The Aggregate Impacts of Tournament Incentives in Experimental Asset Markets*

Debapriya Jojo Paul† Julia Henker‡ Sian Owen§

March 22, 2015

Abstract

Existing studies of the aggregate impacts of tournament incentives find that asset price bubbles in experimental markets are larger and do not dissipate with experience when participants trade under tournament incentives. However, these results potentially overstate the real-world impacts of tournament incentives for two reasons. First, they examine tournaments in a restrictive single-asset market setting, which constrains the risk-taking options available to traders. Second, by purely conferring additional rewards for good relative performance, the tournament contracts used ignore the risk-moderating role played by penalties that are also written into or implicit in real-world counterparts. We address these gaps by examining how prices behave under tournament incentives in experimental markets where participants can trade two assets with differing risk-levels. In addition, we compare price behaviour under tournament incentives with and without an embedded penalty for poor performance. Our findings suggest that the results in the existing literature are driven by the single-asset nature of their markets – we do not find any compelling evidence that prices in two-asset markets are more distorted under tournament incentives than normal incentives. Moreover bubbles in these markets do diminish with experience under tournament incentives. Also, while penalties embedded into tournament contracts are associated with less trading activity in markets compared to reward-only tournament contracts, they are also associated with longer periods of overvaluation and higher prices, albeit only with inexperienced traders. These results are particularly significant in light of the recent debate attributing financial market instability to convex incentive structures such as tournament incentives.

---

* We gratefully acknowledge financial support provided by the Australian Research Council (ARC DECRA grant no.: DE120101523).

† Send correspondence to the School of Banking and Finance, UNSW Business School, UNSW, Sydney 2052. Email: d.paul@unsw.edu.au. All errors are mine.

‡ Faculty of Business, Bond University, QLD 4229

§ School of Banking and Finance, UNSW Business School, UNSW, Sydney 2052

1. Introduction

Professionals in the intensely competitive world of finance routinely vie for 'prizes' such as bonuses, fund flows, and promotions that are tied to their performance relative to others. This gives many of the incentive schemes used in the industry the flavour of a tournament, which is characterised by compensation that depends on an employee's *relative* rather than absolute performance. The sizeable upside provided by these compensation structures, often not matched by an offsetting downside, creates 'convex' incentives. In light of the severe dysfunction that has punctuated financial markets this century – most notably during the collapse of the dot-com bubble and the US sub-prime mortgage crisis – concerns have been raised that such incentives may help precipitate market instability by encouraging excessive risk-taking and short-termism (Rajan 2006; Bebchuk and Spamann 2010; Wagner 2013)[1]. Yet despite the obvious importance of these concerns, the market-level effects of tournament incentives remain relatively unexplored and are not well understood.

On the one hand, that tournaments alter the risk-taking incentives of *individuals* is a well established point in the literature – since the most lucrative prizes go to winners, tournament incentives may encourage some contestants, particularly those who trail the leader, to take more risk in an attempt to 'win' (Bronars 1987; Hvide 2002; Cabral 2003; Tsetlin, Gaba and Winkler 2004). However, the implications of such behaviour for market prices have only been examined in a handful of studies, mostly in confines of the experimental laboratory where traders' incentives and asset fundamentals can be readily manipulated. The experimental studies, all conducted using the continuous double-auction bubble-market design of Smith, Suchanek, and Williams (1988), support the view of tournaments as a distortionary force in markets. They generally find that tournament

[1] See also Rajan (2008) and Blinder (2009) for treatments of this issue in the news media.

incentives exacerbate asset price bubbles – periods of sustained overvaluation – compared to absolute-performance based incentives, an effect that is, alarmingly, only magnified as participants gain more experience (James and Isaac 2000; Cheung and Coleman 2014).

In this paper, we extend the nascent experimental literature on the aggregate impacts of tournament incentives by addressing two issues affecting existing studies that potentially reduce the generalisability and real-world relevance of their results. First, we examine how market prices behave under tournament incentives when subjects can trade more than one type of risky asset. In contrast, existing studies only examine single-asset trading environments, which unduly restrict the risk-taking options available to traders compared to real-world markets. Investors seeking to 'get ahead' in the real world not only have the ability to speculate on a specific security, but also alter the risk profile of their portfolios by shifting into alternative, inherently riskier asset classes or securities. Thus it stands to reason that price behaviour in single-asset markets may differ from multi-asset markets.

Second, we investigate what impact adding a penalty for underperformance to a tournament contract has on market prices. Despite theoretical and empirical evidence suggesting that penalties, or 'sticks' in tournament contracts can curtail risk-taking by contestants (Gilpatric 2009; Qiu 2003; Kempf, Ruenzi and Thiele 2009; Hu, Kale, Pagani, and Subramanian 2011), the experimental literature pays scant attention to the role played by disincentives, focusing instead on the 'carrots', or rewards paid for good relative performance. Thus, rather than strictly being a tournament phenomenon, it may be that the heightened overvaluation seen in existing studies are driven by the absence of consequences attached to poor performance arising from excessive speculation. However, given that the fear of underperformance potentially encourages traders to herd (Rajan 2006; Dass, Massa and Patgiri 2008), the addition of a penalty may actually result in even higher prices, and hence is an open empirical issue.

To examine these issues, we implemented a between-subjects experimental design featuring three treatments that differ only in the way in which participants were remunerated – a normal incentive *Baseline* treatment, and two tournament treatments *Carrot* and *Stick*, where the latter is identical to *Carrot* but includes a penalty for underperformance in the form of a significantly reduced payment (zero!). While the compensation contracts in the *Carrot* and *Stick* treatments rewarded/penalised traders on the basis of their performance relative to the 'average' trader (as in James and Isaac (2000) and Isaac and James (2003)), we also implemented a set of alternative tournament contracts based on a rank-order tournament (i.e. where rank determines payoff), called *GilCarrot* and *GilStick.* Participants in all treatments traded in a Smith et al (1988)-type experimental asset market featuring two risky assets – a low-risk asset called X, which paid a modestly sized divided in each period, and a high-risk asset called Y, which paid a lottery-like dividend, thus allowing participants to more naturally vary risk than earlier studies have allowed.

Our first result suggests that the main conclusions in the existing literature are driven by the single-asset nature of their markets. We do not find any compelling evidence that tournament incentives – whether measured using the *Carrot, Stick, GilCarrot,* or *GilStick* treatments – distort prices more than absolute-performance based incentives (*Baseline*), as gauged by the size and duration of mispricing/bubbles in the markets of the respective treatments. Moreover, we find that bubbles under tournament incentives *do* moderate in size and duration as traders gain experience of the experimental design. In fact, evidence of improvement in price behaviour with once-experienced traders is weaker in the normal-incentive *Baseline* treatment than it is in our two tournament treatments.

On the impact of penalties, our second result finds that, in markets populated with inexperienced traders, embedding a penalty into a tournament contract that rewards traders for above-average performance (i.e. *Stick)* reduces the amount of trading activity compared to the corresponding reward-only contract (*Carrot*). However, consistent with the herding hypothesis, the trading activity that occurs in *Stick* markets is actually characterised by significantly *longer* booms (periods of overvaluation) in both risky assets, in addition to significantly *higher* prices in the high-risk asset. These differences however disappear with experienced participants. Moreover, we do not detect a significant difference in price behaviour between the rank-order tournament treatments, *GilCarrot* and *GilStick*, with inexperienced or experienced traders.

The remainder of this paper is structured as follows. In section 2, we review the related literature and develop our hypotheses. Section 3 details the experimental design, while section 4 describes our results. Finally, we present our conclusions in section 5.

2. Literature review and hypothesis development

2.1 Tournaments at the individual-level

The effect of tournament incentives on the behaviour of individuals is the subject of an extensive academic literature, the theoretical underpinnings of which have its formal beginnings in optimal labour market contracting under moral hazard. Starting with Lazear and Rosen (1981) and developed further by Green and Stokey (1983), Nalebuff and Stiglitz (1983), O'Keeffe, Viscusi and Zeckhauser (1984), Rosen (1986) and others, much of the early literature, along with the associated empirical work (e.g. Bull, Schotter and Weigelt 1987, Ehrenberg and Bognanno 1990) examines the comparative efficiency and optimality of the incentives provided by tournaments vis-à-vis other incentive structures such as piece rates. Importantly, the variable of interest and the only lever available to agents to affect their chances of winning in these early models is the amount of effort they choose to expend. Of course, players in a tournament can often also vary the amount of risk they take, and thus a sub-strand of the literature has emerged that focuses on risk-taking incentives in tournaments.

Bronars (1987, cited in Hvide 2002, p. 880) was the first to introduce risk-taking as a choice variable into a tournament model, finding that in sequential tournaments, it is optimal for leaders to reduce risk, while followers are inclined to increase risk in order to catch-up. The basic intuition underlying this result arises from the convexity of payoffs produced by tournaments. Faced with a win/lose dichotomy, the consequences of losing by a lot are the same as losing by a small margin[2]. Hence, laggards are better off 'going for broke' in the hope of maximising their chances of securing the larger prize earned by the winner(s), whereas leaders should try and 'lock in' their gains by playing conservatively. Results in line with this are also reported in a multi-period setting by Tsetlin et al (2004) and in an infinite-period model by Cabral (2003).

Other models however reveal a more nuanced relationship. Gaba and Kalra (1999) and Gaba, Tsetlin and Winkler (2004) show in a one-period setting that risk-taking incentives are sensitive to the proportion of players deemed 'winners'/'losers'. When the proportion of winners is low (specifically, less than 0.5), players have an incentive to 'break away from the herd' by increasing risk (as measured by variance). Conversely, when the proportion of winners is high (greater than 0.5), the priority is to avoid an especially poor performance, thus making a low-variance strategy optimal. Nieken and Sliwka (2010) demonstrate using a two-player model that the correlation between the outcomes of contestants' risky strategies is another important determinant of risk-taking preferences. When risky outcomes are uncorrelated between players – as is typical of most tournament models in the literature – the leader (laggard) prefers to play it safe (take risks), provided the additional expected return

[2] This describes most tournament models in the literature, where 2 (or more) players compete over two levels of prizes differentiating winner(s) from loser(s), W_1 and W_2, where $W_1 > W_2$.

from the risky strategy is sufficiently small relative to size of the lead. However, as the correlation increases, it becomes more attractive for the leader to mimic the (anticipated) risky strategy of the trailing agent as a means of maintaining their lead. Of course, the trailing agent is aware of this, hence at high correlations (>0.5), a mixed strategy equilibrium may exist in which the leading player chooses the risky strategy with a higher probability than the trailing player.

While the aforementioned studies of risk-taking in tournaments ignore effort as a choice-variable and consider only the risk-level, Hvide (2002) combines the two by examining a one-period symmetric tournament where players simultaneously choose both the mean (effort) and variance (risk) of their output. In equilibrium, all participants adopt the highest possible level of risk and expend low effort; since expending effort is costly, players have a common incentive to take high risk because it induces noise in the level of output, which makes differences in effort less important to their chances of winning/losing[3]. However, Kräkel and Sliwka (2004) show that the uniform preference for high risk and low effort does not necessarily hold when contests are asymmetric. They model a two-player tournament where risk-neutral players differ in ability (or equivalently, their relative starting positions), and choose risk first and effort second. They find that diverse equilibria are possible, with the exact equilibrium depending on the interplay of a number of factors including the magnitude of the difference in abilities, the associated interaction between the effect of risk-taking on effort and on the probability of winning, the shape of the cost-of-effort function, and the prize spread. Although no equilibrium in their model sees the high-ability/leading agent adopt a high-risk strategy whilst the low-ability/trailing agent takes a

[3] In the extreme case where risk is unbounded, players make zero effort and take an infinite amount of risk, causing the tournament compensation scheme to fail. In the case of bounded variance, the prize spread (the difference between the winning and losing prizes) can be adjusted to maintain first-best levels of effort (i.e. an efficient tournament contract) when players are risk-neutral. However in their model, tournaments will be less efficient than piece rates when agents are risk-averse.

low-risk strategy, the reverse does not always hold (i.e. low-risk for high-ability/leaders and high-risk for low-ability/laggards, as in Bronars (1987)).

Furthermore, in a result that holds particular significance to our study, Gilpatric (2009) shows that asymmetry in the prize structure of a tournament can also affect the incentive to take risk. Specifically, Gilpatric demonstrates that adding a third payoff level – an explicit penalty for finishing last (a 'stick') – to the customary prizes for the winner (a 'carrot') and the also-rans in a winner-takes-all contest can curb risk-taking by risk-neutral contestants. In the presence of a penalty for severe underperformance, those who trail the leader no longer increase risk (i.e. variance) with impunity, since increasing risk also entails a greater possibility of finishing with an even lower payoff. In the model, the precise amount of risk-taking in equilibrium can be controlled by adjusting the relative sizes of the carrot (the additional reward to the winner) and the stick (the penalty for coming last) – the larger the carrot relative to the stick, the greater the incentive to engage in risk-seeking behaviour.

In addition to the theoretical literature, a growing body of research has examined risk-taking in tournaments empirically. In finance, the relevance of tournament theory to the funds management industry has attracted much interest[4]. Brown, Harlow and Starks (1996) argue that mutual fund managers engage in annual contests with each other because their compensation is typically tied to the value of funds under their management, which in turn depends on their recent performance relative to other funds – the best-performing funds receive the largest inflows of new funds, while those performing poorly do not experience similar-scaled outflows (Sirri and Tufano 1998). This convexity in the relationship between relative performance and compensation motivates Brown et al to hypothesise that fund

[4] The dominance of relative-performance concerns and the presence of highly incentivised contestants mean that the world of professional sport has also attracted significant empirical interest. A number of studies look at risk-taking by contestants in motorsports (Becker and Huselid 1992; Bothner, Kang and Stuart 2007), golf (Brown and Li 2010), basketball (Grund, Höcker, and Zimmerman 2013), weightlifting (Genakos and Pagliero 2012), and high-stakes poker (Lee 2004).

managers who are 'losing' mid-way through the year will increase the risk of their portfolios more than mid-year 'winners'. Although they find evidence supporting their hypothesis in their sample, subsequent research has provided mixed, often contradictory results[5]. These conflicts can (at least partially) be reconciled by the aforementioned contributions to tournament theory that show that it is not always optimal for laggards (leaders) to be more risk-seeking (conservative)[6]. Moreover, significantly for our study, a number of empirical studies lend support to the theory posited by Gilpatric (2009) that penalties/disincentives serve to moderate risk-taking – Qiu (2003), Kempf, Ruenzi and Thiele (2009), and Hu, Kale, Pagani, and Subramanian (2011) all observe that greater termination risk (the risk of job-loss) has a negative effect on risk-taking by fund managers.

2.2 Tournaments at the market-level

In contrast to the substantive literature on tournament behaviour at the individual-level, the market-level impacts of tournaments have received relatively little attention. To the best of our knowledge, a series of experiments by James and Isaac (2000) and Isaac and James (2003) represent the first and until recently, only attempt by researchers to study the aggregate effects of tournament incentives. Noting that tournaments can alter individuals' risk-taking incentives, the question these studies pose is whether tournament incentives

[5] See for example, Chevalier and Ellison (1997), Koski and Pontiff (1999), Busse (2001), Elton, Gruber and Blake (2003), Qiu (2003), Goriaev, Nijman and Weker (2005), Kempf and Ruenzi (2008), Chen and Pennacchi (2009), Kempf, Ruenzi and Thiele (2009), Elton, Gruber, Blake, Krasny and Ozelge (2010), and Hu, Kale, Pagani, and Subramanian (2011).

[6] Brown et al (1996) have also inspired a number of theoretical tournament models of the mutual fund industry. Taylor (2003) tackles the issue of risk-taking by leaders/laggards in a two-player mutual fund tournament. In a strategic setting where the risky strategy yields the same return for both players, the (mixed-strategy) equilibrium is characterised by the mid-year winner being more likely adopt a high-risk strategy than the loser. Taylor's model is a special case of the model developed by Nieken and Sliwka (2010), where the correlation between contestants' risky strategies is set at 1. Bagnoli and Watts (2000) consider the risk choices of fund managers in the presence of return-chasing investors (i.e. in a tournament). They show that risk-neutral fund managers will invest in riskier portfolios compared to the case where investors don't chase returns, and this behaviour will be amplified if investors select funds based on rankings rather than performance relative to the average. Acker and Duck (2006) examine the propensity of fund managers to take 'extreme' positions (mostly cash or mostly shares) in a 2-period model where one of the managers is an exogenous passive 'benchmark'. They find that trailing funds are more likely to take extreme positions, especially if they are far behind, or as the end of the tournament approaches.

distort market prices by fuelling speculative asset price bubbles. Using a within-subject design and the oft-replicated Smith et al (1988) double-auction market as a baseline, they examine how prices are affected by the introduction of a tournament condition that rewards traders on the basis of their performance relative to the 'average' trader. Typically, when traders are compensated according to their absolute performance, prices converge quickly to fundamental value in markets consisting of (twice) experienced traders. However, this convergence fails to occur once James and Isaac introduce tournament contracts. In fact, repeated exposure to tournament incentives causes prices to deviate further from fundamental value. James and Isaac explain this by showing that it may be mutually advantageous (i.e. rational) for risk-neutral traders to transact at prices above (or below) fundamental value under their tournament contract.

A recent study by Cheung and Coleman (2014) reinforces the main results of James and Isaac (2000). They investigate prices under tournament incentives in both declining (i.e. Smith et al 1988) and constant fundamental value markets but use a different tournament compensation contract, based on the mutual fund industry's convex performance-fund flow relationship; they also use a between-subjects design, which is free of the order effects that can afflict within-subject designs[7]. Somewhat contrastingly, Robin, Straznicka, and Villeval (2012) find that long-term and short-term competitive bonuses have differing effects. Their long-term contract pays a bonus at the end of the market based on relative performance over the course of the entire market, whereas the short-term contract awards a bonus at the end of each trading period. They find that their long-term bonus contract produces less price distortion than both short-term bonus and normal incentive contracts, although it is unclear if

[7] Specifically, Cheung and Coleman (2014) detect significantly larger bubble Amplitudes and Durations under tournament incentives in inexperienced Smith et al (1988) markets. Differences become larger in experienced Smith et al markets, across a wider range of bubble measures. The effect of tournaments in their constant fundamental value markets is milder than in declining fundamental value markets, but nonetheless is still sizeable.

their results are driven by the incentives or the changing liquidity of the market that is induced by the payment of bonuses in their short-term contract (Palan 2013)[8].

A common element of the design of these experimental studies is their use of markets featuring only one type of risky asset. This poses a potential problem for the generalisability of their results because unlike real markets, where investors have the opportunity to alter portfolio risk by shifting between a variety of asset classes and securities that are intrinsically more/less risky or speculative in nature, the risk-taking options for traders looking to 'win' in a single-asset environment are extremely limited – they are restricted to simply acquiring more of the same asset by paying higher prices in the hope of selling at a profit or getting lucky with high dividend payments (or both). Therefore, it is uncertain how applicable the behaviour elicited by single-asset markets is to multi-asset environments, or indeed the real world. Thus, by better approximating real-world markets, an experimental market containing more than one type of risky asset should allow for more natural risk-shifting behaviour, and thus better scope to understand the aggregate impacts of tournament incentives. Our study fills this gap in the literature by examining if tournament contracts distort prices more/less than absolute-performance-based incentives in experimental asset markets where participants can simultaneously trade *two* differentiated risky assets. Thus our first hypothesis, stated in the null, is:

Hypothesis 1: Price behaviour does not differ between tournament markets and normal-incentive markets.

[8] Ang, Diavatopoulos and Schwarz (2010) also implement tournament incentives in an experimental market, but only in every alternate trading period as an additional compensation scheme, with the aim of creating a shortened investment horizon for traders. This significant departure in methodology makes it is difficult to place their results amongst the other experimental literature. Ang et al find that the effect of shortened horizons/tournament incentives on bubbles depend on the risk-attitudes of the traders in their markets and whether participants trade with their own money.

We also examine if the reported tendency for bubbles to worsen with trading experience under tournament conditions is sustained in a two-asset environment, leading to our second (null) hypothesis:

Hypothesis 2: Price behaviour under tournament incentives does not differ between markets containing traders who are inexperienced with regards to the experimental design versus once-experienced traders.

Another attribute that the existing experimental studies have in common lies in the design of their tournament compensation contracts – being solely characterised by the payment of additional rewards for good relative performance, their contracts are all ‘carrot’, no ‘stick’. For instance, James and Isaac (2000) pay a flat fee to traders who perform below average, while above-average performers are rewarded with an additional bonus that is proportional to the degree to which they outperform the average. Similarly, Cheung and Coleman (2014) and Robin et al. (2012) periodically award new funds to traders based on their relative performance in the previous sub-period of the market. In doing so, these studies overlook the importance of disincentives or ‘sticks’ in employment contracts, specifically penalties attached to poor performance, which may serve to moderate risk-taking (Gilpatric 2009). Hence, it may not be tournament incentives driving their results per se, rather the balance, or lack thereof, between ‘carrots’ and ‘sticks’ in their contracts.

