlying causes of price discrepancies, which presents a promising avenue for future research.

References

Angerer, M., T. Neugebauer, and J. Shachat (2023): "Arbitrage bots in experimental asset markets," *Journal of Economic Behavior and Organization*, 206, 262–278.

Asparouhova, E., P. Bossaerts, N. Roy, and W. Zame (2016): ""Lucas" in the laboratory," *Journal of Finance*, 71, 2727–2779.

Ben-David, I., F. Franzoni, and R. Moussawi (2018): "Do ETFs Increase Volatility?" *The Journal of Finance*, 73, 2471–2535.

Carbone, E., J. Hey, and T. Neugebauer (2021): "An Experimental Comparison of Two Exchange Economies: Long-Lived Asset vs. Short-Lived Asset," *Management Science*, 67, 6496–6962.

Chan, K. S., V. Lei, and F. Vesely (2013): "Differentiated assets: An experimental study on bubbles," *Economic Inquiry*, 51, 1731–1749.

Charness, G. and T. Neugebauer (2019): "A Test of the Modigliani-Miller Invariance Theorem and Arbitrage in Experimental Asset Markets," *The Journal of FINANCE*, 74, 493–529.

Childs, J. and S. Mestelman (2006): "Rate-of-return Parity in Experimental Asset Markets," *Review of International Economics*, 14, 331–347.

Chu, Y., D. Hirshleifer, and L. Ma (2020): "The Causal Effect of Limits to Arbitrage on Asset Pricing Anomalies," *Journal of Finance*, 75, 2631–2672.

Crockett, S., J. Duffy, and Y. Izhakian (2019): "An experimental test of the Lucas asset pricing model," *Review of Economic Studies*, 86, 627–667.

Duan, J. and N. Hanaki (2023): "The impact of asset purchases in an experimental market with consumption smoothing motives," *Journal of Economic Dynamics and Control*, 156.

Duffy, J., J. Rabanal, and O. Rud (2022): "Market Experiments with Multiple Assets: A Survey," in *Handbook of Experimental Finance*, ed. by S. Füllbrunn and E. Haruvy, Edward Elgar, chap. 18, 213–224.

Dufwenberg, M., T. Lindqvist, and E. Moore (2005): "Bubbles and Experience: An Experiment," *American Economic Review*, 95, 1731–1737.

Fischbacher, U. (2007): "z-Tree: Zurich toolbox for ready-made economic experiments," *Experimental Economics*, 10, 171–178.

Greiner, B. (2015): "An online recruitment system for economic experiments," *Journal of the Economic Science Association*, 1, 114–125.

Healy, P. J. (2020): "Explaining the bdm—or any random binary choice elicitation mechanism—to subjects," Tech. rep., Department of Economics, The Ohio State University, Columbus, OH.

Hirshleifer, J. (1966): "Investment Decision under Uncertainty: Applications of the State-Preference Approach," *The Quarterly Journal of Economics*, 80, 252–277.

Lei, V., C. N. Noussair, and C. R. Plott (2001): "Nonspeculative Bubbles in Experimental Asset Markets: Lack of Common Knowledge of Rationality vs. Actual Irrationality," *Econometrica*, 69, 831–859.

Levati, M. V., J. Qiu, and P. Mahagaonkar (2012): "Testing the Modigliani-Miller theorem directly in the lab," *Experimental Economics*, 15, 693–716.

Noussair, C. and S. Tucker (2006): "Futures markets and bubble formation in experimental asset markets," *Pacific Economic Review*, 11, 167–184.

Nuzzo, S. and A. Morone (2017): "Asset markets in the lab: A literature review," *Journal of Behavioral and Experimental Finance*, 13, 42–50.

Owen, A. L. and R. H. Thaler (2003): "Anomalies: The law of one price in financial markets," *Journal of Economic Perspectives*, 191–202.

Palan, S. (2013): "A Review of bubbles and crashes in experimental asset markets," *Journal of Economic Surveys*, 27, 570–588.

Powell, O. and N. Shestakova (2016): "Experimental asset markets: A survey of recent developments," *Journal of Behavioral and Experimental Finance*, 12, 14–22.

Rakowski, D., J. D. Diltz, and A. T. Nguyen (2024): "Mutual fund flow-driven trading and the mispricing of cross-listed stocks," *Journal of Multinational Financial Management*, 2631–2672.

Smith, V. L., G. L. Suchanek, and A. W. Williams (1988): "Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets," *Econometrica*, 56, 1119–1151.

Stiglitz, J. E. (1969): "A Re-Examination of the Modigliani-Miller Theorem," *The American Economic Review*, 59, 784–793.

Stöckl, T., J. Huber, and M. Kirchler (2010): "Bubble measures in Experimental Asset Markets," *Experimental Economics*, 13, 284–298.

A The analysis following the pre-registration

As noted in the beginning of Section 3, because some of the measures we have originally pre-registered proved inadequate for capturing specific characteristics of our data, or were found unsuitable for this experimental design, the hypotheses and analyses related to those measures have been excluded from the main body of the paper. However, to maintain transparency, this Appendix presents the complete results based on the measures and hypotheses initially pre-registered.