However, while introducing a penalty into a tournament contract may deter *individuals* from taking risks, the implications for market prices are less clear. On the one hand, penalising underperformance may produce lower prices by discouraging traders from bidding excessively for assets, since doing so makes them more likely to underperform. However, this fear of underperformance may perversely result in *higher* prices. Rajan (2006) argues that relative-performance based compensation encourages investment managers to herd, since herding reduces the chances of underperforming. While this in itself may create a bubble, it also means fund managers may choose to 'ride the bubble' because the alternatives of trading against it or doing nothing expose them to the risk of underperforming if the mispricing persists. As a result, bubbles may 'inflate' further and last longer. In this context, herding becomes the safe strategy while the risky strategy is to deviate (Dass et al 2008). Of course, all of this crucially relies on there being real consequences for underperforming. Since including a financial penalty in a tournament contract makes the consequences more salient compared to a penalty-free contract, the tendency to herd and thus the severity of asset price bubbles may be greater in the presence of a penalty. Furthermore, Dass et al argue that bonuses for outperforming the competition ('carrots') may actually help to deflate bubbles by inducing fund managers to try and win the tournament, something that can only be achieved by leaving the safety of the herd. In support of this, they find that during the dot-com bubble, the more highly incentivised fund managers had smaller holdings of so-called 'bubble stocks'.

Our study extends the experimental literature on tournaments by seeking to resolve the uncertainty surrounding the aggregate-level impacts of penalties. While Isaac and James (2003) also consider a tournament contract with an explicit penalty for severe underperformance, they only run two sessions and unsurprisingly obtain inconclusive results. In contrast, we comprehensively investigate whether including penalties for underperformance affects the severity of mispricing/bubbles compared to tournament contracts that contain no such penalties. Hence our third hypothesis, stated in the null, is:

Hypothesis 3: Prices do not behave differently between 'carrots'-only tournament markets and 'carrots-and-sticks' tournament markets.

In examining the above hypotheses, our study also contributes to the literature on bubbles in multi-asset experimental markets, first studied by Fisher and Kelly (2000). Like our study, participants in these experiments typically trade two different assets in a market that mimics the basic Smith et al (1988) continuous double-auction design. Research following Fisher and Kelly (2000) has examined how prices in these markets behave when assets becomes differentiated by characteristics such the mean and/or the variance of payoffs, or maturity (see Ackert, Charupat, Church and Deaves 2006; Childs and Mestelman 2006; Chan, Lei, and Vesely 2013). We build on this literature by introducing tournament incentives into a two-asset market, whereas all research has hitherto been based on 'normal', or absolute-performance based incentives. In addition, we examine the effect of trading experience on bubbles in multi-asset markets, which to our knowledge has not been investigated before.

Our paper perhaps has most in common with Kleinlercher, Huber and Kirchler (2014), who also examine the effect of different incentive schemes on price behaviour in a two-asset experimental market. Like ours, their incentive schemes include an option-like, reward-only "Bonus" contract and a "Penalty" contract. However, the key difference between their study and ours is that whereas we examine tournament incentives, they do not. Rather, their "Bonus" and "Penalty" treatments are absolute-performance based compensation schemes featuring an *exogenous* benchmark, specifically a pre-defined final cash balance. In contrast, the benchmarks in tournament schemes like ours are *probabilistic*, such as the performance of the "average" trader[9]. Since optimal risk-taking behaviour may differ for individuals faced with an exogenous benchmark versus a contest scenario (Taylor 2003; Tsetlin et al 2004), the aggregate implications may also vary.

3. Experimental Design

The experiment comprises 35 independent markets carried out across 19 sessions at the ASB Experimental Research Laboratory at UNSW Australia between August and November 2013, with 261 subjects taking part across all treatments. Participants were university students with no prior experience in market experiments, recruited using ORSEE (Greiner 2004)[10]. We begin by describing the parameters of the market institution that were common to all sessions before detailing the specific treatment variables. We finish with an overview of the procedures followed in each session.

3.1 Market structure

In each session, participants were given the opportunity to trade two types of assets concurrently, one called "X", the other called "Y". The market for both assets ran for 12 periods, each lasting 3 minutes[11]. Trade occurred according to continuous double-auction

[9] To further highlight the difference, Kleinlercher et al (2014) point out that it is possible under their bonus compensation contract for *all* traders to receive a bonus, since with a favourable dividend outcome, all traders could exceed the benchmark-level of cash. In the absence of collusion, this is not possible in tournament schemes where bonuses are paid for above-average performance, since almost certainly *someone* will perform below average.

[10] In total, 38 markets were run. However, some participants with multiple ORSEE profiles managed to 'slip through' and participated in more than one session of this experiment. To mitigate the potential confounding of treatment effects, we have excluded from our analysis any data from the 3 markets that contained a subject who had participated in an earlier session.

[11] While experimental studies of tournament incentives have largely stuck with the parameters in Smith et al (1988), there is considerable heterogeneity in studies involving multiple assets. The number of trading periods in these studies ranges from 12 (Ackert et al, 2006) to 30 (Chan et al, 2013), while trading period lengths vary between 3 (Chan et al 2013) and 6 minutes (Fisher and Kelly, 2000). The parameters chosen for our sessions are consistent with the lower end of this range, and represent suitable compromise given constraints posed by budgets and time. In particular, we were mindful of avoiding sessions that were "too long" and risked inducing boredom in participants, given the repetitive nature of market experiments.

rules; participants were allowed to post bids and asks for both assets in separate open order books, and accept any posted bid or ask for either asset, subject to the constraints posed by their asset holdings and cash balance. All trade occurred in single units, and short-selling and buying on margin were not permitted. Trade was conducted in experimental currency called ‘francs’, with earnings being paid out at the end of the experiment in Australian dollars at a pre-announced exchange rate of 200 francs to 1 Australian dollar. The market institution was fully computerised using zTree (Fischbacher, 2007) – the trading interface is shown in Figure 1[12].

< Insert Figure 1 about here >

All traders began the market with the same initial endowment of assets and cash – 5 units each of X and Y, and 1950 francs. This ensured that the relative position of any trader in the market was not affected by the composition of their initial allocation, and also that the expected earning opportunities for all traders were initially the same. At the end of each trading period, Asset X paid a cash dividend drawn from the distribution {10, 30} with equal probability, while Asset Y paid a dividend from the distribution {0, 100} with respective probabilities (0.8, 0.2)[13]. These distributions were known to all participants. Dividend draws, which were made by the computer, were independent across trading periods and between the two types of assets. Any dividend earnings were added to the trader’s cash balance, and their end-of-period portfolio carried over to the next trading period.

[12] Note that given the previously documented tendency for trading activity to be biased in favour of the market that appears on the left-hand side of the screen (see Chan et al 2013), the market for Asset X was placed on the left for roughly half of the sessions in each treatment, and on the right for the remainder.

[13] These dividend structures mimic that of Ackert et al (2006), who also use a standard/lottery-asset dichotomy, albeit with a much more pronounced difference in potential payoffs between the two types. Their ‘standard’ asset’s dividend distribution is {0.50, 0.90, 1.2} with respective probabilities (0.48, 0.48, 0.04), while their ‘lottery’ asset pays a dividend from the distribution {0, 18} with associated probabilities (0.96, 0.04). The maximum possible payoff in a period from their lottery asset is 15x the maximum payoff from the standard asset, whereas the corresponding multiple in our study is 3.33x. This is intentional, as we wanted participants to still view Asset Y as a viable “investment” rather than a purely speculative bet.

Note that the expected dividend paid by both X and Y in each period is 20 francs. Hence, the risk-neutral fundamental value (FV) of both assets is the same, and is equal to the expected total future dividend stream, or 20 multiplied by the number of trading periods remaining (including the current period)[14]. As shown by the solid black line in Figure 2, the resulting risk-neutral FV process of both assets begins at 240 in period 1 and declines in steps of 20 in each period, falling to 20 in period 12 before expiring worthless after the final dividend is drawn at the end of period 12. The FV process represents another difference between our study and Kleinlercher et al (2014). Whereas we adopt the declining FV environment of Smith et al (1988) in line with other experimental research on tournament incentives and multi-asset experimental markets, they study experimental assets that have a constant FV. In comparison to declining FV markets, constant FV markets of the type examined by Kleinlercher et al are less prone to bubble under normal incentives (Smith, van Boening, and Wellford 2000).

Figure 2 also illustrates the largest (dotted line) and smallest (solid line) possible cumulative future dividend realisations of each asset in our experiment (X in grey, Y in blue). To keep the other features of the graph from being obscured, the step function for the maximum possible dividends from asset Y – which potentially pays 100 francs in each period – is only partially displayed; the blue dotted line begins at 1200 in period 1 and falls in steps of 100 in each ensuing period. In contrast, the minimum possible cumulative dividend payment from asset Y is zero, while asset X pays at least 10 and possibly 30 francs in each period. These step functions serve to demonstrate that although both asset types have the same expected dividend, the variance of Y's dividend payoff is much greater than X's. As asset X always pays out at least 10 francs in each period, it represents a 'safe' investment,

[14] The expected value of the total future dividend stream was common knowledge, and was communicated to participants in the form of an "average holding value" table contained within the written instructions given to all participants.

whereas Y with it lottery-like characteristic is riskier/more speculative. This presence of a second, risky/speculative asset in the market environment provides a more natural and realistic avenue for traders to increase risk in the hope of greater reward than what the single-asset environments of earlier tournament studies provide.

< Insert Figure 2 about here >

The parameters discussed above determine the initial liquidity of our markets, as measured by the initial cash-to-assets ratio – the ratio of total cash to the total intrinsic value of all assets (X and Y) at the beginning of the market. This ratio was 0.8125 in all sessions, allowing us to control for the effects of liquidity on prices, which is known to be positively associated with the magnitude of bubbles in experimental markets (Caginalp, Porter and Smith 1998, 2000, 2001). While existing experimental studies of tournament incentives and multiple assets have used a variety of initial cash-to-asset ratios, our choice of 0.8125 reflects the cash-to-assets ratio in the most oft-replicated Smith et al (1988) design, as well as being the initial liquidity used by Cheung and Coleman (2014) in their tournament study.

3.2 Treatments

3.2.1 James and Isaac tournament contracts

To examine the influence of rewards (‘carrots’) and penalties (‘sticks’) in tournament contracts, we implemented a between-subjects design with 3 treatments that differ in the way participants were remunerated for their performance in the market. In the *Baseline* or ‘normal’/linear incentives treatment, participants were compensated on the basis of their absolute performance in the market. Since Assets X and Y expired worthless at the end of the market, this means that traders were paid their final cash balance[15].

[15] Ending cash balance = initial cash balance + dividend earnings + sales revenue – expenditure on purchases

The remaining two treatments invoke tournament incentives. In both the *Carrot* and *Stick* treatments, we mirrored the approach taken by James and Isaac (2000) and Isaac and James (2003) by compensating traders on the basis of their performance relative to the 'average' trader. The *Carrot* compensation contract rewarded above-average performance with a bonus payment, while paying all other traders a fixed amount, using the following rule:

$$Earnings_i = \begin{cases} 3000 & if \quad C_i < C^* \\ 3000 + 2(C_i - C^*) & if \quad C_i \geq C^* \end{cases}$$

C_i is the final cash balance of trader *i* and C* is the average of the final cash balances of all traders in the market. All units and amount shown are denominated in francs.

The compensation contract in the *Stick* treatment introduced an additional component to the contract used in the *Carrot* treatment – a penalty intended to reflect the consequences of a scenario where a trader performs so poorly that they lose their job.

$$Earnings_i = \begin{cases} 0 & if \quad C_i < \frac{1}{2}C^* \\ 3000 & if \quad \frac{1}{2}C^* \leq C_i \leq C^* \\ 3000 + 2(C_i - C^*) & if \quad C_i > C^* \end{cases}$$

While our *Carrot* contract and the "Bonus" contract used by Kleinlercher et al (2014) have similar, convex functional forms, our *Stick* contract differs markedly from their "Penalty" contract, which deducts a proportional penalty from a fixed payment, effectively placing a cap on traders' earnings[16]. Hence unlike their study, a comparison between our *Carrot* and *Stick* treatments indicates *only* the effect of introducing a penalty for poor performance.

At the end of each trading period, participants in the two tournament treatments were given information on-screen about their relative performance. Specifically, they were informed of the value of their own Account Total and the average Account Total in their market. Based on a measure of the same name used by Schoenberg and Haruvy (2012), Account Total is akin to the market value of a trader's portfolio, and is defined as the sum of a trader's end-of-period cash balance and the value of their end-of-period asset holdings; the end-of-period holdings of X and Y in our study were valued at their respective median traded prices in that period. Like Cheung and Coleman (2014), we chose the median price in preference to the final trading price or highest bid (as used by Schoenberg and Haruvy) because it is more difficult for traders to manipulate[17]. Since all assets expired worthless after the final dividend payment, the Account Total at the end of period 12 (i.e. at the end of the market) reverted to the final cash balance[18]. Traders in the *Baseline* treatment were also informed of their own Account Totals at the end of each trading period, but were not told the average in their market.

3.2.2 Gilpatric tournament contracts

We also tested two alternative tournament treatments, *GilCarrot* and *GilStick*, which more closely reflect the type of tournament modelled by Gilpatric (2009). Being rank-order tournaments, participants in these treatments were paid a *fixed* amount determined purely by

[16] That is, the "Penalty" contract used by Kleinlercher et al (2014) is a penalty-only contract, whereas our *Stick* contract is a bonus-and-penalty contract.

[17] In periods where there was no trade in an asset, the median transaction price was replaced by the median buy offer for that asset in the period. This was done to avoid misleading fluctuations in the Account Total, and participants were made aware of this before the market began.

[18] This small change in the definition of the Account Total for period 12 was necessary, since otherwise, it would create an incentive for participants to arbitrarily bid up the prices of assets X and Y in period 12 in the hope of maximising their Account Totals.

their relative position, specifically their final rank. Our *GilCarrot* contract paid the trader with the largest final cash balance 10,000 francs, while all other traders received the significantly lower payment of 4000.[19] The *Gilpatric Stick* contract is the same, except the worst performing trader – the trader with the lowest final cash balance – received nothing from the market. Contrast these with the 'James and Isaac' tournaments contracts described above, where payoffs depend not only on being better/worse than average but also the extent to which a trader's absolute performance exceeds the average. By severing any link between absolute performance and compensation, the 'Gilpatric' contracts can be considered 'purer' tournaments, in the Lazear and Rosen (1981) sense, where *only* relative performance matters. Since the appropriate piece of relative-performance information in these treatments is the trader's rank, participants in Gilpatric tournament treatments were informed of their rank at the end of each period (calculated on the basis of Account Total), in addition to the other relative performance information described above.

3.3 Procedures

Each experimental session corresponded to a single treatment to which it (and hence, each subject within it) was randomly assigned[20]. Sessions were designed to run two independent market-groups of (up to) 8 traders each and ran for approximately 2.5 hours[21]. To ensure consistency in the delivery of instructions between sessions and reduce experimenter demand effects, all participants received written instructions, which were also

[19] The minimum payment here was set to 4000 francs compared to 3000 francs in the equivalent James and Isaac tournament contract *Carrot* to ensure that the average compensation per trader in real currency, Australian dollars, was roughly equal across treatments, and to also conform to the ASB Lab ethics protocol which specified an average payment range of $15-20 per hour per participant.

[20] The only exception to this was a single session where a *Carrot* treatment market ran alongside a *GilStick* market. The instructions and procedures were appropriately modified for this session to prevent contamination of the subject pool.

[21] That is, excluding the practice period, participants only traded with other participants who were in the same market-group. Dividends were also drawn independently for each market-group.

communicated verbally by the experiment administrator[22]. Potential interaction effects between participants were mitigated by prohibiting subjects from communicating with each other for the duration of the experiment.

The procedure followed in each session was identical, regardless of the treatment. Sessions began with participants being randomly allocated to a computer/workstation that determined their market-group[23]. They then received training on how to use the trading screen to make and accept bids and offers for each asset (10 minutes), following which they were given 10 minutes to practise trading using the interface. After the practice period, subjects were given further information about the other features of the market environment, including how their earnings would be calculated. After this, the market-proper began. Upon the conclusion of the market, participants were informed that they would be taking part in another 12-period market with the same traders (i.e. market-group). Participants' inventory of assets and cash were reset to their starting levels, and trading commenced for a second round.

After the end of the second round, participants completed an untimed survey consisting of 3 sections[24]. The first section gathered general demographic information about participants and their experiences and thought-processes during the market(s)[25]. The second and third sections, which form part of a related study, comprise the *Cognitive Reflection Test* (CRT) and *Domain-Specific Risk-Taking* (DOSPERT) Scale. The CRT is a measure of cognitive ability developed by Frederick (2005) that consists of 3 problem-solving type questions that assess the ability of respondents to reject an impulsive and intuitive incorrect

[22] To ensure consistency with the procedures used in the existing literature, the written protocol was adapted from those used by Dufwenberg et al (2005), Noussair et al (2001), Noussair and Powell (2010), Lugovskyy et al (2009), Childs and Mestelman (2006), and Cheung and Coleman (2014). Participants were also given time to read the instructions on their own, and to ask any clarifying questions privately (which were also answered privately). The written protocol can be found in Appendix B.

[23] The workstation number also served as a participant's ID, thus ensuring the anonymity of their data.

[24] The survey was initially paper-based (9 sessions), but was computerised using the *Qualtrics* survey software and administered electronically in the October and November sessions (10 sessions).

[25] This is a modified version of the end-of-experiment questionnaire used by Ackert and Church (2001).

answer in favour of a correct answer that requires more deliberation. In addition to general measures of cognitive ability, performance in the CRT is correlated with time and risk preferences (Frederick 2005), as well as certain behavioural biases (Oechssler, Roider, and Schmitz 2009). The 30-item DOSPERT Scale, designed by Blais and Weber (2006), is a psychometric scale that measures risk preferences and perceptions across five separate decision-making domains: Financial (split into Investing and Gambling), Health/Safety, Recreational, Ethical, and Social[26]. Respondents use a 7-point scale to rate the likelihood of their participation (Part 1), the perceived riskiness (Part 2), and the benefits expected to accrue (Part 3) from engaging in 30 different domain-specific risky activities. Of course, administering the DOSPERT Scale after the market stage carries with it the risk that responses may be influenced by participants' experiences during the market. However, given our main objective is to study price behaviour, this is the 'lesser of two evils', as the alternative of implementing the scale before the market could in turn influence participants' trading behaviour. A summary of the demographic characteristics of the subject pool, CRT scores (out of 3), and DOSPERT likelihood/preference scores in the most relevant domain, Financial (ranges from 6 to 42, higher scores indicate greater willingness to take financial risks), is presented in Table 1, categorised by treatment.

Once the surveys were completed, participants were called up individually, paid their earnings (in envelopes) and dismissed. Participants' total earnings from the experiment were calculated as the sum of their earnings from both rounds of the market, converted to Australian dollars, plus a $5 participation fee. The average payment to participants, inclusive of the participation fee, was $49.

< *Insert Table 1 about here* >

[26] Compared to the original 40-item DOSPERT scale (Weber, Blais and Betz 2002), which was developed for American undergraduate college students, the revised 30-item DOSPERT scale (Weber and Blais 2006) is designed to be more readily applicable to a more diverse range of cultures, age groups, and educational levels. Consequently, we chose the latter.

4. Results

4.1 Inexperienced Traders

4.1.1 Descriptive Summary

Panels (a) and (b) of Figure 3 chart the time-path of the median transaction price of assets X and Y respectively in the *Baseline*, *Carrot*, and *Stick* treatments during the first round of the market; for each treatment, the charted price in each period is the median of the median transaction prices from all markets in that treatment. Median prices in Figure 3 for both assets in all treatments broadly follow the pattern associated with Smith et al (1988)-type markets populated with inexperienced traders – prices start below fundamental value and remain there in the initial periods before rising above fundamental value. However, with the possible exception of asset Y in the *Stick* treatment, which experiences a precipitous fall in median price from periods 7 to 8, the characteristic bubble-and-crash is notably missing. In fact, median prices in the *Carrot* treatment can hardly be described to ‘bubble’ at all, though it should be noted that these graphs hide considerable heterogeneity at the individual market level. In fact, bubbles-and-crashes were observed in individual markets of all treatments, although they did not occur with the regularity reported in other studies of multi-asset experimental markets such as Fisher and Kelly (2000).

Perhaps the most notable feature of Figure 3 is the persistently higher median prices/more pronounced overvaluation exhibited by the *Stick* treatment in comparison to the *Carrot* treatment. In fact, median prices are higher in the *Stick* treatment in every trading period for the more speculative asset Y, and in all bar 1 period for asset X. A Wilcoxon Mann-Whitney (WMW) U test – the non-parametric equivalent of the independent samples t-test – reveals that these differences are significant at the 5% level in period 2 though to 6 in asset X, and in periods 2 and 5 for asset Y[27]. Furthermore, it is the *Carrot* treatment where prices appear to most closely conform to fundamental value, even in comparison to the normal-incentive *Baseline* treatment, which can be tentatively described as charting a path in between the two tournament contract treatments, particularly for asset X. These observations run contrary to the notions that tournaments necessarily distort prices, and that rewards (penalties) encourage (discourage) the formation of bubbles. Moreover, they also present a sharp contrast to Kleinlercher et al (2014), who find that average prices for their "high-risk" (equivalent to our asset Y) asset are highest under their "Bonus" treatment and lowest in their "Penalty" treatment.

Figure 3 also reveals a high degree of correlation in the prices of assets X and Y, which is consistent with behaviour observed in earlier studies of multi-asset markets (e.g. Fisher and Kelly 2000, Childs and Mestelman 2006), where *relative* prices between asset-types tend to remain close to the 'correct' value (i.e. risk-neutral value) even when individual assets exhibit severe mispricing. This is more clearly illustrated by Figure 4, which graphs the median *Prediction Error* in each period for each treatment. Like Fisher and Kelly (2000), we define the *Prediction Error* in each period of an individual market as the percentage deviation of the relative price of asset Y (median price of Y divided by the median price of X in that period) from the risk-neutral benchmark (equal to 1 in this study)[28]. More positive (negative) values indicate a greater willingness by market participants to pay a premium to acquire the riskier (less risky) asset Y (X). As Figure 4 illustrates, median prediction errors in all treatments remain relatively close to zero throughout the market.

[27] While we do not report full results here, the two-sided p-values in periods 2-6 for asset X are 0.045, 0.049, 0.048, 0.024, and 0.027. For asset Y, the p-values in period 2 and 5 are 0.014 and 0.046 respectively. In addition, median transaction prices for asset Y are higher in the *Stick* treatment than the *Carrot* treatment at the 10% level in periods 1 and 6 (p-value = 0.094 and 0.066 respectively).

[28] Unlike Fisher and Kelly (2000), we report the median of the Prediction Errors across all sessions/markets rather than the average, due to the lower sensitivity of the median to outliers in small samples.

Furthermore, mirroring the approach of Brown et al (1996) by comparing the first and second half of the market in Figure 4 does not indicate the presence of an obvious 'tournament effect' in the two tournament treatments. The effect we seek to detect here is heightened 'risk-seeking' behaviour by traders in the second half of the market, as evidenced by a substantial rise in the price of Y relative to X[29]. Instead, we see that relative prices in *Carrot* and *Stick* behave similarly in both halves – the average of the median *Prediction Errors* in the *Carrot* treatment in periods 1-6 is 1% vs. 2.2% in periods 7-12, while the corresponding values for the *Stick* treatment are 1% and 2.4%. In contrast, the average of the median *Prediction Errors* during the first half of the market in the *Baseline* treatment is -0.2%, compared to 10% in the second half, potentially indicating that participants were willing to pay more to acquire the riskier asset Y in the latter stages of the market. Note however that the desire to move up the leaderboard is unlikely to be an adequate explanation for this apparent risk-seeking behaviour since relative performance information was not shown to participants in the *Baseline* treatment[30].

< *Insert Figure 3 about here* >

< *Insert Figure 4 about here* >

4.1.2 Statistical Analysis

<u>Bubble Measures</u>

To conduct a more formal comparison of the treatments, we calculate a number of measures of mispricing/bubbles that are frequently used in the experimental asset market

[29] Of course, Brown et al (1996) were concerned with adjustments made by individual fund managers in portfolio risk between the two halves of the year rather than an aggregate metric like relative price.

[30] Having said that, it is possible that even if relative performance information is not provided, participants have an internalised benchmark of what "average" performance looks like, and hence whether they are performing better or worse.

literature. These bubble measures can broadly be categorised into two groups that assess two different dimensions of mispricing – magnitude and length.

Amplitude (Haruvy and Noussair 2006), the first of the magnitude measures, quantifies the extent to which average prices in a market change relative to FV. It is calculated as $\max_t\{(\bar{P}_t - F_t)/F_t\} - min_t\{(\bar{P}_t - F_t)/F_t\}$, where the largest and smallest deviations of average price P_t from fundamental value F_t are normalised by the FV in the respective period *t*. Large values of this measure indicate big swings in price relative to FV and hence the possible presence of a bubble. *Total Dispersion* (Haruvy and Noussair, 2006) measures the aggregate absolute deviation of median price from FV across all trading periods, and is defined as $\sum_t |MedianP_t - F_t|$. Since it treats both positive and negative deviations from FV identically, it is a measure of aggregate mispricing rather than over or undervaluation, with smaller values indicating a closer correspondence between price and fundamental value. *Turnover,* a normalised measure of trading activity, is used as a measure of magnitude since bubble are typically associated with higher trading volumes. We calculate turnover as defined by King, Smith, Williams, and van Boening (1993), namely $\sum_t V_t/(TSU)$, where V_t, the volume of trade in period *t* is normalised by *TSU*, the total number of units of the asset (X or Y) in the market. *Normalised Deviation*, measured by Haruvy, Lahav, and Noussair (2007) as $\sum_t V_t |MedianP_t - F_t|/(TSU)$, combines the preceding two measures to account for both the size of the price deviation and the level of trading activity in a market. To examine how closely prices track changes in FV, we calculate *Haessel-*R^2 (Dufwenberg, Lindqvist, and Moore 2005), which is the R-squared from the regression of average prices on fundamental values. Being a goodness-of-fit measure, it conveys how much of the variation in average price across periods is explained by changes in FV; values closer to 0 (1) suggest the potential existence (absence) of price bubbles. Note that none of the aforementioned measures determine whether the asset is generally overvalued or undervalued. To gauge the degree of overpricing/underpricing, we calculate *Average Bias* (Haruvy and Noussair 2006), which measures how far median prices on average deviate from FV over the course of the market, and is calculated as $\frac{1}{N}\sum_{t=1}^{N}(MedianP_t - F_t)$. Large positive (negative) values suggest that prices tend to stay above (below) FV. Values close to zero may suggest that prices stay close to FV or that the asset experiences equal degrees of over and underpricing in the market; assessing the *Average Bias* in conjunction with *Total Dispersion* helps to shed light in this regard, since observing a small (large) *Total Dispersion* at the same time as a near-zero *Average Bias* would imply the former (latter) (Haruvy and Noussair 2006).

The first of the bubble-length measures, *Duration* (Porter and Smith 1995), calculates the maximum number of consecutive periods where average price increases relative to fundamental value, or $max\{m: \bar{P}_t - F_t < \bar{P}_{t+1} - F_{t+1} < \cdots < \bar{P}_{t+m} - F_{t+m}\}$. Larger values of *Duration* point to sustained periods where changes in (average) transaction price across trading periods do not 'adequately' track changes in the FV, potentially indicating the presence of a bubble. *Boom* (*Bust*) *Duration* (Haruvy and Noussair 2006) is defined as the maximum number of consecutive periods where median prices stay above (stay below) FV; large values indicate long periods of overvaluation (undervaluation), potentially signalling the presence (absence) of a bubble.

The Behaviour of Individual Assets

Panels A and B of Table 2 report the median values of the bubble measures in each treatment for assets X and Y respectively in Round 1, along with the associated median absolute deviations[31]. For each asset-type, each measure produces one observation per

[31] The median absolute deviation (MAD) is measure of spread of a distribution, and is calculated as the median of the absolute deviations of all values in a sample from the median. We report the median value and MAD of each measure in preference to the mean and standard deviation due to the small number of observations involved, and their lower sensitivity to outliers.

market; hence the medians are based on 7 observations in the *Baseline* treatment, and 8 observations each in the *Carrot* and *Stick* treatments[32]. The bottom half of each panel reports two-sided exact p-values from Wilcoxon Mann-Whitney U (WMW) tests of the differences in the measures between treatments, under the null that the values from both treatments come from the same distribution. The WMW test, which is a non-parametric test, is the appropriate statistical test given the small sample size.

We begin by comparing the two tournament treatments, *Carrot* and *Stick* (i.e. Hypothesis 3). In the case of the 'safe' asset, X, the differences between the *Carrot* and *Stick* treatments on most bubble measures are not statistically significant. Of the magnitude measures, only *Turnover* is marginally significantly lower in the *Stick* treatment compared to *Carrot* (p-value = 0.065), which lends some support to the notion that penalties embedded in tournament contracts inhibit speculation. However, the bubble-length measures present the opposite story, with significantly longer *Boom Durations* (p-value = 0.007) and significantly shorter *Bust Durations* (p-value = 0.039) in the *Stick* treatment indicative of more prolonged periods of overvaluation compared to the *Carrot* treatment; the median market in the *Stick* (*Carrot*) treatment experiences 10 (3.5) consecutive periods where the median price of X exceeds fundamental value, and only 1.5 (4.5) consecutive periods below fundamental value.

For the riskier asset Y, the degree of mispricing is comparable to asset X, as suggested by the similarity in the median bubble measure values between the two assets-types in each tournament treatment. Like asset X, *Turnover* for asset Y is significantly higher in the *Carrot* treatment (p-value = 0.038), while *Boom Duration* is again significantly longer in the *Stick* treatment (p-value = 0.022). In addition, prices for Asset Y are also significantly lower in the *Carrot* treatment according to the *Average Bias* measure (p-value = 0.028),

[32] The bubble measure values observed in the individual markets of each treatment are tabled in Appendix A. See Table A1 and A2 for the values from Round 1.

which shows that Asset Y is on average overvalued by 25 francs in each period under *Stick* incentives, whereas *Carrot* incentives are associated with asset Y being *undervalued* on average by almost 10 francs per period. This difference in *Average Bias* between the two tournament treatments is also mirrored in asset X, however the failure to attain statistical significance there is due to greater noise in the *Carrot* treatment.

Taken as a whole, the bubble measures are consistent with our observations from Figure 3. They reveal that when penalties are embedded into tournament contracts that reward participants for beating the 'market', inexperienced traders trade less compared to reward-only contracts. However, the trade that does occur actually happens at higher prices, and periods of overvaluation last longer, especially in the riskier asset. Thus, rather than curtailing the impetus to speculate on riskier ventures, our findings mostly suggest that the addition of a 'stick' achieves the opposite result. While we do not examine individual-level behaviour in this study, these results are consistent with pricing expected under the herding hypothesis (Rajan 2006; Dass et al 2008).

Like Kleinlercher et al (2014), our results appear to be driven by the attendant incentives and not by differences between the treatments in participants' inherent risk attitudes – average DOSPERT scores in the Financial domain (or its subsets), though collected after the market stage, do not differ significantly between *Carrot* and *Stick* markets (WMW test p-value (two-sided) = 0.529, $n_{Carrot} = n_{Stick} = 8$); nor are they driven by differences in cognitive ability, as measured by CRT scores (WMW test p-value (two-sided) = 0.6, $n_{Carrot} = n_{Stick} = 8$)[33]. However, while our findings here coincide with Kleinlercher et al regarding trading volumes, they contrast strongly with respect to the degree of mispricing

[33] Debate surrounding the collection of the DOSPERT/CRT data before or after the market is in some ways moot, since random assignment of participants to treatments should ensure that the treatment groups are on average 'equivalent' at the outset of the experiment. Nonetheless, we tested for differences as an additional safety measure. Though we do not report the full results here (available upon request), we do not find a significant difference between any of the treatments in the CRT scores (market average or individual) or DOSPERT scores (market average or individual).

observed in the riskier asset; overvaluation (calculated similarly to *Average Bias)* in their "high-risk" asset is greatest in their "Bonus" treatment and lowest in their "Penalty" treatment[34]. Since the absence of relative performance evaluation reduces the inclination to herd, a possible explanation for this stark difference between our studies lies in the non-competitive incentives faced by their traders. This, combined with the penalty-only nature/framing of their "Penalty" contract, may focus participants' thoughts on avoiding the uncertainty associated with the riskier asset, as there is no competition to beat or 'reward' to be gained. Moreover, while we do not speculate on the precise mechanism, the constant FV process used by Kleinlercher et al may also play a part.

Comparing the tournament treatments to the normal incentive *Baseline* treatments (Hypothesis 1) in Table 2 is also revealing. For both assets X and Y, we fail to find a significant difference between *Baseline* and the two tournament treatments on any of the bubble-magnitude measures. Of the bubble-length measures, *Boom Duration* is smaller in the *Baseline* treatment than the *Stick* treatment for asset X, but only marginally so (p-value = 0.081), while being significantly larger in the *Baseline* treatment compared to the *Carrot* treatment for asset Y (p-value = 0.013). This mostly runs contrary to much of the evidence from single-asset experimental studies going back to James and Isaac (2000) that find tournament incentives to be associated with significantly larger bubble. Hence, our results suggest that the findings of these earlier studies may be an artefact of speculation in a single-asset environment. When inexperienced traders are given the ability to bet on a higher payoff from an alternate, risky asset, tournament incentives do not distort prices any more than normal incentives.

< Insert Table 2 about here >

[34] For their "low-risk" asset (equivalent to asset X in our study), Kleinlercher et al (2014) report significantly lower (higher) average prices in their 'Penalty' ('Linear' incentives) treatment compared other treatments, although the differences are not economically significant – the price paths in all their treatments for the low-risk asset are very similar.

Relative Prices

Turning to relative prices, Panel A of Table 3 shows the median value of the *Average Prediction Error* in each treatment, along with the associated median absolute deviations. *Average Prediction Error* is identical to the 'overall normalised exchange rate deviation' measure used by Fisher and Kelly (2000), and is calculated by averaging the *Prediction Errors* (defined above) in all periods of a session/market. Like the bubble measures, this yields one observation per market. The table reports median *Average Prediction Errors* based on the entire duration of a market ("Avg PredErr"), as well as in each half of the market ("Avg PredErr_p1to6" and "Avg PredErr_p7to12).

Looking at the whole-of-market measure ("Avg PredErr"), we see that the median values are very similar – around 3% – in all three treatments. In fact, the median values are not significantly different from zero in any of the treatments, which is consistent with other multi-asset studies that find relative prices do not significantly deviate from the risk-neutral theoretical value when assets are differentiated by the variance of payoffs alone (Ackert et al 2006, Childs and Mestelman 2006)[35,36]. However, this contrasts again with Kleinlercher et al (2014), whose high-risk asset sells at a significant premium to the low-risk asset in their "Bonus" and the normal-incentive "Linear" treatments, while the opposite holds true in their "Penalty" treatment.

To examine if relative prices behave differently between the first and second half of the market – specifically due to heightened speculation on the risky asset in the second half – we compare the measures corresponding to each half *within* treatments. We use a Wilcoxon

---

[35] The statistical significance of the median *Average Prediction Error* in each treatment is assessed using the one-sample Wilcoxon Signed-rank test, under the null that the median is equal to zero. It is the non-parametric equivalent of the one-sample t-test.

[36] However, this result contrasts with Fisher and Kelly (2000) who report that their riskier asset sells at a slight premium to the safer asset, although they do not obtain enough observations to make a formal statistical comparison. In addition, their results are potentially confounded by the differing levels of experience of some traders in their markets. Ackert et al (2006) find a preference for assets with lottery-like payoffs only when trade occurs with borrowed money.

Signed-rank test to examine the one-sided alternative hypothesis that *Average Prediction Errors* in the second half of the market are higher than in the first half; the corresponding p-values (one-sided) are reported in the right-most column of Panel A[37]. The results do not provide compelling evidence of a 'tournament effect' in relative prices in Round 1. Although median *Average Prediction Errors* are larger in all treatments in the second half, these differences are only statistically significant in the normal-incentive *Baseline* treatment (median *Average Prediction Error* is -0.32% in 1st half vs. 7.88% in 2nd median), but only at the 10% level (p-value (one-sided) = 0.064).

The bottom of Panel A reports exact p-values (two-sided) from WMW tests comparing the measures between treatment-pairs. The null hypothesis is that values in both treatments come from the same distribution. The failure to achieve statistical significance on any of the tests means that we do not find support for the conjecture that tournament incentives have a significant impact on relative prices compared to normal incentives (Hypothesis 1), or that relative prices in *Carrot* markets behave differently to *Stick* markets (Hypothesis 3).

< Insert Table 3 about here >

4.2 The Effect of Experience

<u>Individual Assets</u>

Figure 5 depicts the evolution of median transaction prices for assets X and Y (panels (a) and (b) respectively) in the *Baseline, Carrot,* and *Stick* treatments during the second round of trading. With once-experienced traders, we see that the difference in price behaviour, in particular between the two tournament treatments is much less obvious compared to Round 1 (cf. Figure 3). The convergence between two is especially pronounced in the case of the risky

[37] The (paired-sample) Wilcoxon Signed-rank test is the non-parametric equivalent of the paired t-test. The null hypothesis is that values from both groups come from the same distribution.

asset Y, where in a reversal of Round 1, the *Stick* treatments appears to conform more closely to fundamental value than *Carrot.* And although median prices for asset X are again higher in the *Stick* treatment compared to the *Carrot* treatment in most trading periods, the differences do not appear to be as great as in Round 1, especially in the early and latter stages of the market. In fact, in contrast to Round 1, the differences in median price between the *Carrot* and *Stick* treatment are not statistically significant in any trading period, for both assets (unreported WMW tests). Furthermore, notwithstanding the considerable heterogeneity in price behaviour at the individual market level, it is the *Baseline* treatment where median prices exhibit the most obvious bubble – Asset Y – resulting in median prices for Y that are significantly higher than the *Carrot* treatment at the peak of the bubble in periods 7 and 8.[38]

< *Insert Figure 5 about here* >

The bubble measures corroborate these observations. Median values of the bubble measures in each treatment in Round 2, along with exact p-values (two-sided) from the associated WMW tests are detailed in Table 4[39]. For asset X (Panel A), relative median values on most bubble measures point to greater mispricing and overvaluation in the *Stick* treatment compared to the *Carrot* treatment, whereas the opposite holds true for asset Y (Panel B). However, on most measures, these differences in median values between the two treatments are smaller in Round 2 compared to Round 1. More importantly, in contrast to Round 1, we fail to reject the null of no difference between the *Carrot* and *Stick* treatments (Hypothesis 3) on any of the measures for either asset X or Y.

Comparing the *Baseline* treatment against the tournament treatments (Hypothesis 1), we do not find a significant difference in Round 2 between the *Baseline* treatment and the

[38] WMW test p-values are 0.045, and 0.048 in periods 7 and 8 respectively. The *Baseline* treatment also registers a significantly higher median price than *Carrot* in asset Y in period 3 (p-value = 0.049).
[39] Refer to Table A3 and A4 in Appendix A for the values of the bubble measures in each individual market of these treatments.

two tournament treatments in any of the bubble measures for asset X. For the riskier asset Y, all of the median bubble-measure values, with the exception of *Turnover*, indicate *greater* mispricing/bubble behaviour in the *Baseline* treatment than in either the *Carrot* or *Stick* treatments. Of these however, the only difference that is significant at the 5% level is in *Boom Duration*, which is smaller in the *Stick* treatment ($median_{Baseline}$ = 7 vs. $median_{Stick}$ = 3, p-value = 0.035). *Average Bias* is also higher in *Baseline* than in the *Carrot* treatment, but is only marginally significant ($median_{Baseline}$ = 24.33 vs. $median_{Carrot}$ = 1.08, p-value = 0.094).

< *Insert Table 4 about here* >

Hence, aggregate-level differences between the incentive schemes seem to largely dissipate when traders are experienced in relation to the experimental design (and their trading cohort). To help understand what drives this, Figures 6 and 7 compare, by treatment, the evolution of median prices in the two rounds for assets X and Y respectively. While it is difficult to make strong conclusions based on these figures, it is notable that the most striking change between rounds occurs in the *Stick* treatment for asset Y, where median prices are lower in most periods and adhere much more closely to FV in Round 2. Improved adherence to fundamental value in Round 2, particularly in the early stages of the market, is also evident for asset X in both the *Carrot* and *Stick* treatments. Median prices for asset X in the latter treatment also appear to adjust more quickly and successfully to fundamental value towards the end of the market. In contrast, median prices for asset Y in the *Baseline* and *Carrot* treatments seem to conform less well to FV with experienced traders, especially in the second half of the market.