In the pre-registration, we have defined six measures: $\Delta_t^{r,g}$, $PD^{r,g}$, $\beta_t^{r,g}$, $|\beta^{r,g}|$, $RD_k^{g,r}$, and $DF^{r,g}$.[17] The first four assess the price deviations from the parity, while the remaining two evaluate the price deviations from fundamentals. The definitions of $\Delta_t^{r,g}$, $PD^{r,g}$, $RD_k^{g,r}$, and $DF^{r,g}$ are provided in Eq.(1), Eq.(2), Eq(4), and Eq.(6). The measures $\beta_t^{r,g}$, and $|\beta^{r,g}|$ are defined as follows:

$$\beta_t^{r,g} = P_{B,t}^{r,g} - \frac{E(D_{B,t})}{E(D_{A,t})} P_{A,t}^{r,g} \tag{A.1}$$

$$|\beta|^{r,g} = \frac{1}{T} \sum_{t=1}^{T} |\beta_t^{r,g}| \tag{A.2}$$

It should be noted that $\beta_t^{r,g}$ is associated with the equation $\frac{P_{B,t}^{r,g}}{P_{A,t}^{r,g}} = \frac{E(D_{B,t})}{E(D_{A,t})}$, reflecting rate-of-return parity, where $E(D_{A,t})$ and $E(D_{B,t})$ denote the expected dividends per unit for A-shares and B-shares, respectively. This formulation articulates the price relationship between A-shares and B-shares as $P_{B,t}^{r,g} = \frac{E(D_{B,t})}{E(D_{A,t})} P_{A,t}^{r,g}$. Therefore, $\beta_t^{r,g}$ as defined by Equation A.1 evaluates the degree of deviation from this parity, alongside $|\beta^{r,g}|$ as outlined in Equation A.2. However, our experimental design, guided by Charness and Neugebauer (2019), informed participants that "the dividend amount

[17] As noted in footnote 5, the pre-registration contained a typo in defining RD. Namely, $\overline{FV}_k$ in the denominator of Equation 4 was shown as $FV_{k,t}$.

for each unit of B-shares is always 24 cents higher than that for each unit of A-shares" rather than "the expected dividend amount for each unit of B-shares is always equals the 2 times of that for each unit of A-shares." This distinction implies that Δ_t and PD are more suitable than $\beta_t^{r,g}$ and $|\beta^{r,g}|$ for analyzing price discrepancies in our paper. Consequently, we have excluded the analysis concerning $\beta_t^{r,g}$ and $|\beta^{r,g}|$ from the main body of the paper.

Using the above defined measures, the following four hypotheses were given in the pre-registration:

Hypothesis A0 (Risk-neutral pricing) *There are no deviations of prices from fundamentals for either type of shares, which means that RD_A, RD_B and DF equal zero.*

Hypothesis A1 (Rate-of-return parity):
Hypothesis A1-1 *The ratio of market prices between twin-shares will equal the ratio of their expected dividend, which implies that β_t and $|\beta|$ equal zero.*
Hypothesis A1-2 *The adjusted market prices for twin-shares will be identical, which implies that Δ_t and PD are both equal to zero.*

Hypothesis A2 (Hedging effect) *The cross-asset price discrepancies vary depending on whether arbitrage between twin-shares is permitted, indicating that β_t, $|\beta|$, Δ_t and PD are different between W and WO treatments.*

Testing **Hypothesis A2**, which relates to comparing price discrepancies, requires a series of two-sided two-sample tests. However, the primary objective of this paper is to test whether limiting the opportunity for risk-free arbitrage across different assets exacerbates the degree of violation of the law of one price, as suggested by Charness and Neugebauer (2019). This objective is outlined in the "Introduction" section of the main body of the paper. Consequently, we have reformulated this hypothesis (Hypothesis 1 in the main text) to state that "(the measures assessing the degree of price discrepancies, namely LPD) will be higher in the WO treatment than in the W treatment." which are more suitable for the purpose of this experiment and requires a one-sided two-sample test. We consider the new hypothesis to be better aligned with the goals of this experiment.

Hypothesis A3 (Experience effect) *The RD_k, DF, β_t, $|\beta_t|$, Δ_t, and PD will decrease over time (in consecutive market).*

During the analysis, we observed significant deviation in the pricing within a specific group from the W treatment, where, for example, the average PD and LPD for this group across three sequences was 0.295 and 0.241. These values are 3 and 2.8 times their respective averages (average PD is 0.096 and the average LPD is 0.086) for the W treatment, and 2.5, and 2.2 times higher than the respective second highest observations (second highest PD is 0.117, and the second higher LPD is 0.108) within the treatment (The Smirnoff-Grubbs test yielded $t = 2.28, p = 0.005$ for PD and $t = 2.21, p = 0.011$ for LPD). Consequently, we excluded this outlier from the analysis presented in the main body of the paper to avoid biasing the results. However, our pre-registration did not specify exclusion criteria based on the Smirnov-Grubbs test. As such, we will report results both with and without this outlier included.

Table A.1 presents the results of the Mann-Whitney tests conducted both between treatments and between median values and zero for each measure. In this table, columns two through four display the results with the outlier omitted, while columns five through seven show the results with the outlier included. Furthermore, Table A.2 provides the corresponding OLS results, controlling for group average risk preferences.

Table A.1: The results of the two-sample Mann-Whitney tests conducted between treatments for each measure

Data	After omitting the outlier			Before omitting the outlier		
Treatments	W (n=7)	WO (n=8)	W = WO[a]	W (n=8)	WO (n=8)	W = WO[a]
Panal A. The median value of average Δ						
Sequence 1	0.025**	0.028	p = 0.536	0.052***	0.028	p = 0.328
Sequence 2	0.032	0.024	p = 0.951	0.050*	0.024	p = 0.505
Sequence 3	0.029**	-0.006	p = 0.463	0.039**	-0.006	p = 0.165
Total	0.057**	0.009	p = 0.694	0.060**	0.009	p = 0.442
Panal B. The meidan value of PD						
Sequence 1	0.080**	0.125***	p = 0.094	0.095***	0.125***	p = 0.279
Sequence 2	0.063**	0.104***	p = 0.232	0.073***	0.104***	p = 0.505
Sequence 3	0.036**	0.084***	p = 0.121	0.055***	0.084***	p = 0.161
Total	0.065**	0.107***	p = 0.152	0.074***	0.107***	p = 0.382
Panal C. The median value of average β_t						
Sequence 1	15.12**	0.84	p = 0.281	18.89***	0.84	p = 0.195
Sequence 2	13.29**	2.05	p = 0.397	18.51**	2.05	p = 0.328
Sequence 3	13.54**	-8.10	p = 0.336	14.06***	-8.10	p = 0.328
Total	14.52**	-6.84	p = 0.281	20.07***	-6.84	p = 0.235
Panal D. The median value of $\|\beta\|$						
Sequence 1	22.81**	31.97***	p = 0.054	24.03***	31.97***	p = 0.130
Sequence 2	16.38**	28.22***	p = 0.121	20.96***	28.22***	p = 0.235
Sequence 3	16.73**	30.87***	p = 0.121	18.85***	30.87***	p = 0.105
Total	16.28**	29.04***	p = 0.072	20.94***	29.04***	p = 0.161
Panal E. The median value of RD_A						
Sequence 1	0.004	-0.003	p = 0.694	0.001	-0.003	p = 0.798
Sequence 2	-0.011	-0.020	p = 0.955	-0.002	-0.020	p = 0.879
Sequence 3	-0.009	-0.021	p = 0.955	-0.012	-0.021	p = 0.959
Total	-0.005	-0.001	p = 0.955	-0.001	-0.001	p = 1.000
Panal F. The median value of RD_B						
Sequence 1	0.024*	0.020	p = 0.336	0.063**	0.020	p = 0.195
Sequence 2	0.019	0.004	p = 1.000	0.038	0.004	p = 0.721
Sequence 3	0.009	-0.003	p = 0.281	0.024	-0.003	p = 0.195
Total	0.016	-0.004	p = 0.281	0.042	-0.004	p = 0.160
Panal G. The median value of DF						
Sequence 1	0.064**	0.124***	p = 0.054	0.067***	0.124***	p = 0.195
Sequence 2	0.059**	0.093***	p = 0.189	0.060***	0.093***	p = 0.441
Sequence 3	0.031**	0.102***	p = 0.014	0.037***	0.102***	p = 0.014
Total	0.064**	0.118***	p = 0.054	0.060***	0.115***	p = 0.195

*, **, and *** indicate a significant difference from 0 at the 10, 5, and 1% significance levels, respectively.

a: p-values are based on the two-tailed, one-sample Mann-Whitney tests. H_0: = 0.

Table A.2: The regressions on each measure

Panel A. The regressions results after omitting the outlier

	Dependent variables						
	$\Delta_t^{g,r}$	$PD^{r,g}$	$\beta_t^{r,g}$	$\lvert\beta\rvert^{r,g}$	$RD_A^{g,r}$	$RD_B^{g,r}$	$DF^{g,r}$
Intercept	0.078	0.015	-12.15	3.75	0.005	-0.067	0.001
	(0.542)	(0.366)	(173.32)	(127.01)	(0.648)	(0.501)	(0.333)
Treatment	0.003	0.059**	-2.16	23.36*	-0.020	-0.029	0.059**
(variable=1: WO)	(0.046)	(0.029)	(18.39)	(13.05)	(0.070)	(0.031)	(0.025)
The average score	-0.0001	0.002	0.797	0.647	-0.001	0.002	0.002
on risk attitude	(0.011)	(0.008)	(3.66)	(2.646)	(0.014)	(0.011)	(0.007)
Sequence	-0.013	-0.031***	-4.86	-7.51*	0.006	-0.012	-0.023**
	(-0.013)	(0.009)	(4.42)	(4.10)	(0.021)	(0.010)	(0.009)
Adjusted R^2	-0.057	0.200	-0.053	0.156	-0.066	0.022	0.178
Observations				45			