< *Insert Figure 6 about here* >

< *Insert Figure 7 about here* >

We formally assess if price behaviour changes significantly between rounds within each treatment (i.e. Hypothesis 2) by conducting Wilcoxon Signed-rank tests on the various bubble measures; the null hypothesis is that there is no difference in the bubble measure between rounds. The two-sided p-values from these tests are shown in Table 5. Beginning with Panel A, which corresponds to the 'safe' asset X, we see that the *Baseline* treatment shows no significant change in behaviour between the two rounds on any of the measures. In contrast, the results point to an 'improvement' in price behaviour in the two tournament treatments, with a number of bubble measures indicative of significantly reduced mispricing/bubble behaviour. This is especially the case in the *Stick* treatment, where *Boom Duration* and *Turnover* are both significantly smaller in Round 2 (p-value = 0.014 and 0.03 respectively), *Haessel-*R^2 is significantly larger (p-value = 0.036), while *Normalised Deviation* and *Bust Duration* show improvements that are marginally significant (p-value = 0.093 and 0.078 respectively)[40]. These improvements help drive the trend to insignificance between the *Stick* and *Carrot* treatments in Round 2 for asset X. Even in the *Carrot* treatment, where median prices conform relatively well to FV in the first round of trading and hence the scope for 'improvement' is more limited, we see that *Turnover* and *Normalised Deviation* are both significantly lower in Round 2 for asset X (p-value = 0.012 for both measures). The significance of the latter measure appears to be driven by the decline in *Turnover* however, since *Total Dispersion* is not significantly different between the two rounds.

The moderating effect of experience on mispricing/bubbles under tournament incentives is also seen in the riskier asset, Y (Panel B). They confirm what Figure 7 strongly suggests – that the price behaviour of asset Y in the *Stick* treatment shows marked

[40] Even though the median value of *Normalised Deviation* for asset X in the *Stick* treatment is higher in Round 2 than in Round 1, the sign-rank test nonetheless reveals a marginally significant *improvement* because the Round 2 value of this measure is actually *lower* than the corresponding Round 1 value in 6 out of 8 markets.

improvement in its adherence to FV in Round 2. We see improvement in all of the bubble measures over the two rounds, significantly in the case of *Turnover* (p-value = 0.036) and the bubble-length measures, *Duration* (p-value = 0.031), *Boom Duration* (p-value = 0.011), and *Bust Duration* (p-value = 0.019). The adjustment is particularly large in the case of *Boom Duration*, where the median value falls from 8 to 3 periods. This primarily drives the shift to insignificance (significance) for the *Stick* treatment on this measure with respect to the *Carrot* (*Baseline*) treatment, where the same measure does not change significantly between rounds.

Despite the impression created by Figure 7 that prices for asset Y in the *Carrot* treatment are distorted more by experience, we do not find any evidence supporting this in the bubble measures. In fact, the median values of most bubble measures suggest *less* distortion in Round 2, significantly in the case of *Duration* (p-value = 0.013) and *Turnover* (p-value = 0.013), and marginally significantly for *Total Dispersion* (p-value = 0.093). The composite measure of *Turnover* and *Total Dispersion*, *Normalised Deviation*, is also significantly smaller with experienced traders (p-value = 0.012).

Once again, the effect of experience is least pronounced in the *Baseline* treatment. Indeed, consistent with Figure 7, most bubble measures for asset Y in this treatment actually 'worsen' in Round 2, although the deterioration is only significant in one measure – *Total Dispersion* – and that too only marginally so (p-value = 0.091). The only measure to achieve significance at the 5% level is *Turnover*, which like the two tournament treatments, is actually significantly reduced by experience (p-value = 0.018).

< Insert Table 5 about here >

Thus, the trend towards convergence in price behaviour between the treatments as participants gain experience is primarily driven by the reduction in mispricing/bubbles in the two tournament treatments, especially in the *Stick* treatment. This result contrasts strongly with James and Isaac (2000) and Cheung and Coleman (2014), who find that prices under tournament incentives diverge more from fundamental value (i.e. bubbles become larger) as traders gain experience. The most likely explanation for this discrepancy lies in a crucial difference between these earlier studies and ours – they examine single-asset environments, whereas participants in our markets trade two differentiated risky assets. Hence, our result suggests that the number of assets available for trade also plays an important role in determining how trading experience interacts with tournament incentives in affecting prices.

Relative Prices

Having examined how experience affects the prices of individual assets, we now turn to its impact on relative prices. Figure 8 shows the median *Prediction Error* in each period for each treatment in Round 2. The most interesting aspect of this chart and the most obvious change from Round 1 (cf. Fig. 4) is that the *Carrot* treatment exhibits a pronounced upward trajectory in the second half of the market, particularly in the last 3 periods. Participants in the median *Carrot* market were willing to pay a 63% premium to acquire asset Y relative the price paid for X in the final period. Furthermore, the average of the median *Prediction Errors* in the *Carrot* treatment in the first 6 periods is -12.2% compared to 16.4% in the final six. This is consistent with tournament-induced risk-seeking by traders who are hoping to improve their rankings as the end of the market approaches by betting on receiving the relatively large dividend that asset Y provides. In contrast, relative prices do not seem to behave in an overtly similar manner in the *Baseline* and *Stick* treatments in Figure 8, with median *Prediction Errors* for both treatments staying in the region of zero in all periods.

< Insert Figure 8 about here >

Panel B of Table 3, which reports the results of statistical tests on the *Average Prediction Errors* in Round 2, confirms the changing behaviour of relative prices within the

Carrot treatment – the median *Average Prediction Error* in the second half of the market (15.04%) is significantly larger (one-sided p-value = 0.013) than the corresponding value for the first half of the market (-12.70%). The fact that we also observe a similar effect in the *Baseline* treatment (one-sided p-value = 0.032), albeit one that is smaller – median *Average Prediction Error* in periods 1-6 is -0.12% vs. 5.23% in periods 7-12 – indicates that this not a purely 'tournament' phenomenon. However, the larger magnitude of the difference in the *Carrot* treatment suggests that competitive incentives may play an amplifying role. As for why such an effect appears with experienced traders even though it is missing in Round 1, we posit that a possible explanation could be that traders become more aware of the strategic use of the riskier asset Y as they become more familiar with the trading environment and dividend structures of the two assets[41].

Unlike the two other treatments, the *Stick* treatment does not show a significant difference in relative price behaviour between the two halves of the market. Moreover, when relative prices are examined over course of the entire market, the whole-of-market *Average Prediction Error* ('Avg PredErr') in all three treatments is not significantly different from zero (i.e. relative prices conform to the theoretical value 'on average'). For the *Carrot* treatment, this result arises because the statistically significant and large relative discount for asset Y in the first half of the market (one-sample Wilcoxon Signed-rank p-value = 0.0499) is offset by the (marginally) significant and large premium in the second (one-sample Wilcoxon Signed-rank p-value = 0.093). Furthermore, we do not detect a significant difference between any of the treatments on the whole-of-market measure, and we also fail to reject the null of no difference between the treatments in the measures corresponding to each half of the market. Hence, like Round 1, relative prices in our treatments do not appear to behave significantly differently from each other.

[41] Of course, it is possible that there *is* a similar difference in the behaviour of relative prices in Round 1, but our test lacks the power to detect it.

4.3 Gilpatric Contracts

Individual Assets

Panels A and B of Figure 9 (10) compare the price behaviour of asset X (Y) in the *Carrot* and *Stick* treatments against their respective rank-order tournament equivalents, *GilCarrot* and *GilStick* in Round 1. These figures reveal that median prices in the rank-order tournaments and their James & Isaac tournament counterparts are generally closely associated, especially in the case of the penalty-based contracts (*Stick* and *GilStick*). However, the relationship does not appear to be as close between the reward-only treatments, where *GilCarrot* produces higher median prices than the *Carrot* treatment in most periods. Except for the first period, these differences are generally small or negligible in asset Y but are larger and more persistent in asset X[42]. As a consequence, the noticeable difference in median prices that exists between the *Carrot* and *Stick* treatments in inexperienced markets (see Fig. 3) is greatly diminished in the case of *GilCarrot* and *GilStick*, as shown in panels A and B of Figure 11.

< *Insert Figure 9 about here* >

< *Insert Figure 10 about here* >

< *Insert Figure 11 about here* >

[42] Median prices for asset X in the *GilCarrot* treatment are significantly higher than *Carrot* in periods 9 and 12, but only at the 10% level (WMW p-value = 0.07 and 0.092 respectively). Unreported WMW tests comparing bubble measures between the *Carrot* and *GilCarrot* treatments show that *GilCarrot* has a longer *Boom Duration* that is marginally significant (p-value = 0.064) for asset X; all other measures for asset X return insignificant differences, while there are no significant differences between *Carrot* and *GilCarrot* on any measures for asset Y. Similarly, *Stick* and *GilStick* do not show significant differences on any of the bubble measures for either asset, except the trading activity measure *Turnover*, which is marginally significantly higher in the *GilStick* treatment for asset X (p-value = 0.053).

The bubble measures from the rank-order tournaments in Round 1 are summarised in Table 6, along with exact p-values (two-sided) from the corresponding WMW tests[43]. Consistent with the visual data, and in contrast to the James & Isaac tournament contracts (cf. Table 2), we do not detect, for either asset-type, a significant difference between *GilCarrot* and *GilStick* on any of the bubble measures (Hypothesis 3). Relative to the James & Isaac tournaments, Table 6 is also somewhat more supportive of the argument that tournament contracts distort prices more than normal incentives (Hypothesis 1) – *Haessel-*R^2 is higher, and *Turnover* and *Normalised Deviation* are both significantly lower in the *Baseline* treatment than in the *GilCarrot* treatment, albeit only marginally (p-values of 0.051, 0.073, and 0.073 respectively), and only for asset X. Given that we fail to find a significant difference in *Total Dispersion* between *Baseline* and *GilCarrot*, it is also likely that the difference in *Normalised Deviation* simply reflects the same effect as *Turnover*.

< Insert Table 6 about here >

In markets with experienced traders (Round 2), we see in Figure 12 that differences in median price between *GilCarrot* and *GilStick* appear to be greater for asset Y than X. Indeed, median prices in the *GilStick* treatment exhibit a sizeable bubble in asset Y. However as Table 7 reveals, like Round 1, none of the bubble measures for asset Y in Round 2 differ significantly between the two rank-order tournament treatments. Also, with the exception of *Amplitude*, which is marginally significantly greater in the *GilCarrot* treatment than in *GilStick* (p-value = 0.065), all other bubble measures for asset X return insignificant differences. In regards to the *Baseline* treatment, we fail to find a significant difference on any of the bubble measures with respect to the *GilCarrot* or *GilStick* treatments for either asset.

[43] Refer to Table A1 and A2 in Appendix A for the values of the bubble measures in the individual markets of these treatments. Round 2 equivalents can be found in Tables A3 and A4.

< *Insert Figure 12 about here* >

< *Insert Table 7 about here* >

As with the James & Issac tournaments, we find that greater trading experience is associated with smaller bubbles under rank-order tournament conditions, especially in the *GilStick* treatment (Hypothesis 2). The results (two-sided p-values) of Signed-rank tests carried out on the bubble measures of each rank-order tournament treatment are shown in Table 8. The table reveals that for asset X (Panel A) in the *GilStick* treatment, *Amplitude* and *Total Dispersion* are significantly smaller in Round 2 (p-value = 0.028 and 0.046 respectively), while *Normalised Deviation*, *Duration*, and *Boom Duration* are also smaller but only at the 10% level (p-value = 0.075, 0.091, and 0.058 respectively). For asset Y (Panel B) in the *GilStick* treatment, *Turnover* is significantly reduced by experience (p-value = 0.046), while smaller *Amplitudes* and *Durations* in Round 2 are marginally significant (p-value = 0.075 and 0.058 respectively). On the other hand, the evidence that experience reduces mispricing/bubbles is weaker in the *GilCarrot* treatment, where the only measure that changes significantly between rounds is the trading activity measure, *Turnover*, which is smaller in Round 2 for both assets (p-value = 0.028 in both cases). Importantly however, none of the bubble measures in either rank-order tournament treatment indicate significantly *more* mispricing as participants gain experience.

< *Insert Table 8 about here* >

<u>Relative Prices</u>

Relative prices in the Gilpatric tournaments in Round 1 behave in a qualitatively similar manner to the James and Isaac tournaments. That is, on ‘average’, relative prices conform to the theoretical value and do not differ significantly between the two treatments, *GilCarrot* and *GilStick*. This can be seen in Figure 13(a), where the median *Prediction Error* in both treatments is close to zero in most periods. Consistent with this, in both treatments we fail to reject the null hypothesis that the median *Average Prediction Error*, summarised in Panel A of Table 9, is equal to zero. This is also the case for the two half-market measures. Moreover, within each rank-order tournament treatment, we do not detect a significant difference in relative-price behaviour between the first and second halves of the market. Furthermore, the differences between *GilCarrot* and *GilStick,* assessed using the WMW test, are not statistically significant on any of the measures, nor do the rank-order tournaments differ significantly from the *Baseline* condition.

The results in Round 2 are similar to Round 1. In both rank-order tournaments, median *Prediction Errors* stay in the region of zero, as shown in Figure 13(b), while *Average Prediction Errors* (for the whole market and in each half) are generally not significantly different from zero (see Panel B of Table 9). The only exception to this is the *GilCarrot* treatment, in which according to the *Average Prediction Error* measure, asset Y sells at a statistically significant discount to asset X of around 4% in the first half of the median market, but only at the 10% level (one-sample Signed-rank test p-value = 0.075). In addition, while *Average Prediction Error* is higher in the second half of the market in the *GilStick* treatment compared to the first half, the statistical significance of the difference is only marginal (one-sided p-value = 0.058) and the economic significance even less so (3.46% in the first half vs. 3.85% in the second). Notably, relative prices in the *GilCarrot* treatment display none of the signs of heightened speculation in asset Y that is evident in the *Carrot* treatment in Round 2 (cf. Fig. 8). Furthermore, comparing the *GilCarrot* and *GilStick* treatments, we see in Panel B of Table 9 that differences in *Average Prediction Error* between the two treatments generally fail to attain statistical significance. The first half of the market again presents the exception; there is some evidence that the relative price of asset Y is higher in *GilStick* than in *GilCarrot* during this period, although significance here is only at the 10% level (p-value = 0.093) and the difference is economically quite small (-4.14% in *GilCarrot* vs. 3.46% in *GilStick)*.

< Insert Figure 13 about here >

< Insert Table 9 about here >

5. Conclusion

Tournament incentives have been accused in the experimental literature of distorting the efficient functioning of markets by exacerbating asset price bubbles. While this narrative tallies with mooted concerns regarding the link between market instability and the proliferation of convex incentive structures in the financial industry, the real-world relevance of these experimental results is limited by their examination of single-asset markets that preclude the ability to trade in securities with dissimilar risk characteristics, an option that is available to investors in real markets. Moreover, the reward-centric focus of existing studies means the role of penalties for poor performance in tournament contracts have been largely ignored, despite the fact that they may help to moderate risk-taking behaviour. We address these gaps in the literature by examining how rewards ('carrots') and penalties ('sticks') embedded in tournament contracts affect price behaviour in experimental asset markets where participants can trade in two differentiated assets. Each asset has the same risk-neutral fundamental value, but one asset is intrinsically riskier by virtue of a lottery-like dividend structure that generates potentially higher payoffs, thus allowing traders to naturally vary their risk by shifting in/out of the asset.

Our results challenge the main conclusions of the existing literature. In two-asset experimental markets, we do not find any compelling evidence to suggest that asset price bubbles are larger under tournament incentives than normal, absolute-performance based incentives. Moreover, unlike earlier studies, bubbles under tournament incentives in our markets do dissipate with experienced traders. Hence, the results of earlier studies appear to be driven by the single-asset nature of their markets.

Furthermore, penalties embedded into tournament contracts that reward traders for ‘beating the market’ reduce the volume of trading activity in inexperienced markets compared to reward-only contracts. However, the trade that does occur happens at higher prices, and periods of overvaluation last longer, especially in the case of the riskier asset. Thus, in markets with inexperienced traders, ‘sticks’ or penalties for underperformance are associated with *greater* mispricing, not less. While this may seem a counterintuitive and surprising result, it is consistent with price behaviour under tournament incentives when traders are prone to herd; the inclusion of a penalty for underperformance makes traders more likely to herd as a way to minimise the risk of being an underperformer, thus perversely exacerbating and prolonging mispricing. However, this effect does not appear to survive in our markets when participants are once-experienced. Moreover, we do not observe a significant difference in price behaviour between carrot-only and carrot-and-stick contracts when we implement a rank-order tournament, either with inexperienced or experienced traders.

In light of the on-going debate surrounding compensation practices in the financial industry, our results are particularly relevant to policymakers and regulators. Our findings suggest that, at the aggregate-level, tournament incentives may not be as disruptive a force as earlier studies indicate. Furthermore, regulatory initiatives such as placing caps on finance professionals’ bonuses may be misplaced – shifting the balance between carrots and sticks further towards the stick-end may reduce (increase) the incentive to (herd) trade against the herd, thereby having the perverse effect of fuelling the instability that such actions seek to prevent.

Although we make important contributions towards better understanding the aggregate effects of tournament incentives, the laboratory environment in which we conduct our study is obviously considerably less complex than real markets and the real world. As such, our study is subject to the limitations of experimentation as a methodology. Foremost amongst these is the 'penalty' in our tournament contracts – a zero payment, which some may reasonably protest is not a 'real' penalty since it does not impose actual losses on traders. Whilst true, a zero payment represents the most an experimenter can penalise experiment participants, given that ethical considerations preclude experimenters from enforcing financial losses/liabilities on subjects. Even if it were possible, potential selection biases make it undesirable, since only certain types of subjects may volunteer for the experiment. Moreover, the gap between our non-penalty payment and the zero payment still represents a sizeable disincentive for university student participants, given the time commitment made. (2.5 hours). Hence, while the impact of 'sticks' may be diminished in an experimental setting compared to the real world – where professionals face the more sever risks of job termination and/or reputational damage – this suggests the differences that we do observe with inexperienced participants are likely to be *underestimated*.

The fact that we consider markets with only two risky assets, whereas real-world markets are characterised by a myriad of potential investments, may be considered another limitation of our study. Furthermore, unlike our treatment groups, market participants in the real world do not all trade under the same incentives. These represent simplifications of the real world required to build a workable experimental design and isolate the effects of different incentive schemes. Thus the extents to which our results can be generalised when these restrictions are relaxed is an open question, and as such, represent potential avenues for future research.

References:

Acker, D., Duck, N.W., 2006. A Tournament Model of Fund Management. *Journal of Business Finance & Accounting* 33, 1460-1483

Ackert, L.F., Charupat, N., Church, B.K., Deaves, R., 2006. Margin, Short Selling, and Lotteries in Experimental Asset Markets. *Southern Economic Journal* 73, 419-436

Ackert, L.F., Church, B.K., 2001. The Effects of Subject Pool and Design Experience on Rationality in Experimental Asset Markets. *Journal of Psychology and Financial Markets* 2, 6 - 28

Ang, J., Diavatopoulos, D., Schwarz, T., 2010. The Creation and Control of Speculative Bubbles in a Laboratory Setting. In: Lee C-F, Lee A & Lee J (eds.) *Handbook of Quantitative Finance and Risk Management*. Springer US, pp. 137-164.

Bagnoli, M., Watts, S.G., 2000. Chasing Hot Funds: The Effects of Relative Performance on Portfolio Choice. *Financial Management* 29, 31-50

Bebchuk, L.A., Spamann, H., 2009. Regulating bankers' pay. *geo. Lj* 98, 247

Becker, B.E., Huselid, M.A., 1992. The Incentive Effects of Tournament Compensation Systems. *Administrative Science Quarterly* 37, 336-350

Blais, A.-R., Weber, E.U., 2006. A Domain-Specific Risk-Taking (DOSPERT) scale for adult populations. *Judgment and Decision Making* 1, 33

Blinder, A.S., 2009. 'Crazy Compensation and the Crisis', *The Wall Street Journal*, 28 May, accessed 16 January 2015, < http://www.wsj.com/articles/SB124346974150760597>

Bothner, M.S., Kang, J.-h., Stuart, T.E., 2007. Competitive Crowding and Risk Taking in a Tournament: Evidence from NASCAR Racing. *Administrative Science Quarterly* 52, 208-247

Brown, J., Li, J., 2010. Going for it: the adoption of risky strategies in tournaments. Working paper

Brown, K.C., Harlow, W.V., Starks, L.T., 1996. Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry. *The Journal of Finance* 51, 85-110

Bull, C., Schotter, A., Weigelt, K., 1987. Tournaments and Piece Rates: An Experimental Study. *Journal of Political Economy* 95, 1-33

Busse, J.A., 2001. Another Look at Mutual Fund Tournaments. *The Journal of Financial and Quantitative Analysis* 36, 53-73

Cabral, L.M.B., 2003. R&D Competition when firms Choose Variance. *Journal of Economics & Management Strategy* 12, 139-150

Caginalp, G., Porter, D., Smith, V., 1998. Initial cash/asset ratio and asset prices: An experimental study. *Proceedings of the National Academy of Sciences of the United States of America* 95, 756-761

Caginalp, G., Porter, D., Smith, V., 2001. Financial bubbles: Excess cash, momentum, and incomplete information. *The Journal of Psychology and Financial Markets* 2, 80-99

Caginalp, G., Porter, D., Smith, V.L., 2000. Overreactions, Momentum, Liquidity, and Price Bubbles in Laboratory and Field Asset Markets. *Journal of Psychology & Financial Markets* 1, 24-48

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

Chen, H.-l., Pennacchi, G.G., 2009. Does Prior Performance Affect a Mutual Fund's Choice of Risk? Theory and Further Empirical Evidence. *The Journal of Financial and Quantitative Analysis* 44, 745-775

Cheung, Stephen L., and Andrew Coleman, 2014, Relative performance incentives and price bubbles in experimental asset markets, *Southern Economic Journal* 81, 345-363.