Panel B. The regressions results before omitting the outlier

	Dependent variables						
	$\Delta_t^{g,r}$	$PD^{r,g}$	$\beta_t^{r,g}$	$\lvert\beta\rvert^{r,g}$	$RD_A^{g,r}$	$RD_B^{g,r}$	$DF^{g,r}$
Intercept	-0.369	-0.431	-88.46	-79.53	-0.0.109	-0.540	-0.282
	(0.570)	(0.484)	(161.80)	(132.06)	(0.544)	(0.495)	(0.342)
Treatment	-0.003	0.053	-3.23	22.20	-0.022	-0.036	0.055*
(variable=1: WO)	(0.050)	(0.033)	(18.66)	(13.40)	(0.070)	(0.038)	(0.028)
The average score	0.010	0.012	2.53	2.54	0.002	0.013	0.009
on risk attitude	(0.012)	(0.011)	(3.45)	(2.81)	(0.116)	(0.011)	(0.007)
Sequence	-0.024	-0.042***	-7.01	-9.67**	0.005	-0.021	-0.029**
	(0.015)	(0.014)	(4.69)	(4.44)	(0.020)	(0.013)	(0.011)
Adjusted R^2	0.019	0.159	0.000	0.133	-0.056	0.017	0.168
Observations				48			

Note: The independent variable, "The average score on risk attitude" quantifies the average risk aversion level of each group, with a lower score indicating a higher degree of risk aversion. Robust standard errors, enclosed in parentheses, are clustered by group. In all panels, *, **, and *** indicate a significant difference from 0 at the 10, 5, and 1% significance levels, respectively.

In this table, Panel A reports the OLS results with the outlier omitted, and Panel B presents the results with the outlier included.

Let us first see the results related to the rate-of-return parity, which are displayed in Panels A through D of Table A.1, as well as columns two through five of Table A.2. Initially, we observe that the average Δ_t (Second and sixth columns in Panel A, Table A.1) and average β_t (Second and sixth columns Panel C, Table A.1) in W treatment are consistently and significantly greater than 0 in most of sequences, regardless of whether the outlier is included or omitted. This indicates that B-shares are systematically overvalued relative to A-shares in the W treatment. Furthermore, Table A.1 reports that PD (Panel B) and $|\beta|$ (Panel D) in both treatments are significantly greater than 0 in all sequences, regardless of outlier inclusion. This finding demonstrates a strong violation of rate-of-return parity in this experiment, thereby refuting **Hypothesis A1**.

For the results related to the comparison of Δ_t, PD, β_t, and $|\beta|$ between treatments, the findings vary depending on whether the analysis includes or omits the outlier. On one hand, when the outlier is omitted, we observe some treatment effects: the median value of PD in the WO treatment is significantly higher than that in the W treatment in Sequence 1 (second row and fourth column, Panel D, Table A.1) at 10% significant level. Similarly, the median value of $|\beta|$ in the WO treatment is significantly higher than in the W treatment in Sequence 1 (Second row and fourth column, Panel D, Table A.1) and overall (Fifth row and fourth column, Panel D, Table A.1), at 10% significant level. This effect is more clearly observed when controlling for group average risk preference in regression: the treatment dummy (= 1: WO) presents a significant positive coefficient for the dependent variable PD (second row and second column, Panel A Table A.2) at 5% significant level, and for the dependent variable $|\beta|$ (second row and fourth column, Panel A Table A.2) at 10% significant level. These observations suggest that limiting opportunities for arbitrage between twin shares increases the degree of price discrepancy violation, providing some support for **Hypothesis A2**.

However, on the other hand, if the outlier is included in the analysis, no treatment effect is observed, as shown in the last column of Panels B and D in Table A.1, and in the second row, second and fourth columns in Panel B of Table A.2.

Let us now focus on the results concerning asset mispricing from fundamentals. In Panels E and F of Table A.1, we find that the median values of RD_A and RD_B

are not significantly differ from 0 in any of the treatments and sequences, except for Sequence 1 of the W treatment where the median RD_B is higher than 0 at 10% significant level when omitting the outlier and at 5% significant level when including it. Furthermore, Panel G reveals that the median DF in both treatments is significantly higher than 0 in all sequences, irrespective of the inclusion of the outlier. These findings partially reject **Hypothesis A0**.

Focusing on the comparison between treatments, we observe treatment effects related to the degree of mispricing from the fundamentals in some sequences. Specifically, Panel G of Table A.1 reports that, when the outlier is omitted, the median value of DF in the WO treatment is significantly higher than in the W treatment in Sequence 1 (second row and fourth column), Sequence 3 (fourth row and fourth column) and overall (fifth row and fourth column) at 10%, 5%, and 10% significance levels, respectively. However, when the outlier is included, this difference is only significant in Sequence 3 at 5% significance level. The OLS regression yields consistent results: in Table A.2, for the dependent variable DF, the treatment dummy (= 1: WO) shows a significant positive coefficient at 5% significance level when the outlier is omitted (second row and last column, Panel A), and at 10% significance level when the outlier is included (second row and last column, Panel B). This finding suggests that limiting arbitrage opportunities between twin shares may increase the degree of asset mispricing from fundamentals.

Finally, Table A.2 reports significant negative coefficients for the variable "Sequence" for PD, $|\beta|$, and DF at significant levels of 1%, 10%, and 5% respectively, when the outlier is excluded (Panel A). When the outlier is included, the significance levels are 1%, 5%, and 5% respectively (Panel B), respectively. These findings suggest that PD, $|\beta|$, and DF decrease over time (in consecutive markets), which supports **Hypothesis A3**.

Throughout Tables A.1 and A.2, we observe that measures derived from absolute values, such as PD, $|\beta|$, and DF, tend to show significant treatment differences more readily than measures that are not, such as average Δ_t and RD. This suggests, once again, that the differences between treatments are more pronounced in the amplitude of fluctuations in the deviation from parity and the pricing deviation from fundamentals, rather than in their mean values. This is the reason why we changed the measures in the main body of the paper.

B The analysis of the measures of LPD, RAD_k, and DF when no data is omitted

This Appendix reports the results of LPD, RAD_k, and DF when the outlier is not omitted.

Table B.1 presents the results for LPD, assessing the degree of price discrepancy between twin-shares. Specifically, Panel A displays the median LPD for each treatment and the results of the one-sided Mann-Whitney test between treatments. Panel A indicates that the median LPD values in the WO treatment are significantly higher than those in the W treatment in Sequence 3 (0.053 in the W treatment vs. 0.088 in the WO treatment, $p = 0.080$) and Overall (0.071 in the W treatment vs. 0.105 in the WO treatment, $p = 0.097$) at 10% significant level.[18] Panel B, which presents the regression results on the dependent variable $LPDs$, reports a significant positive coefficient ($= 0.050$) for the dummy variable "treatment ($= 1 : WO$)" at 10% significant level in Model (1), while it is not significant in Model (2). These findings

[18] If we employ a two-sided Mann-Whitney test, the $p-value$ for the differences in LPD between treatment are $p = 0.235$ for Sequence 1, $p = 0.505$ for Sequence 2, $p = 0.161$ for Sequence 3, and $p = 0.195$ for Overall. No statistical differences are observed, and **Hypothesis 1** is not rejected.

Table B.1: The Log Absolute Deviation from Parity, LPD

Panel A. the median value of LPD

	W ($n = 8$)	WO ($n = 8$)	The p-values from one-sided Mann-Whitney test between treatments (H_0: W = WO)
Sequence 1	0.088	0.122	0.117
Sequence 2	0.077	0.104	0.253
Sequence 3	0.053	0.088	0.080*
Overall	0.071	0.105	0.097*

Panel B. The regressions on $LPD^{g,r}$

Independent variables	(1)	(2)
Intercept	-0.349	-0.0351
	(0.375)	(0.368)
Treatment	0.050*	0.053
(variable=1: WO)	(0.027)	(0.059)
The average score	0.010	0.010
on risk attitude	(0.008)	(0.008)
Sequence	-0.035***	-0.033*
	(0.010)	(0.017)
Treatment×		-0.002
Sequence		(0.021)
Adjusted R^2	0.195	0.176
Observations	48	

Note: Panel A presents the median value of LPD for each sequence and treatment. Panel B details the linear regressions conducted on $LPD^{g,r}$. In Panel B, the independent variable, "The average score on risk attitude" quantifies the average risk aversion level of each group, with a lower score indicating a higher degree of risk aversion. Robust standard errors, enclosed in parentheses, are clustered by group. In both panels, *, **, and *** indicate a significant difference from 0 at the 10, 5 and 1% significance levels.

provide some evidence rejecting **Hypothesis 1**, though this evidence is weaker than when the outlier is omitted, where the median $LPDs$ was higher in the WO treatment than in the W treatment in Sequences 1, 3, and Overall at 5%, 10%, and 5% significance levels respectively. Additionally, the variable "Sequence" shows a significant negative coefficient in both (1) ($= -0.035$) and Model (2) ($= -0.033$) at 1% and 10% significant levels, respectively. These results support **Hypothesis 3**.

The results of RAD_k, and DF, which assess the degree of mispricing, are reported in Table B.2. Specifically, Panel A, B, and C present the median values of RAD_A, RAD_B, and DF for each treatment, along with the comparison results between treatments using the one-sided Mann-Whitney test for each measure, respectively. Panel D shows the regression results for each measure.

In Panel A, we observe that the values of RAD_A in the WO treatment are significantly higher than those in the W treatment at 5%, 10%, 1%, and 5% significant levels at Sequence 1 (0.041 in the W treatment vs. 0.116 in the WO treatment, $p = 0.041$), Sequence 2 (0.028 in the W treatment vs. 0.084 in the WO treatment, $p = 0.065$), Sequence 3 (0.019 in the W treatment vs. 0.130 in the WO treatment, $p = 0.003$), and Overall (0.045 in the W treatment vs. 0.112 in the WO treatment, $p = 0.025$), respectively.[19] However, as shown in Panel B, no statistical difference in the values of RAD_B is observed between treatments.[20] Compared to the results when the outlier is omitted, there is no change in the significance levels for these two measures, although the median value of RAD_A for Sequence 1 in the W treatment

[19] If we employ a two-sided Mann-Whitney test, the $p-value$ for the differences in RAD_A between treatment are $p = 0.083$ for Sequence 1, $p = 0.130$ for Sequence 2, $p = 0.007$ for Sequence 3, and $p = 0.050$ for Overall. The effect of restricting arbitrage between twin-shares on the pricing of A-share is observed, rejecting **Hypothesis 2**.