Chevalier, J., Ellison, G., 1997. Risk Taking by Mutual Funds as a Response to Incentives. *Journal of Political Economy* 105, 1167-1200

Childs, J., Mestelman, S., 2006. Rate-of-return Parity in Experimental Asset Markets*. *Review of International Economics* 14, 331-347

Dass, N., Massa, M., Patgiri, R., 2008. Mutual Funds and Bubbles: The Surprising Role of Contractual Incentives. *Review of Financial Studies* 21, 51-99

Dufwenberg, M., Lindqvist, T., Moore, E., 2005. Bubbles and Experience: An Experiment. *The American Economic Review* 95, 1731-1737

Ehrenberg, R.G., Bognanno, M.L., 1990. Do Tournaments Have Incentive Effects? *Journal of Political Economy* 98, 1307-1324

Elton, E.J., Gruber, M.J., Blake, C.R., 2003. Incentive Fees and Mutual Funds. *The Journal of Finance* 58, 779-804

Elton, E.J., Gruber, M.J., Blake, C.R., Krasny, Y., Ozelge, S.O., 2010. The effect of holdings data frequency on conclusions about mutual fund behavior. *Journal of Banking & Finance* 34, 912-922

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

Fisher, E.O.N., Kelly, F.S., 2000. Experimental Foreign Exchange Markets. *Pacific Economic Review* 5, 365-387

Frederick, S., 2005. Cognitive Reflection and Decision Making. *The Journal of Economic Perspectives* 19, 25-42

Gaba, A., Kalra, A., 1999. Risk Behavior in Response to Quotas and Contests. *Marketing Science* 18, 417-434

Gaba, A., Tsetlin, I., Winkler, R.L., 2004. Modifying Variability and Correlations in Winner-Take-All Contests. *Operations Research* 52, 384-395

Genakos, C., Pagliero, M., 2012. Interim Rank, Risk Taking, and Performance in Dynamic Tournaments. *Journal of Political Economy* 120, 782-813

Gilpatric, S.M., 2009. Risk taking in contests and the role of carrots and sticks. *Economic Inquiry* 47, 266-277

Goriaev, A., Nijman, T.E., Werker, B.J.M., 2005. Yet another look at mutual fund tournaments. *Journal of Empirical Finance* 12, 127-137

Green, J.R., Stokey, N.L., 1983. A Comparison of Tournaments and Contracts. *Journal of Political Economy* 91, 349-364

Greiner, B., 2004. The Online Recruitment System ORSEE 2.0 - A Guide for the Organization of Experiments in Economics. University of Cologne, Department of Economics

Grund, C., Höcker, J.A.N., Zimmermann, S., 2013. Incidence and consequences of risk-taking behavior in tournaments — evidence from the NBA. *Economic Inquiry* 51, 1489-1501

Haruvy, E., Lahav, Y., Noussair, C.N., 2007. Traders' Expectations in Asset Markets: Experimental Evidence. *The American Economic Review* 97, 1901-1920

Haruvy, E., Noussair, C.N., 2006. The Effect of Short Selling on Bubbles and Crashes in Experimental Spot Asset Markets. *The Journal of Finance* 61, 1119-1157

Hu, P., Kale, J.R., Pagani, M., Subramanian, A., 2011. Fund Flows, Performance, Managerial Career Concerns, and Risk Taking. *Management Science* 57, 628-646

Hvide, Hans K., 2002. Tournament Rewards and Risk Taking. *Journal of Labor Economics* 20, 877-898

Isaac, R.M., James, D., 2003. Boundaries of the Tournament Pricing Effect in Asset Markets: Evidence from Experimental Markets. *Southern Economic Journal* 69, 936-951

James, D., Isaac, R.M., 2000. Asset Markets: How They Are Affected by Tournament Incentives for Individuals. *The American Economic Review* 90, 995-1004

Kempf, A., Ruenzi, S., 2008. Tournaments in Mutual-Fund Families. *Review of Financial Studies* 21, 1013-1036

Kempf, A., Ruenzi, S., Thiele, T., 2009. Employment risk, compensation incentives, and managerial risk taking: Evidence from the mutual fund industry. *Journal of Financial Economics* 92, 92-108

King, R., Smith, V.L., Williams, A.W., Boening, M.V., 1993. The robustness of bubbles and crashes in experimental stock markets. In: Day R & Chen P (eds.) *Non Linear Dynamics and Evolutionary Economics*. Oxford Press, Oxford, pp. 183-2999.

Kleinlercher, D., Huber, J., Kirchler, M., 2014. The impact of different incentive schemes on asset prices. *European Economic Review* 68, 137-150

Koski, J.L., Pontiff, J., 1999. How Are Derivatives Used? Evidence from the Mutual Fund Industry. *The Journal of Finance* 54, 791-816

Kräkel, M., Sliwka, D., 2004. Risk Taking in Asymmetric Tournaments. *German Economic Review* 5, 103-116

Lazear, E.P., Rosen, S., 1981. Rank-Order Tournaments as Optimum Labor Contracts. *Journal of Political Economy* 89, 841-864

Lee, J., 2004. Prize and risk-taking strategy in tournaments: evidence from professional poker players. Working paper.

Lugovskyy, V., Puzzello, D., Tucker, S., 2009. An Experimental Study of Bubble Formation in Asset Markets Using the Tâtonnement Pricing Mechanism. Working paper. University of Canterbury, Department of Economics and Finance

Nalebuff, B.J., Stiglitz, J.E., 1983. Prizes and Incentives: Towards a General Theory of Compensation and Competition. *The Bell Journal of Economics* 14, 21-43

Nieken, P., Sliwka, D., 2010. Risk-taking tournaments – Theory and experimental evidence. *Journal of Economic Psychology* 31, 254-268

Noussair, C., Robin, S., Ruffieux, B., 2001. Price Bubbles in Laboratory Asset Markets with Constant Fundamental Values. *Experimental Economics* 4, 87-105

Noussair, C.N., Powell, O., 2010. Peaks and valleys: Price discovery in experimental asset markets with non-monotonic fundamentals. *Journal of Economic Studies* 37, 152-180

Oechssler, J., Roider, A., Schmitz, P.W., 2009. Cognitive abilities and behavioral biases. *Journal of Economic Behavior & Organization* 72, 147-152

O'Keeffe, M., Viscusi, W.K., Zeckhauser, R.J., 1984. Economic Contests: Comparative Reward Schemes. *Journal of Labor Economics* 2, 27-56

Palan, Stefan, 2013, A review of bubbles and crashes in experimental asset markets, *Journal of Economic Surveys* 27, 570-588.

Porter, D.P., Smith, V.L., 1995. Futures contracting and dividend uncertainty in experimental asset markets. *Journal of Business* 68, 509

Qiu, J., 2003. Termination Risk, Multiple Managers and Mutual Fund Tournaments. *European Finance Review* 7, 161-190

Rajan, R.G., 2006. Has Finance Made the World Riskier? *European Financial Management* 12, 499-533

Rajan, R.G., 2008. 'Bankers' Pay is Deeply Flawed', *Financial Times*, 9 January, accessed 16 January 2015, < http://www.ft.com/intl/cms/s/0/18895dea-be06-11dc-8bc9-0000779fd2ac.html#axzz3T9uTJuUR>

Robin, S., Straznicka, K., Villeval, M.-C., 2012. Bubbles and incentives: an experiment on asset markets. GATE-Groupe d'Analyse et de Théorie Économique Lyon-St Étienne Working Paper

Rosen, S., 1986. Prizes and Incentives in Elimination Tournaments. *The American Economic Review* 76, 701-715
Schoenberg, E.J., Haruvy, E., 2012. Relative performance information in asset markets: An experimental approach. *Journal of Economic Psychology* 33, 1143-1155
Sirri, E.R., Tufano, P., 1998. Costly Search and Mutual Fund Flows. *The Journal of Finance* 53, 1589-1622
Smith, V.L., Suchanek, G.L., Williams, A.W., 1988. Bubbles, Crashes, and Endogenous Expectations in Experimental Spot Asset Markets. *Econometrica* 56, 1119-1151
Smith, V.L., van Boening, M., Wellford, C.P., 2000. Dividend Timing and Behavior in Laboratory Asset Markets. *Economic Theory* 16, 567-583
Taylor, J., 2003. Risk-taking behavior in mutual fund tournaments. *Journal of Economic Behavior & Organization* 50, 373-383
Tsetlin, I., Gaba, A., Winkler, R., 2004. Strategic Choice of Variability in Multiround Contests and Contests with Handicaps. *Journal of Risk and Uncertainty* 29, 143-158
Wagner, W., 2013. Performance Evaluation and Financial Market Runs. *Review of Finance* 17, 597-624
Weber, E.U., Blais, A.-R., Betz, N.E., 2002. A domain-specific risk-attitude scale: measuring risk perceptions and risk behaviors. *Journal of Behavioral Decision Making* 15, 263-290

Table 1: General Demographic Information

This table reports general demographic information on the subject pool, categorised by the experimental treatment to which participants were randomly assigned. 'Business student' is defined as someone studying (self-reported) Finance, Economics, Actuarial, Accounting, or "Commerce". In a post-experiment survey, all participants completed the Cognitive Reflection Test (CRT) developed by Frederick (2005), which measures cognitive ability; CRT scores are out of 3 and higher scores indicate better performance. Participants also completed the Domain-Specific Risk-Taking (DOSPERT) Scale (Blais and Weber 2006). The score reported here relates to participants' (self-reported) likelihood of engaging in risky financial activities. Scores range from 6 to 42, with higher scores indicating a greater likelihood of engaging in risky activities.

	Baseline	*Carrot*	*Stick*	*GilCarrot*	*GilStick*
No. markets	7	8	8	6	6
No. subjects	51	58	61	45	46
Average age	22.3	22.4	22.2	22.7	22.6
Male (%)	65	52	52	42	46
Business students (%)	29	40	31	36	43
Avg. CRT score	1.6	1.3	1.5	1.5	1.3
Avg. DOSPERT Fin. score	19.2	19.4	18.5	19.8	18.8

Table 2: Summary of bubble measures for assets X and Y in Round 1

This table reports median values of each bubble measure in the *Baseline*, *Carrot*, and *Stick* treatments during Round 1 of the market (i.e. with inexperienced traders); the associated median absolute deviations are displayed in parentheses. Markets contaminated by subjects who had participated in an earlier session are excluded. Panel A (B) reports bubble measure data relating to Asset X (Y). For definitions of the relevant bubble measures, refer to section 4.1.2. The statistical significance of the difference between treatments in each measure is assessed using a two-sided Wilcoxon Mann-Whitney U Test, under the null that values from both treatments come from the same distribution. Exact p-values are reported. Differences that are significant at the 10%, 5% and 1% level are denoted by *, **, and *** respectively.

Panel A: Asset X, Round 1:

Treatment [N]	*Amplitude*	*Total Dispersion*	*Average Bias*	*Haessel R^2*	*Turnover*	*Normalised Deviation*	*Duration*	*Boom Duration*	*Bust Duration*
Baseline [7]	3.31	298.00	1.65	0.78	2.45	83.43	5.00	5.00	4.00
	(1.53)	*(181.50)*	*(15.73)*	*(0.16)*	*(0.72)*	*(45.08)*	*(2.00)*	*(2.00)*	*(1.00)*
Carrot [8]	1.51	569.50	-9.42	0.56	2.93	171.09	5.00	3.50	4.50
	(1.06)	*(330.25)*	*(28.77)*	*(0.36)*	*(0.66)*	*(104.79)*	*(2.50)*	*(2.00)*	*(1.50)*
Stick [8]	2.63	607.00	25.63	0.46	2.16	98.89	4.50	10.00	1.50
	(1.55)	*(267.75)*	*(25.56)*	*(0.35)*	*(0.34)*	*(47.69)*	*(1.00)*	*(1.50)*	*(0.50)*
WMW U-Test p-values (two-sided):									
Baseline vs Carrot	0.867	0.779	0.536	0.536	0.281	0.397	0.799	0.317	0.290
Baseline vs Stick	0.694	0.281	0.232	0.121	0.779	0.613	0.465	0.081*	0.400
Carrot vs Stick	0.382	0.878	0.105	0.328	0.065*	0.878	0.576	0.007***	0.039**

Panel B: Asset Y, Round 1

Treatment [N]	*Amplitude*	*Total Dispersion*	*Average Bias*	*Haessel R^2*	*Turnover*	*Normalised Deviation*	*Duration*	*Boom Duration*	*Bust Duration*
Baseline [7]	1.63	530.50	15.38	0.77	2.03	99.06	4.00	7.00	4.00
	(1.01)	*(150.50)*	*(25.33)*	*(0.05)*	*(0.35)*	*(54.11)*	*(1.00)*	*(1.00)*	*(2.00)*
Carrot [8]	1.76	584.75	-9.96	0.32	2.85	194.11	6.00	4.00	4.00
	(1.20)	*(299.50)*	*(13.67)*	*(0.27)*	*(0.39)*	*(71.02)*	*(1.00)*	*(2.00)*	*(1.00)*
Stick [8]	1.99	529.25	24.85	0.56	1.89	96.18	4.00	8.00	3.00
	(0.96)	*(270.00)*	*(27.04)*	*(0.30)*	*(0.38)*	*(39.99)*	*(1.00)*	*(2.50)*	*(1.00)*
WMW U-Test p-values (two-sided):									
Baseline vs Carrot	0.955	0.955	0.232	0.463	0.281	0.694	0.421	0.013**	0.405
Baseline vs Stick	0.779	0.955	0.336	0.613	0.613	0.867	1.000	0.755	0.669
Carrot vs Stick	0.721	0.798	0.028**	0.721	0.038**	0.279	0.329	0.022**	0.124

Table 3: Average Prediction Errors

Median values of the *Average Prediction Error* in the *Baseline, Carrot, and Stick* treatments in Round 1 and 2 are shown below in Panels A and B respectively, with the associated median absolute deviations in parentheses. Markets contaminated by subjects who had participated in an earlier session are excluded. *Average Prediction Error* is calculated using all periods in a market, the first 6 periods, and the final 6 periods in *AvgPredErr*, *AvgPredErr_p1to6*, and *AvgPredErr_p7to12* respectively. The statistical significance of the individual measures is assessed using a (two-sided) one-sample Wilcoxon Signed-rank test, under the null that the median is equal to zero. The statistical significance of the difference between treatments is assessed using the Wilcoxon Mann-Whitney U Test under the null that values from both treatments come from the same distribution. The statistical significance of the difference between *AvgPredErr_p1to6* and *AvgPredErr_p7to12* within each treatment is assessed using a (paired-sample) Wilcoxon Signed-rank test, against the one-sided alternative hypothesis that *AvgPredErr_p7to12* > *AvgPredErr_p1to6*. Differences that are significant at the 10%, 5% and 1% level are denoted by *, **, and *** respectively.

Panel A: Round 1

Treatment [N]	*Avg PredErr (%)*	*AvgPredErr _p1to6 (%)*	*AvgPredErr _p7to12 (%)*	*Signed-rank p-value (1-sided) 1-6 vs 7-12*
Baseline [7]	3.78	-0.32	7.88	0.064*
	(9.46)	*(3.36)*	*(9.26)*	
Carrot [8]	3.17	-1.04	4.56	0.242
	(4.43)	*(7.20)*	*(8.01)*	
Stick [8]	3.23	0.14	4.03	0.200
	(4.08)	*(6.85)*	*(1.37)*	
WMW U-Test p-values (2-sided):				
Baseline vs Carrot	0.694	0.867	0.463	
Baseline vs Stick	0.955	0.955	0.463	
Carrot vs Stick	0.878	1.000	0.959	

Panel B: Round 2

Treatment [N]	*Avg PredErr (%)*	*AvgPredErr _p1to6 (%)*	*AvgPredErr _p7to12 (%)*	*Signed-rank p-value (1-sided) 1-6 vs 7-12*
Baseline [7]	2.72	-0.12	5.23	0.032**
	(11.33)	*(2.27)*	*(21.08)*	
Carrot [8]	2.59	-12.70**	15.04*	0.013**
	(14.92)	*(8.01)*	*(24.94)*	
Stick [8]	-5.00	-3.68	-6.62	0.556
	(8.41)	*(6.76)*	*(7.08)*	
WMW U-Test p-values (2-sided):				
Baseline vs Carrot	1.000	0.189	0.779	
Baseline vs Stick	0.397	0.281	0.397	
Carrot vs Stick	0.279	0.505	0.105	

Table 4: Summary of bubble measures for Assets X and Y in Round 2

This table reports median values of each bubble measure in the *Baseline*, *Carrot*, and *Stick* treatments during Round 2 (i.e. with experienced traders); median absolute deviations are displayed in parentheses. Markets contaminated by subjects who had participated in an earlier session are excluded. Panel A (B) reports bubble measure data relating to Asset X (Y). For definitions of the relevant bubble measures, refer to section 4.1.2. The statistical significance of the difference between treatments in each measure is assessed using a two-sided Wilcoxon Mann-Whitney U Test, under the null that values from both treatments come from the same distribution. Exact p-values are reported. Differences that are significant at the 10%, 5% and 1% level are denoted by *, **, and *** respectively.