[20] If we employ a two-sided Mann-Whitney test, the $p-value$ for the differences in RAD_B between treatment are $p = 0.876$ for Sequence 1, $p = 0.442$ for Sequence 2, $p = 0.798$ for Sequence 3, and for Overall. No statistical differences are observed between treatments, consistent with the results obtained using a one-sided test.

Table B.2: The price deviation from the fundamental

Panel A. the median value of RAD_A

	W ($n = 7$)	WO ($n = 8$)	The p−*values* from the one-sided Mann-Whitney test between treatments (H_0: W = WO)
Sequence 1	0.041	0.116	0.041**
Sequence 2	0.028	0.084	0.065**
Sequence 3	0.019	0.130	0.003***
Overall	0.045	0.112	0.025**

Panel B. the median value of RAD_B

	W ($n = 8$)	WO ($n = 8$)	The *p*-values from the one-sided Mann-Whitney test between treatments (H_0: W = WO)
Sequence 1	0.110	0.110	0.439
Sequence 2	0.071	0.044	0.809
Sequence 3	0.062	0.034	0.640
Overall	0.078	0.063	0.640

Panel C. the median value of DF

	W ($n = 8$)	WO ($n = 8$)	The *p*-values from the one-sided Mann-Whitney test between treatments (H_0: W = WO)
Sequence 1	0.067	0.124	0.097*
Sequence 2	0.060	0.093	0.221
Sequence 3	0.038	0.102	0.007***
Overall	0.061	0.115	0.097*

Panel D. The regressions on $RAD_A^{g,r}$, $RAD_B^{g,r}$, and $DF^{g,r}$.

	Dependent variables		
	$RAD_A^{g,r}$	$RAD_B^{g,r}$	$DF^{g,r}$
Intercept	-0.108	-0.426	-0.282
	(0.412)	(0.450)	(0.342)
Treatment	0.092*	0.003	0.055*
(variable=1: WO)	(0.055)	(0.028)	(0.028)
The average score	0.004	0.013	0.008
on risk attitude	(0.008)	(0.010)	(0.007)
Sequence	-0.018	-0.046***	-0.029**
	(0.018)	(0.014)	(0.011)
Adjusted R^2	0.130	0.191	0.168
Observations		48	

Note: Panels A, B, and C present the median value of RAD_A, RAD_B, and DF for each sequence and treatment, respectively. Panel D provides the results or linear regression analyses on RAD_A, RAD_B, and DF. Within Panel D, the independent variable, "The average score on risk attitude" quantifies the average risk aversion level of each group, with a lower score indicating a higher degree of risk aversion. Robust standard errors, enclosed in parentheses, are clustered by group. In all panels, *, **, and *** indicate a significant difference from 0 at the 10, 5, and 1% significance levels, respectively.

has changed significantly.

From Penal C, we see that the values of DF are significantly higher in the WO treatment than in the W treatment in Sequence 1 (0.067 in the W treatment vs. 0.124 in the WO treatment, $p = 0.097$), Sequence 3 (0.038 in the W treatment vs. 0.102 in the WO treatment, $p = 0.007$), and Overall (0.061 in the W treatment vs. 0.115 in the WO treatment, $p = 0.097$) at 10%, 1%, and 10% significant level, respectively. Compared with when the outlier is omitted, the difference in Sequence 2 becomes insignificant, and the significance levels of the differences in Sequence 1 and Overall are reduced.[21]

Moreover, Panel D presents significant positive coefficients of the dummy variable "treatment $(= 1 : WO)$" for the dependent variables RAD_A $(= 0.092)$ and DF $(= 0.055)$ at 10% significant levels, consistent with the results reported in Panels A through C, partially rejecting **Hypothesis 2** of the main body of the paper.

Finally, In Panel D, the variable "Sequence" exhibits negative coefficients for the dependent variables RAD_B $(= -0.046)$ and DF $(= -0.029)$ at 1% and 5% significant levels, respectively. These findings suggest that the degree of mispricing from the fundamentals for B-shares and the whole market decreases over sequences, which partially supports **Hypothesis 3** and aligns with the analysis when the outlier is omitted.

[21] If we employ a two-sided Mann-Whitney test, the $p - value$ for the differences in DF between treatment are $p = 0.195$ for Sequence 1, $p = 0.442$ for Sequence 2, $p = 0.015$ for Sequence 3, and $p = 0.195$ for Overall. The treatment effect is observed only in Sequence 3, making it difficult to reject **Hypothesis 2** under this situation.

C Analysis of the Effects of Trading Volume and Turnover Rate on Mispricing and Price Discrepancy

Given that trading frequency may potentially influence asset pricing—for example, (i) a more expensive asset may be traded less frequently, revealing less market information and thereby increasing mispricing, or (ii) a more expensive asset might be less susceptible to overpricing due to its higher cost—this section examines whether trading volume and turnover rate differ between treatments and between share types. Subsequently, we investigate whether and how trading volume or turnover rate influences the measures of mispricing (RD, RAD, and DF) and price discrepancy (PD, and LPD).

We first analyze whether total trading volumes or turnover rates differ between the two types of shares. Both measures are considered because their initial endowments differ (the endowment of A-shares is twice that of B-shares), as this difference could lead to distinct analytical outcomes depending on the metric employed. Results from Wilcoxon signed-rank tests are presented in Panel A of Table C.3.

Observation C.1 *The turnover rate of B-shares is higher than that of A-shares in the W treatment, whereas the total trading volume of A-shares is higher than that of B-shares in the WO treatment.*

Support: The left section of Panel A in Table C.3 compares total trading volumes between A-shares and B-shares for each sequence and treatment, while the right section compares turnover rates. As shown in the upper-left section, no statistically significant differences in total trading volume between A-shares and B-shares were

Table C.3: The comparison of the turnover rate and total trading volumes between treatments.

Panel A. The comparison of the trading volume and turnover rate in each treatement between share types.

		Trading volume			Turnover rate		
		A-share	B-share	p-values (H_0: $Vol_A = Vol_B$)	A-share	B-share	p-values (H_0: $TO_A = TO_B$)
W	Seq 1	79	62	0.125	0.188	0.295	0.078
W	Seq 2	38.42	33.28	0.125	0.091	0.158	0.031
W	Seq 3	46.29	36.71	0.406	0.110	0.174	0.078
WO	Seq 1	85.25	43.75	0.031	0.203	0.208	0.945
WO	Seq 2	42.88	21.25	0.016	0.102	0.101	1.000
WO	Seq 3	41.88	26.38	0.219	0.100	0.125	0.313

Panel B. The comparison of the trading volume and turnover rate of each type of shares between treatments.

		Trading volume			Turnover rate		
		W ($n = 7$)	WO ($n = 8$)	p-values (H_0: W = WO)	W ($n = 7$)	WO ($n = 8$)	p-values (H_0: W = WO)
A-shares	Seq 1	79	85.25	0.633	0.188	0.203	0.633
A-shares	Seq 2	38.42	42.88	0.757	0.091	0.102	0.757
A-shares	Seq 3	46.29	41.88	0.631	0.110	0.100	0.631
B-shares	Seq 1	62	43.75	0.231	0.295	0.208	0.231
B-shares	Seq 2	33.28	21.25	0.030	0.158	0.101	0.030
B-shares	Seq 3	36.71	26.38	0.067	0.174	0.125	0.067

Panel C. The comparison of the total trading volumes between treatments.

	W ($n = 7$)	WO ($n = 8$)	p-values (H_0: W = WO)
Sequence 1	141	129	0.802
Sequence 2	71.71	64.13	0.414
Sequence 3	83	68.25	0.220

Note: The p-values in Panel A are derived from two-sided Wilcoxon signed-rank tests, while those in Panels B and C are obtained from two-sided Mann–Whitney U tests.

found in the W treatment. However, as reported in the upper-right section, turnover rates of B-shares were significantly higher than those of A-shares at the 10%, 5%, and 10% significance levels in Sequences 1, 2, and 3, respectively. In contrast, the WO treatment exhibited the opposite pattern. Specifically, as shown in the lower-left section, total trading volumes for A-shares were significantly higher than those for B-shares in Sequences 1 and 2 at the 5% significance level, whereas the lower-right section showed no statistically significant differences in turnover rates between A-shares and B-shares in any sequence.

Observation C.2 *For A-shares, no statistically significant differences in trading volume or turnover rate exist between treatments. However, for B-shares, both trading volume and turnover rate are significantly higher in the* W *treatment than in the* WO *treatment during Sequences 2 and 3.*

Support: Panel B of Table C.3 reports comparisons of total trading volume and turnover rate between treatments for each sequence and share type. The p$-$*values* obtained from two-sided Mann–Whitney U tests. The upper half of Panel B indicates no statistically significant differences in trading volumes or turnover rates of A-shares between treatments. Conversely, the lower half demonstrates that both trading volumes and turnover rates of B-shares are significantly higher in the W treatment compared to the WO treatment in Sequences 2 and 3, at the 5% and 10% significance levels, respectively.

Observation C.3 *No statistically significant differences exist in total trading volume between treatments across all sequences.*

Support: Panel C of Table C.3 presents comparative results for total trading volumes (the sum of trading volumes for A-shares and B-shares) across treatments. The

p-values obtained from two-sided Mann–Whitney U tests. Although total trading volumes are slightly higher in the W treatment compared to the WO treatment in all sequences, these differences are not statistically significant at conventional levels.

To further explore the potential impact of trading volume and turnover rate on asset pricing, linear regression analyses were conducted, regressing RD, RAD, DF, PD, and LPD on trading volumes and turnover rates. The results of these analyses are shown in Table C.4.