Panel A: Asset X, Round 2

Treatment [N]	*Amplitude*	*Total Dispersion*	*Average Bias*	*Haessel R^2*	*Turnover*	*Normalised Deviation*	*Duration*	*Boom Duration*	*Bust Duration*
Baseline [7]	1.15	323.50	10.96	0.82	1.60	39.45	3.00	8.00	3.00
	(0.79)	*(247.00)*	*(13.55)*	*(0.18)*	*(0.69)*	*(32.05)*	*(1.00)*	*(1.00)*	*(1.00)*
Carrot [8]	0.80	302.00	2.94	0.85	2.04	64.30	4.00	5.50	3.00
	(0.35)	*(151.50)*	*(11.25)*	*(0.10)*	*(1.00)*	*(52.34)*	*(2.00)*	*(2.50)*	*(2.00)*
Stick [8]	1.97	689.00	40.17	0.77	1.66	101.15	5.00	8.00	3.00
	(1.59)	*(503.50)*	*(47.38)*	*(0.22)*	*(0.45)*	*(66.21)*	*(1.50)*	*(2.50)*	*(1.50)*
WMW U-Test p-values (two-sided):									
Baseline vs Carrot	0.463	0.694	0.336	0.779	0.779	0.694	0.411	0.271	0.540
Baseline vs Stick	0.955	0.867	1.000	0.779	0.536	0.955	0.797	1.000	0.717
Carrot vs Stick	0.645	0.505	0.279	1.000	0.382	0.645	0.345	0.456	0.917

Panel B: Asset Y, Round 2

Treatment [N]	*Amplitude*	*Total Dispersion*	*Average Bias*	*Haessel R^2*	*Turnover*	*Normalised Deviation*	*Duration*	*Boom Duration*	*Bust Duration*
Baseline [7]	2.16	626.00	24.33	0.52	1.50	127.50	5.00	7.00	3.00
	(0.72)	*(492.50)*	*(68.88)*	*(0.13)*	*(0.47)*	*(64.98)*	*(3.00)*	*(4.00)*	*(2.00)*
Carrot [8]	1.91	391.75	1.08	0.78	1.89	71.52	4.50	5.50	5.00
	(0.83)	*(56.25)*	*(12.58)*	*(0.12)*	*(0.63)*	*(48.75)*	*(1.00)*	*(1.50)*	*(1.50)*
Stick [8]	1.23	481.50	11.25	0.84	1.26	60.95	3.00	3.00	3.50
	(0.86)	*(216.75)*	*(47.93)*	*(0.13)*	*(0.18)*	*(40.14)*	*(1.00)*	*(3.00)*	*(2.00)*
WMW U-Test p-values (two-sided):									
Baseline vs Carrot	0.867	0.121	0.094*	0.867	1.000	0.779	0.715	0.183	0.184
Baseline vs Stick	0.613	0.281	0.536	0.613	0.463	0.281	0.282	0.035**	0.378
Carrot vs Stick	0.574	0.505	0.721	0.574	0.487	0.798	0.209	0.197	0.530

Table 5: Comparing bubble measures between rounds

This table reports the results of within-treatment comparisons of the bubble measures between market rounds in the *Baseline*, *Carrot*, and *Stick* treatments. Markets contaminated by subjects who had participated in an earlier session are excluded. The values shown below are p-values from a two-sided Wilcoxon signed-rank test of the null hypothesis that bubble measure values do not differ significantly between rounds 1 and 2. Differences that are significant at the 10%, 5% and 1% level are denoted by *, **, and *** respectively

Panel A: Asset X, Round 1 vs Round 2:

Treatment [N]	*Amplitude*	*Total Dispersion*	*Average Bias*	*Haessel R^2*	*Turnover*	*Normalised Deviation*	*Duration*	*Boom Duration*	*Bust Duration*
Baseline [7]	1.000	0.499	0.128	0.499	0.176	0.866	0.317	0.230	0.333
Carrot [8]	0.575	0.327	0.575	0.674	0.012**	0.012**	0.160	0.323	0.256
Stick [8]	0.674	0.674	0.779	0.036**	0.030**	0.093*	0.574	0.014**	0.078*

Panel B: Asset Y, Round 1 vs Round 2:

Treatment [N]	*Amplitude*	*Total Dispersion*	*Average Bias*	*Haessel R^2*	*Turnover*	*Normalised Deviation*	*Duration*	*Boom Duration*	*Bust Duration*
Baseline [7]	0.735	0.091*	0.128	0.612	0.018**	0.866	0.475	0.932	0.795
Carrot [8]	0.779	0.093*	0.779	0.401	0.012**	0.012**	0.013**	0.477	0.725
Stick [8]	0.208	0.401	0.161	0.124	0.036**	0.124	0.031**	0.011**	0.019**

Table 6: Summary of bubble measures in Round 1 using rank-order tournaments

This table reports median values of each bubble measure in the *Baseline*, *GilCarrot*, and *GilStick* treatments during Round 1; median absolute deviations are displayed in parentheses. Markets contaminated by subjects who had participated in an earlier session are excluded. Panel A (B) reports bubble measure data relating to Asset X (Y). For definitions of the relevant bubble measures, refer to section 4.1.2. The statistical significance of the difference between treatments in each measure is assessed using a two-sided Wilcoxon Mann-Whitney U Test, under the null that values from both treatments come from the same distribution. Exact p-values are reported. Differences that are significant at the 10%, 5% and 1% level are denoted by *, **, and *** respectively.

Panel A: Asset X, Round 1:

Treatment [N]	*Amplitude*	*Total Dispersion*	*Average Bias*	*Haessel R^2*	*Turnover*	*Normalised Deviation*	*Duration*	*Boom Duration*	*Bust Duration*
Baseline [7]	3.31	298.00	1.65	0.78	2.45	83.43	5.00	5.00	4.00
	(1.53)	*(181.50)*	*(15.73)*	*(0.16)*	*(0.72)*	*(45.08)*	*(2.00)*	*(2.00)*	*(1.00)*
GilCarrot[6]	3.05	758.50	23.42	0.46	3.91	262.55	5.00	8.00	3.50
	(1.71)	*(202.25)*	*(19.52)*	*(0.11)*	*(0.85)*	*(62.06)*	*(2.50)*	*(1.00)*	*(1.00)*
GilStick [6]	1.99	614.50	28.85	0.63	2.79	156.44	6.00	8.00	2.50
	(0.58)	*(222.00)*	*(34.25)*	*(0.08)*	*(0.29)*	*(64.60)*	*(1.00)*	*(2.00)*	*(1.50)*
WMW U-Test p-values (two-sided):									
Baseline vs GilCarrot	0.945	0.366	0.836	0.051*	0.073*	0.073*	0.736	0.178	0.950
Baseline vs GilStick	0.731	0.445	0.181	0.234	0.509	0.234	0.457	0.229	0.668
GilCarrot vs GilStick	0.589	0.937	0.589	0.240	0.167	0.485	0.558	0.864	0.381

Panel B: Asset Y, Round 1

Treatment [N]	*Amplitude*	*Total Dispersion*	*Average Bias*	*Haessel R^2*	*Turnover*	*Normalised Deviation*	*Duration*	*Boom Duration*	*Bust Duration*
Baseline [7]	1.63	530.50	15.38	0.77	2.03	99.06	4.00	7.00	4.00
	(1.01)	*(150.50)*	*(25.33)*	*(0.05)*	*(0.35)*	*(54.11)*	*(1.00)*	*(1.00)*	*(2.00)*
GilCarrot[6]	3.44	676.75	-5.05	0.35	2.79	169.82	4.00	5.00	3.50
	(1.99)	*(217.50)*	*(31.55)*	*(0.25)*	*(0.44)*	*(46.82)*	*(2.00)*	*(1.00)*	*(1.50)*
GilStick [6]	2.52	518.50	12.20	0.61	2.62	117.21	5.50	6.00	2.50
	(1.01)	*(103.25)*	*(32.79)*	*(0.10)*	*(0.93)*	*(53.74)*	*(1.00)*	*(3.50)*	*(1.00)*
WMW U-Test p-values (two-sided):									
Baseline vs GilCarrot	0.731	0.836	0.366	0.731	0.445	0.534	0.871	0.508	0.530
Baseline vs GilStick	0.731	1.000	0.836	0.945	0.731	0.731	0.508	0.458	0.751
GilCarrot vs GilStick	0.818	0.699	0.699	0.699	0.937	0.937	0.675	0.894	0.374

Table 7: Summary of bubble measures in Round 2 using rank-order tournaments

This table reports median values of each bubble measure in the *Baseline*, *GilCarrot*, and *GilStick* treatments during Round 2; median absolute deviations are displayed in parentheses. Markets contaminated by subjects who had participated in an earlier session are excluded. Panel A (B) reports bubble measure data relating to Asset X (Y). For definitions of the relevant bubble measures, see section 4.1.2. The statistical significance of the difference between treatments in each measure is assessed using a two-sided Wilcoxon Mann-Whitney U Test, under the null that values from both treatments come from the same distribution. Exact p-values are reported. Differences that are significant at the 10%, 5% and 1% level are denoted by *, **, and *** respectively.

Panel A: Asset X, Round 2:

Treatment [N]	*Amplitude*	*Total Dispersion*	*Average Bias*	*Haessel R^2*	*Turnover*	*Normalised Deviation*	*Duration*	*Boom Duration*	*Bust Duration*
Baseline [7]	1.15	323.50	10.96	0.82	1.60	39.45	3.00	8.00	3.00
	(0.79)	*(247.00)*	*(13.55)*	*(0.18)*	*(0.69)*	*(32.05)*	*(1.00)*	*(1.00)*	*(1.00)*
GilCarrot[6]	2.33	457.75	36.81	0.87	2.55	115.18	5.00	7.50	2.00
	(1.12)	*(119.75)*	*(22.92)*	*(0.02)*	*(0.48)*	*(45.08)*	*(0.50)*	*(2.50)*	*(1.00)*
GilStick [6]	0.92	446.25	25.54	0.85	2.59	81.13	3.50	6.00	3.00
	(0.49)	*(234.25)*	*(30.01)*	*(0.12)*	*(0.86)*	*(59.78)*	*(1.00)*	*(2.00)*	*(0.50)*
WMW U-Test p-values (two-sided):									
Baseline vs Gil-Carrot	0.366	0.628	0.366	0.445	0.421	0.628	0.864	0.810	0.804
Baseline vs Gil-Stick	0.731	0.836	0.945	0.731	0.219	0.945	0.493	0.650	0.935
Gil-Carrot vs Gil-Stick	0.065*	0.485	0.485	0.818	0.784	0.394	0.210	0.303	0.498

Panel B: Asset Y, Round 2

Treatment [N]	*Amplitude*	*Total Dispersion*	*Average Bias*	*Haessel R^2*	*Turnover*	*Normalised Deviation*	*Duration*	*Boom Duration*	*Bust Duration*
Baseline [7]	2.16	626.00	24.33	0.52	1.50	127.50	5.00	7.00	3.00
	(0.72)	*(492.50)*	*(68.88)*	*(0.13)*	*(0.47)*	*(64.98)*	*(3.00)*	*(4.00)*	*(2.00)*
GilCarrot[6]	1.99	462.50	23.10	0.84	1.74	76.56	6.00	7.00	4.00
	(0.86)	*(143.25)*	*(18.19)*	*(0.02)*	*(0.26)*	*(28.16)*	*(2.00)*	*(2.00)*	*(1.50)*
GilStick [6]	1.75	694.25	52.35	0.57	2.25	96.66	2.50	7.00	1.50
	(0.60)	*(228.25)*	*(18.20)*	*(0.23)*	*(0.94)*	*(39.36)*	*(1.00)*	*(4.00)*	*(0.50)*
WMW U-Test p-values (two-sided):									
Baseline vs GilCarrot	0.945	0.295	0.628	0.181	0.628	0.628	0.386	0.833	0.705
Baseline vs GilStick	0.628	0.534	0.731	1.000	0.313	0.836	0.422	0.756	0.755
GilCarrot vs GilStick	0.818	0.589	0.310	0.240	0.589	0.589	0.106	0.985	0.284

Table 8: Comparing rank-order tournament bubble measures between rounds

This table reports the results of within-treatment comparisons of bubble measures between market rounds in the *GilCarrot* and *GilStick* treatments. Markets contaminated by subjects who had participated in an earlier session are excluded. Panel A (B) reports for asset X (Y). The values shown below are p-values from a two-sided Wilcoxon signed-rank test of the null hypothesis that bubble measure values do not differ significantly between rounds 1 and 2. Differences that are significant at the 10%, 5% and 1% level are denoted by *, **, and *** respectively

Panel A: Asset X, Round 1 vs Round 2:

Treatment [N]	*Amplitude*	*Total Dispersion*	*Average Bias*	*Haessel* R^2	*Turnover*	*Normalised Deviation*	*Duration*	*Boom Duration*	*Bust Duration*
GilCarrot [6]	0.917	0.917	0.116	0.173	0.028**	0.463	0.674	0.916	0.190
GilStick [6]	0.028**	0.046**	0.249	0.173	0.463	0.075*	0.091*	0.058*	0.593

Panel B: Asset Y, Round 1 vs Round 2:

Treatment [N]	*Amplitude*	*Total Dispersion*	*Average Bias*	*Haessel* R^2	*Turnover*	*Normalised Deviation*	*Duration*	*Boom Duration*	*Bust Duration*
GilCarrot [6]	0.917	0.917	0.463	0.116	0.028**	0.173	0.461	0.597	0.665
GilStick [6]	0.075*	0.600	0.173	0.917	0.046**	0.249	0.058*	0.525	0.597

Table 9: Average Prediction Errors – Rank-order tournaments

Median values of the *Average Prediction Error* in the *Baseline, Carrot, and Stick* treatments in Round 1 and 2 are shown below in Panels A and B respectively, with the associated median absolute deviations in parentheses. Markets contaminated by subjects who had participated in an earlier session are excluded. *Average Prediction Error* is calculated using all periods in a market, the first 6 periods, and the final 6 periods in *Avg PredErr*, *AvgPredErr_p1to6*, and *AvgPredErr_p7to12* respectively. The statistical significance of the individual measures is assessed using a (two-sided) one-sample Wilcoxon Signed-rank test, under the null that the median is equal to zero. The statistical significance of the difference between treatments is assessed using the Wilcoxon Mann-Whitney U Test under the null that values from both treatments come from the same distribution. The statistical significance of the difference between *AvgPredErr_p1to6* and *AvgPredErr_p7to12* within each treatment is assessed using a (paired-sample) Wilcoxon Signed-rank test, against the one-sided alternative hypothesis that *AvgPredErr_p7to12* > *AvgPredErr_p1to6*. Differences that are significant at the 10%, 5% and 1% level are denoted by *, **, and *** respectively.

Panel A: Round 1

Treatment [N]	*Avg PredErr (%)*	*AvgPredErr _p1to6 (%)*	*AvgPredErr _p7to12 (%)*	*Signed-rank p-value (1-sided) 1-6 vs 7-12*
Baseline [7]	3.78	-0.32	7.88	0.064*
	(9.46)	*(3.36)*	*(9.26)*	
GilCarrot [6]	-3.22	-3.21	-3.09	0.377
	(8.79)	*(2.78)*	*(15.48)*	
GilStick [6]	2.05	-0.92	6.00	0.377
	(7.17)	*(4.64)*	*(9.49)*	
WMW U-Test p-values (2-sided):				
Baseline vs Gil-Carrot	0.295	0.836	0.295	
Baseline vs Gil-Stick	0.295	0.945	0.534	
Gil-Carrot vs Gil-Stick	0.937	0.699	0.937	

Panel B: Round 2

Treatment [N]	*Avg PredErr (%)*	*AvgPredErr _p1to6 (%)*	*AvgPredErr _p7to12 (%)*	*Signed-rank p-value (1-sided) 1-6 vs 7-12*
Baseline [7]	2.72	-0.12	5.23	0.032**
	(11.33)	*(2.27)*	*(21.08)*	
GilCarrot [6]	-4.51	-4.14*	-3.63	0.377
	(4.95)	*(4.27)*	*(6.68)*	
GilStick [6]	3.80	3.46	3.85	0.058*
	(9.64)	*(4.29)*	*(8.95)*	
WMW U-Test p-values (2-sided):				
Baseline vs Gil-Carrot	0.295	0.181	0.366	
Baseline vs Gil-Stick	0.836	0.836	1.000	
Gil-Carrot vs Gil-Stick	0.180	0.093*	0.180	

Figure 1: Trading Interface

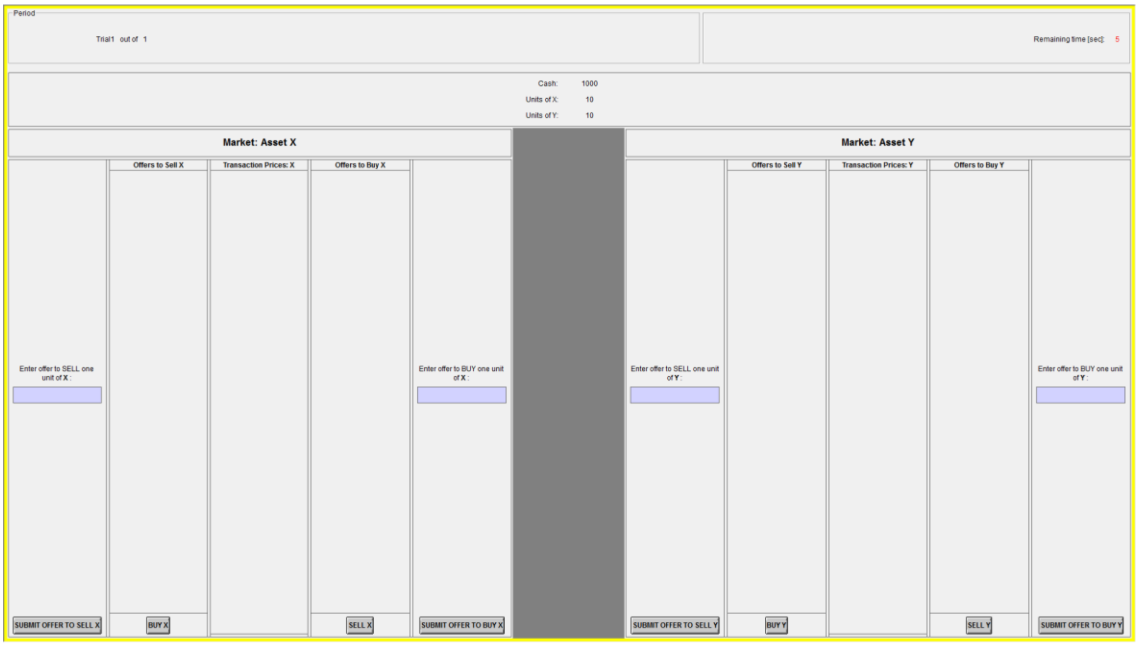


Figure 2: Fundamental value process, assets X and Y

The solid black line in the graph below depicts the risk-neutral fundamental value process of assets X and Y. Both assets pay an expected dividend of 20 per period. The dashed and solid grey (blue) lines depict the largest and smallest possible cumulative future dividend realisations of asset X (Y) respectively. Asset X pays a minimum of 10 francs in dividends each period, and a maximum of 30 per period. Asset Y pays a minimum of zero every period and a maximum of 100 every period. Hence, the blue dotted line, which is only partially graphed, starts at 1200 in period 1 and falls in steps of 100 in each ensuing period.

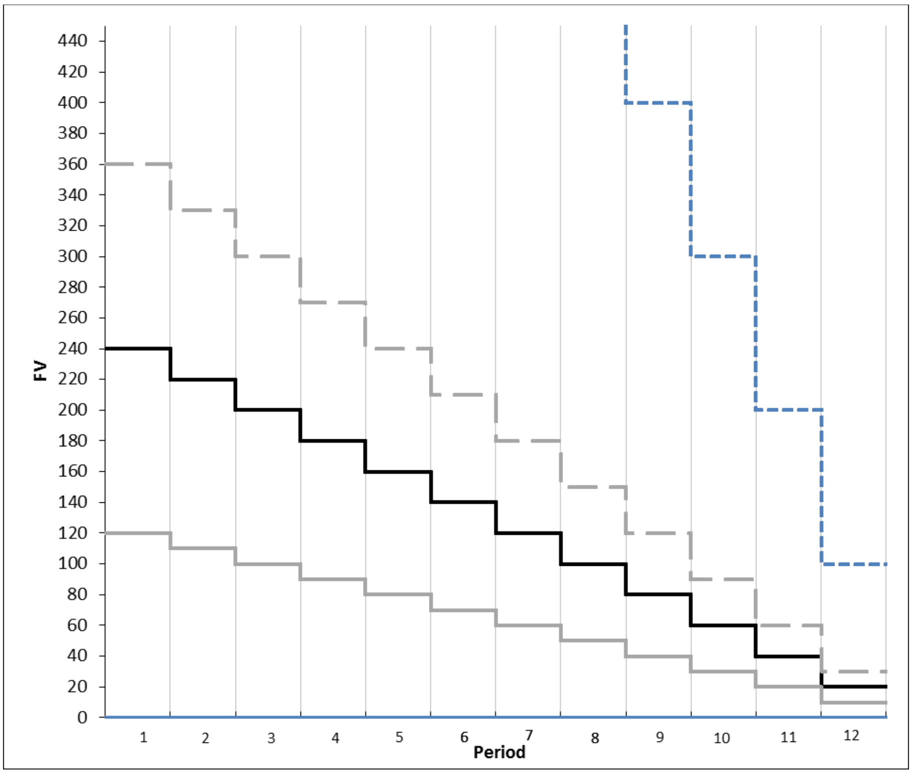
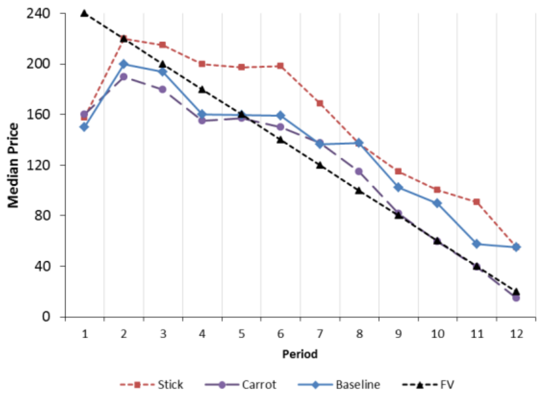
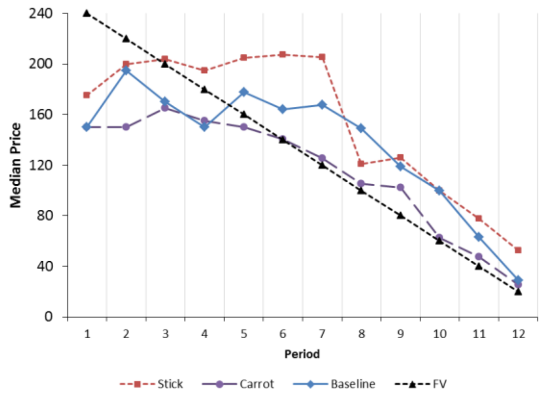


Figure 3: Median prices in Round 1

Median transaction prices in the *Baseline* (solid blue line), *Carrot* (dashed purple line), and *Stick* (dotted red line) treatments during the first round of the market (i.e. with inexperienced traders) are shown below for the 'low-risk' asset X (panel (a)) and 'high-risk' asset Y (panel (b)), along with the risk-neutral fundamental value process for each asset (dashed black line). For each treatment, the plotted median price in each period is the median of the median transaction prices from all markets belonging to that treatment. Any markets that were 'contaminated' by the presence of subjects who had participated in an earlier session of the experiment are excluded.