Observation C.4 *Neither trading volume nor turnover rate has a statistically significant effect on* RD*,* RAD*,* DF*,* PD*, or* LPD*.*

Support: In Table C.4, Models (1) and (2) use RD as the dependent variable, Models (3) and (4) use RAD, Model (5) uses DF, Model (6) uses PD, and Model (7) uses LPD. The table indicates that coefficients for the independent variables "Trading volume of each share type" (Models 1 and 3), "Turnover rate of each share type" (Models 2 and 4), and "Total trading volume of both share types" (Models 5, 6, and 7) are not statistically significantly different from zero. These results imply no evidence that trading volume or turnover rate significantly impacts measures of mispricing from fundamentals (RD, RAD, and DF) or measures of price discrepancies relative to parity pricing (PD and LPD).

In conclusion, although statistical differences exist in trading volume and turnover rate between treatments and share types, our analysis does not reveal any significant impact of trading volume or turnover rate on asset mispricing or price discrepancy.

Table C.4: The linear regressions relative to the turnover and total trading volume.

Dependent variables	*RD*		*RAD*		*DF*	*PD*	*LPD*
Model	(1)	(2)	(3)	(4)	(5)	(6)	(7)
Intercept	-0.008	-0.022	-0.027	-0.040	0.006	0.036	0.005
	0.505	(0.512)	(0.354)	(0.358)	(0.355)	(0.391)	(0.312)
Trading volume of	0.0001	-	0.0003	-	-	-	-
each type shares	(0.001)	-	(0.001)	-	-	-	-
Turnover rate of	-	0.076	-	0.001	-	-	-
each type shares	-	(0.234)	-	(0.001)	-	-	-
Total trading volume	-	-	-	-	-0.00004	-0.0001	-0.0001
of both type shares	-	-	-	-	(0.0003)	(0.0003)	(0.0002)
Treatment	-0.024	-0.023	0.053*	0.055*	0.058**	0.057*	0.053**
(variable=1: WO)	(0.046)	(0.048)	(0.031)	(0.029)	(0.024)	(0.030)	(0.024)
A-shares dummy	-0.057**	-0.052	0.017	0.028	-	-	-
(A-shares=1)	(0.024)	(0.038)	(0.018)	(0.032)	-	-	-
The average score	0.001	0.001	0.002	0.002	0.002	0.002	0.003
on risk attitude	(0.010)	(0.010)	(0.006)	(0.007)	(0.007)	(0.007)	(0.006)
Sequence	-0.001	0.001	-0.022***	-0.023**	-0.024**	-0.036**	-0.031**
	(0.011)	(0.011)	(0.008)	(0.010)	(0.010)	(0.015)	(0.012)
Adjusted R^2	0.029	0.032	0.107	0.108	0.193	0.189	0.233
Observations	90	90	90	90	45	45	45

Note: "The average score on risk attitude" quantifies the average risk aversion level of each group, with a lower score indicating a higher degree of risk aversion. Robust standard errors, enclosed in parentheses, are clustered by group. In both panels, *, **, and *** indicate a significant difference from 0 at the 10, 5 and 1% significance levels.

Table D.5: Comparison of Total Number and Ratio of Short-Selling Transactions Between Treatments

	The total number of short-selling transactions			Ratio of short-selling transactions to total transactions		
	W ($n = 7$)	WO ($n = 8$)	p-value (H_0: W = WO)	W ($n = 7$)	WO ($n = 8$)	p-value (H_0: W = WO)
Seq 1	24	16	0.128	0.090	0.052	0.152
Seq 2	11	2	0.002	0.085	0.018	0.006
Seq 3	12	11	0.554	0.070	0.069	0.717

Note: The p-values are derived from two-sided Mann–Whitney U tests.

D Analysis of Short-Selling Behavior

This section examines whether and how restricting cross-asset arbitrage opportunities affects participants' short-selling behavior.

Table D.5 reports comparisons between treatments of the total number and ratio of short-selling transactions. The left side of the table presents comparisons of the total number of short-selling transactions. As indicated, the total number of short-selling transactions in Sequence 2 is significantly higher in the W treatment compared to the WO treatment at the 1% significance level. Although the total number of short-selling transactions is also slightly higher in the W treatment compared to the WOtreatment in Sequences 1 and 3, these differences are not statistically significant.

The right side of Table D.5 compares the ratio of short-selling transactions to total transactions. The results follow a similar pattern to the total number of transactions: the short-selling ratio is significantly higher in the W treatment than in the WO treatment in Sequence 2 at the 1% significance level, while differences are statistically insignificant in Sequences 1 and 3.

In summary, we observe a moderate treatment effect on short-selling behavior: allowing cross-asset arbitrage opportunities (W treatment) tends to increase short-selling activities. We interpret the higher frequency of short selling in the W treatment compared to the WO treatment as resulting from differences in endowment structure and trading restrictions. Specifically, each type-A participant initially holds 6 units of A-shares and no B-shares, while each type-B participant holds no A-shares and 3 units of B-shares. Participants in the W treatment, who can trade both share types, can immediately short sell the shares they do not possess. Conversely, in the WO treatment, participants must first sell their endowed shares entirely before engaging in short selling, limiting their immediate short-selling activities.