(a) Asset X, Round 1



(b) Asset Y, Round 1

Figure 4: Median value of *Prediction Error*, Round 1

The figure below plots the evolution of the median *Prediction Error* in the *Baseline* (solid blue line), *Carrot* (dashed purple line), and *Stick* (dotted red line) treatments during the first round of the market (i.e. with inexperienced traders). For each treatment, the plotted value in each period is the median of the *Prediction Errors* from all markets in that treatment. *Prediction Error* is defined as the percentage difference between the relative price of Y (i.e. median price of asset Y divided by median price of asset X) and the risk-neutral benchmark of 1. Any markets that were 'contaminated' by the presence of subjects who had participated in an earlier session of the experiment are excluded.

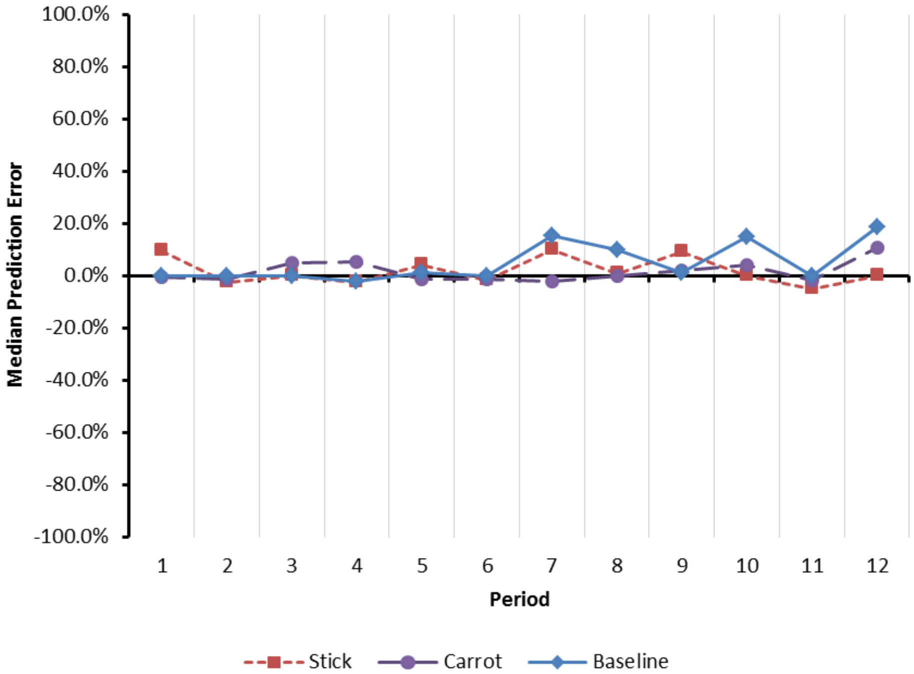
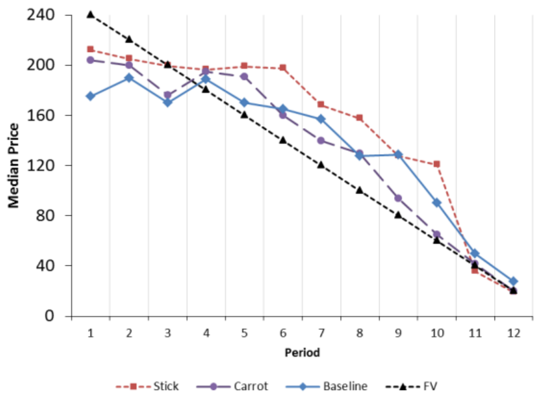
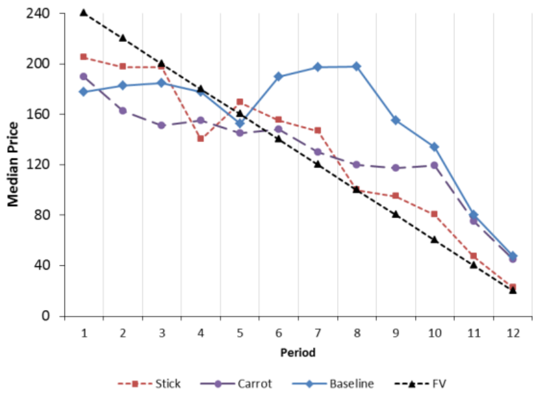


Figure 5: Median prices in Round 2

Median transaction prices in the *Baseline* (solid blue line), *Carrot* (dashed purple line), and *Stick* (dotted red line) treatments during the second round of the market (i.e. with experienced traders) are shown below for the 'low-risk' asset X (panel (a)) and 'high-risk' asset Y (panel (b)), along with the risk-neutral fundamental value process for each asset (black dotted line). For each treatment, the plotted median price in each period is the median of the median transaction prices from all markets belonging to that treatment. Any markets that were 'contaminated' by the presence of subjects who had participated in an earlier session of the experiment are excluded.



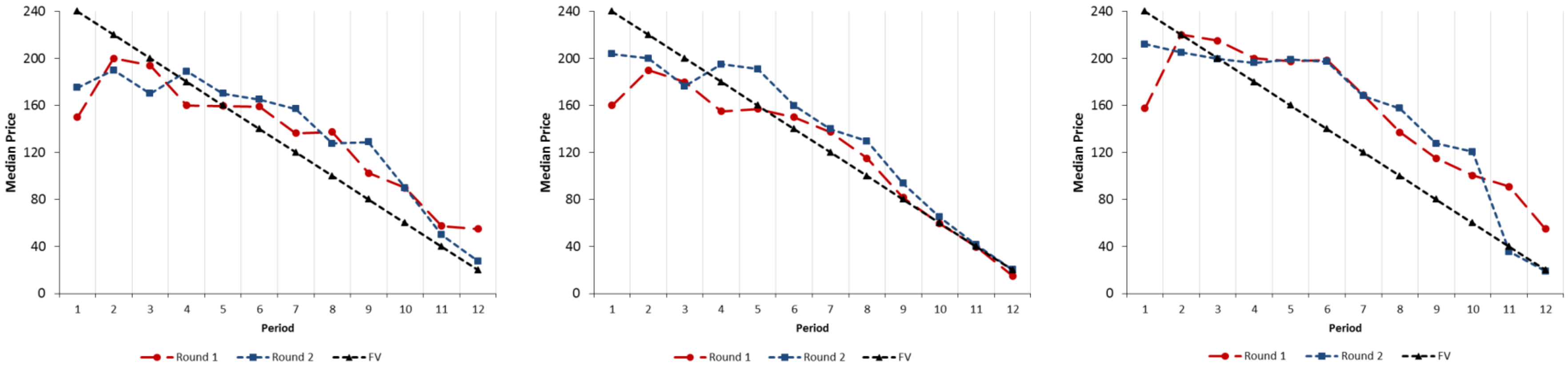
(a) Asset X, Round 2



(b) Asset Y, Round 2

Figure 6: Median prices for asset X, Round 1 vs. Round 2

This figure compares the median-price behaviour of the ‘low-risk’ asset X between the two rounds of the market in the *Baseline* (panel (a)), *Carrot* (panel (b), and *Stick* (panel (c)) treatment. The red dashed line depicts Round 1 prices; the blue dashed line depicts Round 2 prices, while the black dotted line is the fundamental value process. The plotted median price in each period is the median of the median transaction prices from all markets in that treatment. Any markets that were ‘contaminated’ by the presence of subjects who had participated in an earlier session of the experiment are excluded.



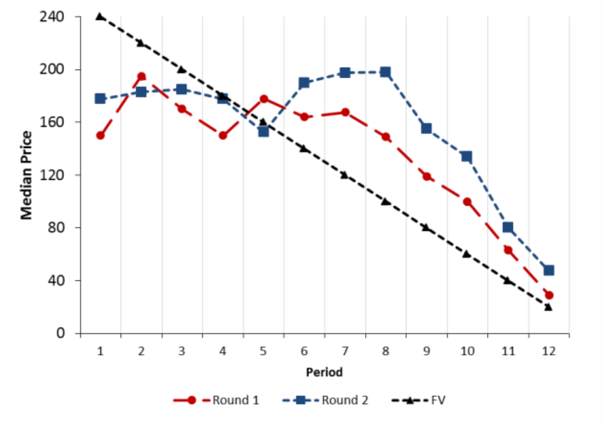
(a) Baseline, Asset X

(b) Carrot, Asset X

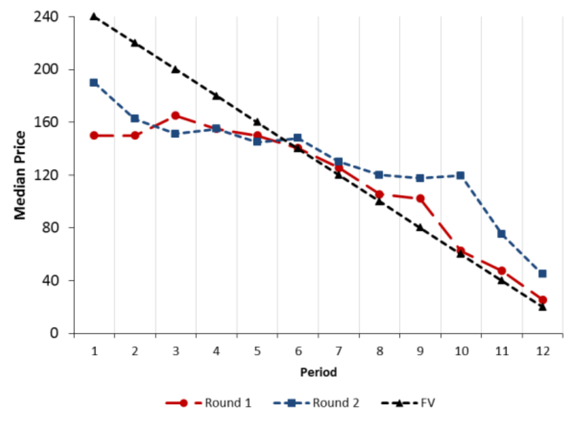
(c) Stick, Asset X

Figure 7: Median prices for asset Y, Round 1 vs. Round 2

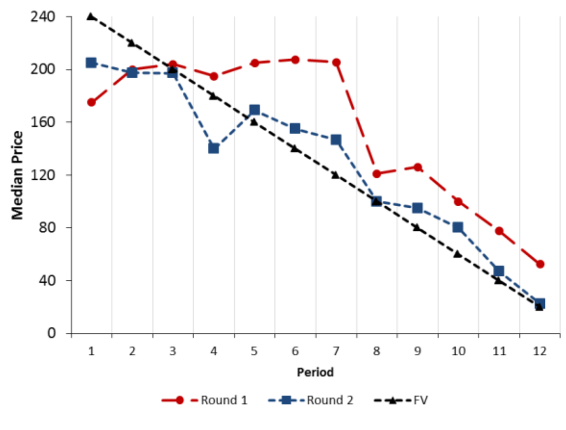
This figure compares the median-price behaviour of the 'high-risk' asset Y between the two rounds of the market in the *Baseline* (panel (a)), *Carrot* (panel (b), and *Stick* (panel (c)) treatment. The red dashed line depicts Round 1 prices; the blue dashed line depicts Round 2 prices, while the black dotted line is the fundamental value process. The plotted median price in each period is the median of the median transaction prices from all markets in that treatment. Any markets that were 'contaminated' by the presence of subjects who had participated in an earlier session of the experiment are excluded.



(a) Baseline, Asset Y



(b) Carrot, Asset Y



(c) Stick, Asset Y

Figure 8: Median value of *Prediction Error*, Round 2

The figure below plots the evolution of the median *Prediction Error* in the *Baseline* (solid blue line), *Carrot* (dashed purple line), and *Stick* (dashed red line) treatment during the second round of the market (i.e. with experienced traders). For each treatment, the plotted value in each period is the median of the *Prediction Errors* from all markets in that treatment. *Prediction Error* is defined as the percentage difference between the relative price of Y (i.e. median price of asset Y divided by median price of asset X) and the risk-neutral benchmark of 1. Any markets that were 'contaminated' by the presence of subjects who had participated in an earlier session of the experiment are excluded.

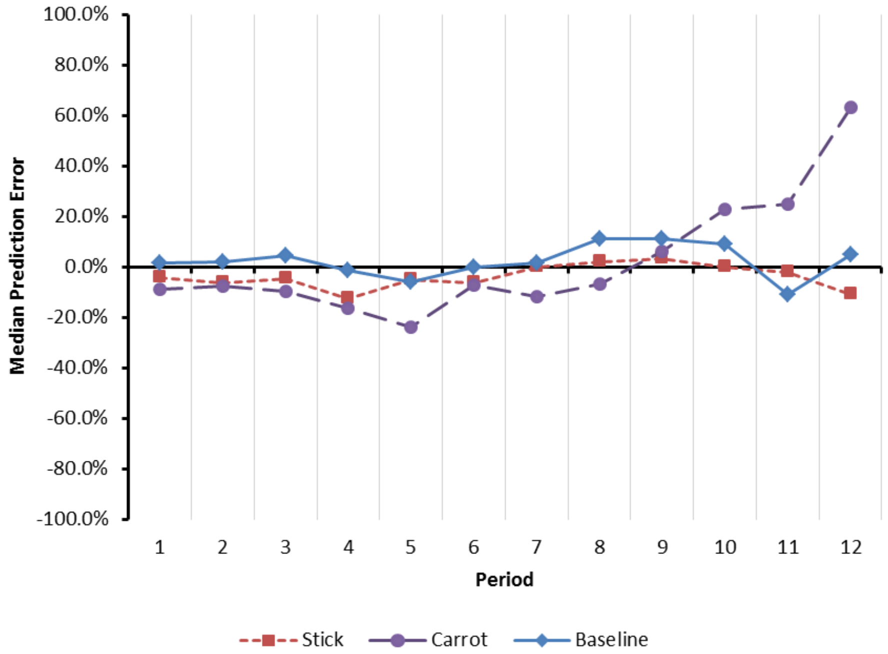
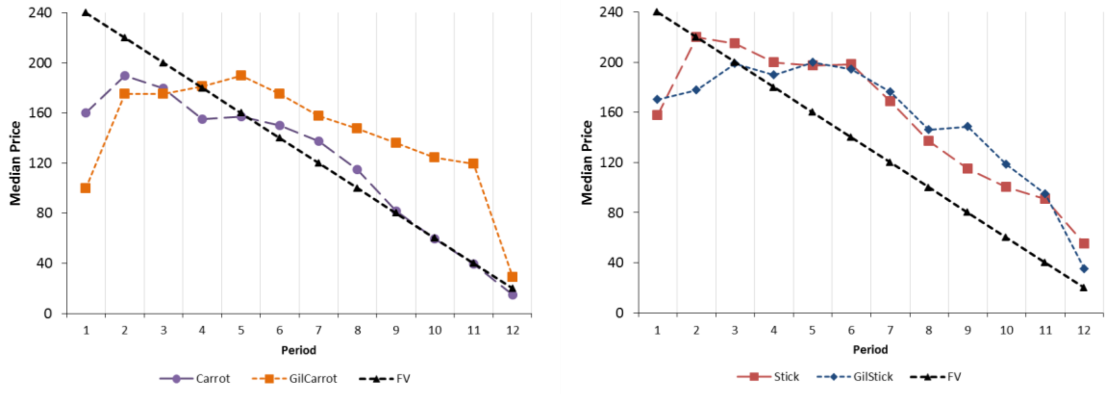


Figure 9: Median prices in J&I tournament vs. Gilpatric tournament, Asset X

The median-price behaviour of the 'low-risk' asset X in Round 1 (i.e. inexperienced traders) is compared between James and Isaac (200)-based tournament treatments and the corresponding Gilpatric (2009)-based rank-order tournament treatment below. Panel (a) depicts median prices in the *Carrot* treatment (purple dashed line) and the *GilCarrot* treatment (yellow dotted line), while Panel (b) shows medians prices in the *Stick* treatment (red dashed line) and the *GilStick* (blue dotted line) treatment. Also shown is the risk-neutral fundamental process (black dashed line). For each treatment, the plotted median price in each period is the median of the median transaction prices from all markets belonging to that treatment. Any markets that were 'contaminated' by the presence of subjects who had participated in an earlier session of the experiment are excluded.

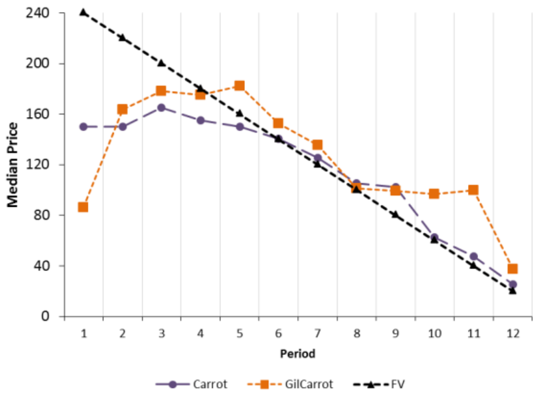


(a) Asset X, Round 1 - Carrot vs GilCarrot

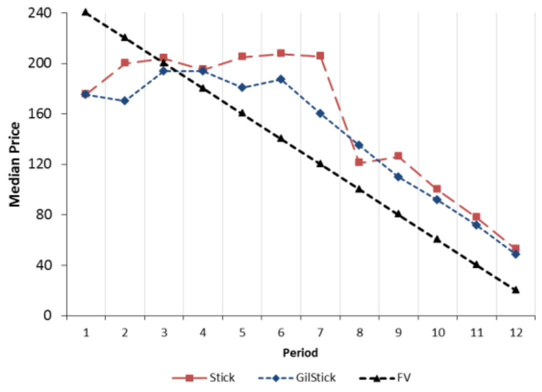
(b) Asset X, Round 1 - Stick vs GilStick

Figure 10: Median prices in J&I tournament vs. Gilpatric tournament, Asset Y

The median-price behaviour of the ‘high-risk’ asset Y in Round 1 (i.e. inexperienced traders) is compared between James and Isaac (200)-based tournament treatments and the corresponding Gilpatric (2009)-based rank-order tournament treatment below. Panel (a) depicts median prices in the *Carrot* treatment (purple dashed line) and the *GilCarrot* treatment (yellow dotted line), while Panel (b) shows medians prices in the *Stick* treatment (red dashed line) and the *GilStick* (blue dotted line) treatment. Also shown is the risk-neutral fundamental process (black dashed line). For each treatment, the plotted median price in each period is the median of the median transaction prices from all markets belonging to that treatment. Any markets that were ‘contaminated’ by the presence of subjects who had participated in an earlier session of the experiment are excluded.



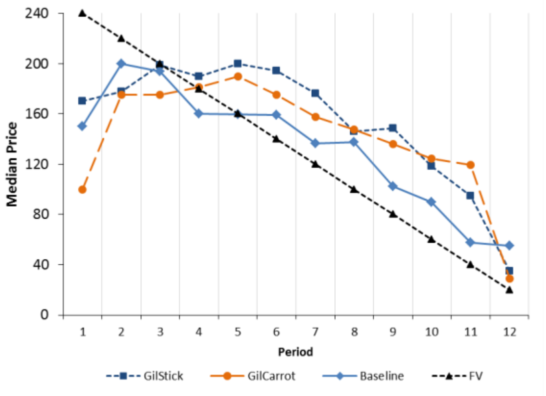
(a) Asset Y, Round 1 - Carrot vs GilCarrot



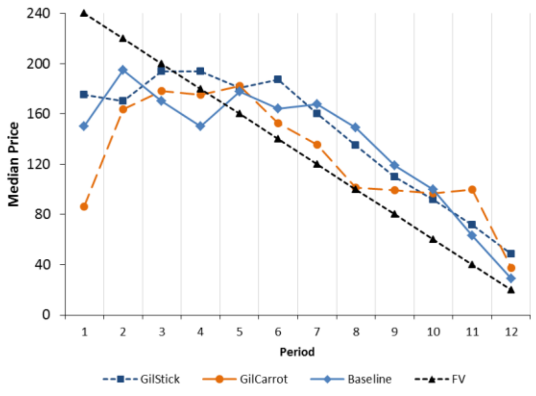
(b) Asset Y, Round 1 - Stick vs GilStick

Figure 11: Median prices in rank-order tournaments, Round 1

Median transaction prices in the *Baseline* (solid blue line), *GilCarrot* (dashed yellow line), and *GilStick* (blue dotted line) treatments during the first round of the market (i.e. with inexperienced traders) are shown below for the 'low-risk' asset X (panel (a)) and 'high-risk' asset Y (panel (b)), along with the risk-neutral fundamental value process (dashed black line). For each treatment, the plotted median price in each period is the median of the median transaction prices from all markets belonging to that treatment. Any markets that were 'contaminated' by the presence of subjects who had participated in an earlier session of the experiment are excluded.



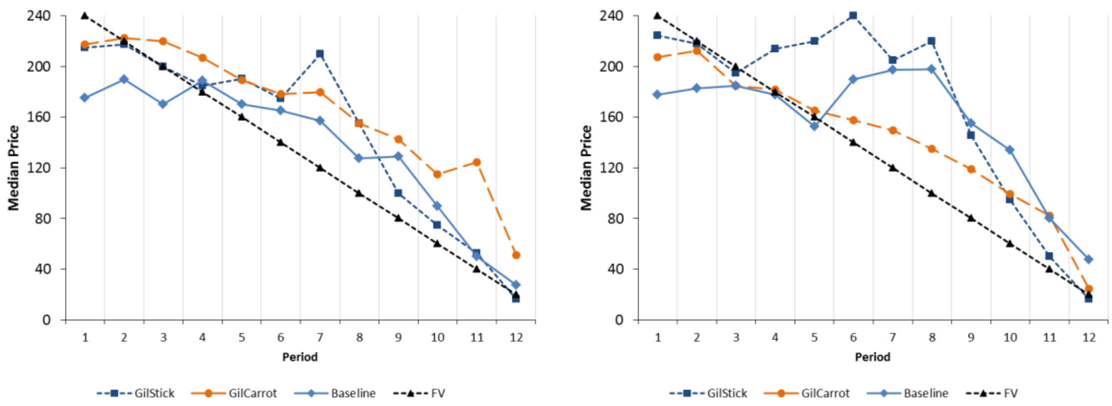
(a) Asset X, Round 1



(b) Asset Y, Round 1

Figure 12: Median prices in rank-order tournaments, Round 2

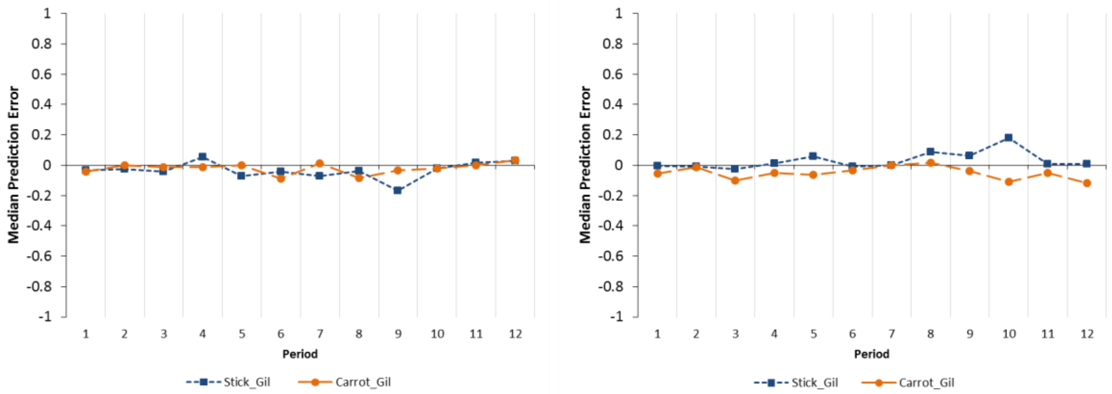
Median transaction prices in the *Baseline* (solid blue line), *GilCarrot* (dashed yellow line), and *GilStick* (blue dotted line) treatments during the second round of the market (i.e. with experienced traders) are shown below for the 'low-risk' asset X (panel (a)) and 'high-risk' asset Y (panel (b)), along with the risk-neutral fundamental value process (dashed black line). For each treatment, the plotted median price in each period is the median of the median transaction prices from all markets belonging to that treatment. Any markets that were 'contaminated' by the presence of subjects who had participated in an earlier session of the experiment are excluded.



(a) Asset X, Round 2 (b) Asset Y, Round 2

Figure 13: Median value of *Prediction Error*, rank-order tournaments

The evolution of the median *Prediction Error* in the *GilCarrot* (dashed yellow line), and *GilStick* (dashed blue line) treatment is shown below for Round 1 of the market in panel (a) and Round 2 in panel (b). For each treatment, the plotted value in each period is the median of the *Prediction Errors* from all markets in that treatment. *Prediction Error* is defined as the percentage difference between the relative price of Y (i.e. median price of asset Y divided by median price of asset X) and the risk-neutral benchmark of 1. Any markets that were 'contaminated' by the presence of subjects who had participated in an earlier session of the experiment are excluded.



(a) Round 1 (b) Round 2

Appendix A: Additional Tables

Tables A1-A4 below display the individual bubble measure values from each market of each treatment. Table A1 (A2) reports for asset X (Y) in Round 1 of the market. Table A3 (A4) reports for asset X (Y) in Round 2. The relevant bubble measures are defined in section 4.1.2.

Table A1: Bubble measures for asset X in Round 1

Panel A: Baseline

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*H-R^2*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
Baseline	B1	3.36	797.50	66.46	0.78	1.91	128.51	5	12	0
	B2	1.79	291.50	17.38	0.94	2.45	53.45	5	5	1
	B3	3.31	298.00	-7.83	0.56	1.53	43.43	4	4	5
	B4	3.42	553.00	-3.92	0.47	3.33	167.78	11	7	5
	B6	0.82	286.50	-15.54	0.81	2.93	83.43	3	5	4
	B7	7.30	947.00	18.58	0.68	3.17	263.97	11	7	5
	B8	0.40	116.50	1.65	0.94	1.26	17.40	3	2	2
	Median:	**3.31**	**298.00**	**1.65**	**0.78**	**2.45**	**83.43**	**5.00**	**5.00**	**4.00**

Panel B: Carrot

	Market	*Amp*	*Tot. Disp.*	*Avg. Bias*	*H-R^2*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
Carrot	C1	0.41	115.50	-6.71	0.89	2.93	41.53	3	1	3
	C2	4.82	1527.50	99.29	0.09	4.00	331.88	6	5	4
	C3	1.50	868.50	-67.54	0.15	3.80	265.09	10	2	9
	C4	9.02	997.50	42.29	0.20	2.94	267.43	3	8	1
	C6	1.53	593.50	-12.13	0.23	4.40	238.53	9	5	5
	C7	5.13	545.50	26.29	0.92	2.49	103.66	11	9	3
	C8	0.55	228.50	-16.54	0.96	2.69	57.86	4	2	6
	C9	0.50	250.00	-31.25	0.91	0.63	20.67	2	0	7
	Median:	**1.51**	**569.50**	**-9.42**	**0.56**	**2.93**	**171.09**	**5.00**	**3.50**	**4.50**

Panel C: Stick

	Market	*Amp*	*Tot. Disp.*	*Avg. Bias*	*H-R^2*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
Stick	S1	9.58	1435.00	119.58	0.02	3.74	458.57	10	12	0
	S2	7.89	961.00	58.58	0.01	1.88	154.08	5	10	1
	S3	2.19	474.50	0.88	0.66	2.38	112.43	3	8	4
	S4	4.33	527.00	35.58	0.65	2.13	85.35	3	11	1
	S5	0.57	198.00	-0.75	0.91	1.78	23.73	5	3	6
	S6	2.97	875.50	55.46	0.27	1.28	78.68	4	11	1
	S7	1.22	340.00	15.67	0.81	2.20	58.69	5	8	2
	S8	2.28	687.00	2.75	0.12	2.86	184.57	2	10	2
	Median:	**2.63**	**607.00**	**25.63**	**0.46**	**2.16**	**98.89**	**4.50**	**10.00**	**1.50**

Table A1 cont.

Panel D: GilCarrot

	Market	*Amp*	*Tot. Disp.*	*Avg. Bias*	*H-R²*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
GilCarrot	GC1	3.30	913.00	32.75	0.08	3.83	284.38	7	9	3
	GC2	2.81	747.00	43.92	0.39	4.95	349.18	2	10	2
	GC3	4.47	770.00	-31.50	0.33	4.58	300.05	11	4	8
	GC4	0.63	229.00	14.08	0.91	4.00	54.13	2	7	4
	GC5	1.04	353.50	-8.29	0.55	1.63	101.17	6	8	2
	GC6	10.45	1008.50	41.96	0.52	2.87	240.73	4	8	4
	Median:	**3.05**	**758.50**	**23.42**	**0.46**	**3.91**	**262.55**	**5.00**	**8.00**	**3.50**

Panel E: GilStick

	Market	*Amp*	*Tot. Disp.*	*Avg. Bias*	*H-R²*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
GilStick	GS1	1.49	682.50	-12.46	0.09	2.75	225.20	6	6	4
	GS2	2.65	899.00	73.25	0.60	2.43	201.89	7	11	1
	GS3	0.79	280.00	1.67	0.74	2.83	67.68	5	5	3
	GS4	1.62	455.00	24.58	0.75	3.00	96.00	4	9	1
	GS5	9.71	1606.00	85.33	0.57	3.85	302.95	9	7	5
	GS7	2.37	546.50	33.13	0.67	2.43	111.00	6	10	2
	Median:	**1.99**	**614.50**	**28.85**	**0.63**	**2.79**	**156.44**	**6.00**	**8.00**	**2.50**

Note: Market B5 in the *Baseline* treatment, C5 in the *Carrot* treatment, and GS6 in the *GilStick* treatment are excluded because they contain subjects who participated in an earlier session of the experiment.

Table A2: Bubble measures for asset Y in Round 1

Panel A: Baseline

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*Haessel R^2*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
Baseline	B1	1.63	681.00	53.58	0.78	2.06	99.06	4	10	1
	B2	2.66	514.50	46.77	0.82	1.75	77.20	3	7	0
	B3	0.62	310.50	-16.88	0.77	1.55	44.95	4	3	6
	B4	4.25	629.50	0.29	0.05	4.68	240.83	6	7	5
	B6	1.12	530.50	-9.96	0.53	2.03	109.70	2	8	4
	B7	7.35	1089.00	22.42	0.54	3.29	316.63	11	7	4
	B8	0.81	289.50	15.38	0.81	1.69	38.23	5	9	2
	Median:	**1.63**	**530.50**	**15.38**	**0.77**	**2.03**	**99.06**	**4.00**	**7.00**	**4.00**

Panel B: Carrot

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*$H\text{-}R^2$*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
Carrot	C1	0.47	168.50	-9.38	0.88	2.80	60.85	3	2	3
	C2	4.60	1570.50	98.96	0.05	2.85	265.10	6	7	2
	C3	1.61	886.00	-66.42	0.12	3.60	265.17	10	3	9
	C4	4.61	882.50	-10.54	0.05	2.66	224.34	6	5	3
	C6	1.27	559.50	-5.96	0.32	4.45	241.93	5	7	4
	C7	4.54	610.00	18.33	0.33	2.86	163.89	7	6	4
	C8	0.66	200.00	-14.50	0.86	3.43	60.57	7	2	6
	C9	1.92	469.00	-32.75	0.66	0.87	28.57	2	2	5
	Median:	**1.76**	**584.75**	**-9.96**	**0.32**	**2.85**	**194.11**	**6.00**	**4.00**	**4.00**

Panel C: Stick

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*$H\text{-}R^2$*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
Stick	S1	5.74	1582.00	131.83	0.10	1.74	209.09	6	12	0
	S2	7.49	1121.00	58.58	0.02	1.98	211.43	5	9	3
	S3	1.66	473.50	-4.88	0.74	2.98	132.93	4	6	6
	S4	1.05	363.50	10.96	0.70	2.40	85.05	3	8	3
	S5	0.93	228.50	4.46	0.88	1.00	16.63	4	3	3
	S6	2.66	996.50	62.21	0.25	1.33	107.30	8	11	1
	S7	1.00	290.00	0.50	0.85	2.14	52.94	3	8	4
	S8	2.32	585.00	38.75	0.43	1.80	75.86	4	5	2
	Median:	**1.99**	**529.25**	**24.85**	**0.56**	**1.89**	**96.18**	**4.00**	**8.00**	**3.00**

Table A2 cont.

Panel D: GilCarrot

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*H-R²*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
GilCarrot	GC1	2.91	1000.50	44.13	0.06	2.65	216.85	9	9	3
	GC2	3.96	704.00	48.67	0.49	3.93	216.43	2	10	2
	GC3	4.68	788.00	-33.17	0.14	2.93	163.28	11	4	8
	GC4	0.71	194.50	9.54	0.92	3.18	51.63	2	6	4
	GC5	0.43	274.50	-19.65	0.93	1.63	101.57	3	4	2
	GC6	21.86	649.50	-40.04	0.21	2.30	176.37	5	4	7
	Median:	**3.44**	**676.75**	**-5.05**	**0.35**	**2.79**	**169.82**	**4.00**	**5.00**	**3.50**

Panel E: GilStick

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*H-R²*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
GilStick	GS1	1.34	631.00	-29.17	0.07	3.58	235.20	8	5	3
	GS2	3.02	889.50	79.05	0.60	1.71	116.97	4	7	1
	GS3	1.00	322.00	-4.30	0.63	1.45	41.33	5	2	3
	GS4	3.36	424.50	28.71	0.82	2.75	85.63	6	10	1
	GS5	4.24	520.00	-31.33	0.59	4.38	220.95	6	2	10
	GS7	2.03	517.00	36.42	0.78	2.49	117.46	3	9	2
	Median:	**2.52**	**518.50**	**12.20**	**0.61**	**2.62**	**117.21**	**5.50**	**6.00**	**2.50**

Note: Market B5 in the *Baseline* treatment, C5 in the *Carrot* treatment, and GS6 in the *GilStick* treatment are excluded because they contain subjects who participated in an earlier session of the experiment.

Table A3: Bubble measures for asset X in Round 2

Panel A: Baseline

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*H-R²*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
Baseline	B1	4.76	1324.00	120.36	0.02	2.29	261.09	7	11	0
	B2	0.53	279.50	19.63	0.93	1.50	31.80	3	8	2
	B3	0.62	221.00	-13.42	0.82	1.60	39.45	2	2	4
	B4	2.03	571.50	10.96	0.46	3.05	156.18	9	8	4
	B6	1.15	323.50	1.38	0.82	1.43	39.03	3	7	3
	B7	92.26	3083.00	234.75	0.32	3.00	510.40	11	9	3
	B8	0.35	76.50	-2.59	0.99	0.89	7.40	3	3	3
	Median:	**1.15**	**323.50**	**10.96**	**0.82**	**1.60**	**39.45**	**3.00**	**8.00**	**3.00**

Panel B: Carrot

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*H-R²*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
Carrot	C1	0.48	124.00	-10.36	0.92	0.75	10.60	2	1	6
	C2	0.83	265.00	6.25	0.93	3.65	71.55	2	4	5
	C3	1.95	829.50	-0.38	0.08	2.74	200.63	3	8	3
	C4	11.63	1346.50	110.54	0.05	2.00	228.09	5	9	1
	C6	0.76	339.00	-12.75	0.78	3.40	115.28	6	7	5
	C7	0.41	168.50	12.14	0.94	1.83	28.23	5	3	1
	C8	1.05	396.00	8.42	0.67	2.09	57.06	7	7	3
	C9	0.35	132.50	-16.19	0.98	0.47	6.70	2	1	3
	Median:	**0.80**	**302.00**	**2.94**	**0.85**	**2.04**	**64.30**	**4.00**	**5.50**	**3.00**

Panel C: Stick

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*H-R²*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
Stick	S1	3.35	1742.00	139.67	0.48	2.40	288.69	6	10	2
	S2	8.07	989.50	67.46	0.03	1.88	161.28	11	9	3
	S3	0.54	480.00	-40.00	0.87	1.55	77.33	5	0	12
	S4	7.21	1865.50	169.59	0.67	1.18	124.98	8	11	0
	S5	0.21	96.00	-8.00	0.99	0.78	7.65	2	0	5
	S6	0.56	248.00	18.83	0.95	1.78	28.85	5	7	3
	S7	0.59	123.00	-6.42	0.98	1.26	16.43	3	4	5
	S8	9.38	898.00	61.50	0.06	2.17	161.20	4	9	2
	Median:	**1.97**	**689.00**	**40.17**	**0.77**	**1.66**	**101.15**	**5.00**	**8.00**	**3.00**

Table A3 cont.

Panel D: GilCarrot

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*H-R²*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
GilCarrot	GC1	1.34	416.00	13.83	0.90	2.80	103.05	3	7	5
	GC2	1.08	434.50	34.38	0.86	3.00	109.70	5	8	2
	GC3	18.34	1629.50	59.68	0.86	2.30	366.90	5	5	5
	GC4	3.29	481.00	39.25	0.87	3.05	120.65	6	11	1
	GC5	1.37	260.00	11.50	0.94	1.43	37.14	4	10	2
	GC6	5.12	1065.00	98.50	0.08	2.03	217.00	5	5	1
	Median:	**2.33**	**457.75**	**36.81**	**0.87**	**2.55**	**115.18**	**5.00**	**7.50**	**2.00**

Panel E: GilStick

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*H-R²*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
GilStick	GS1	0.67	370.50	-22.46	0.78	3.90	107.00	7	4	6
	GS2	1.05	522.00	38.83	0.91	1.31	55.26	3	9	3
	GS3	0.18	57.00	-5.18	0.99	1.75	13.40	2	0	3
	GS4	0.78	199.00	12.25	0.95	1.75	29.30	4	8	2
	GS5	2.44	912.00	54.83	0.31	3.48	241.40	4	8	3
	GS7	1.95	667.50	64.55	0.67	3.43	165.80	2	4	1
	Median:	**0.92**	**446.25**	**25.54**	**0.85**	**2.59**	**81.13**	**3.50**	**6.00**	**3.00**

Note: Market B5 in the *Baseline* treatment, C5 in the *Carrot* treatment, and GS6 in the *GilStick* treatment are excluded because they contain subjects who participated in an earlier session of the experiment.

Table A4: Bubble measures for asset Y in Round 2

Panel A: Baseline

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*H-R^2*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
Baseline	B1	2.70	1118.50	93.21	0.58	1.97	150.60	5	12	0
	B2	2.16	1310.00	109.17	0.68	1.50	127.50	3	12	0
	B3	0.80	572.50	-46.46	0.46	1.23	62.53	5	1	7
	B4	2.87	560.00	11.83	0.52	2.63	148.53	8	7	5
	B6	1.54	626.00	24.33	0.39	0.97	47.87	2	7	2
	B7	104.24	3689.00	286.75	0.40	3.09	416.91	11	9	3
	B8	0.51	113.00	0.40	0.97	1.03	11.31	1	3	5
	Median:	**2.16**	**626.00**	**24.33**	**0.52**	**1.50**	**127.50**	**5.00**	**7.00**	**3.00**

Panel B: Carrot

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*H-R^2*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
Carrot	C1	0.43	271.50	-23.77	0.86	1.00	23.38	2	1	7
	C2	0.80	364.50	5.46	0.85	2.25	73.88	3	2	9
	C3	2.10	766.00	-14.42	0.01	2.00	169.54	5	7	4
	C4	8.56	667.00	21.25	0.04	2.03	140.66	5	7	5
	C6	2.26	422.50	-3.29	0.37	3.93	162.88	5	7	5
	C7	3.30	379.00	10.75	0.94	1.77	69.17	6	7	3
	C8	1.35	404.50	-15.05	0.79	1.00	45.40	4	4	5
	C9	1.71	310.00	6.25	0.78	0.50	22.17	1	2	3
	Median:	**1.91**	**391.75**	**1.08**	**0.78**	**1.89**	**71.52**	**4.50**	**5.50**	**5.00**

Panel C: Stick

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*H-R^2*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
Stick	S1	3.42	1565.00	134.73	0.42	1.94	248.77	2	6	2
	S2	4.51	841.00	71.00	0.16	1.40	109.18	4	7	3
	S3	0.59	453.50	-41.23	0.92	1.48	81.65	3	0	7
	S4	1.29	509.50	-41.77	0.76	1.13	38.60	2	2	8
	S5	0.37	34.50	-3.14	1.00	1.10	6.50	2	0	4
	S6	1.16	322.00	25.64	0.93	1.08	28.90	4	6	1
	S7	0.36	353.50	-32.14	0.91	1.09	40.26	4	0	9
	S8	8.23	755.50	48.55	0.01	1.46	109.77	3	4	3
	Median:	**1.23**	**481.50**	**11.25**	**0.84**	**1.26**	**60.95**	**3.00**	**3.00**	**3.50**

Table A4 cont.

Panel D: GilCarrot

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*H-R²*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
GilCarrot	GC1	1.03	277.00	-3.00	0.92	1.53	33.40	2	5	4
	GC2	1.51	448.50	33.38	0.84	2.23	76.55	6	10	2
	GC3	16.01	1654.50	57.96	0.86	1.75	260.08	10	6	5
	GC4	5.21	476.50	31.38	0.84	2.50	94.15	5	9	1
	GC5	2.46	302.00	14.83	0.81	1.43	37.83	6	8	4
	GC6	1.23	588.50	-34.21	0.39	1.73	76.57	9	3	7
	Median:	**1.99**	**462.50**	**23.10**	**0.84**	**1.74**	**76.56**	**6.00**	**7.00**	**4.00**

Panel E: GilStick

	Market	*Amp.*	*Tot. Disp.*	*Avg. Bias*	*H-R²*	*Turn.*	*Norm. Dev*	*Dur.*	*Boom Dur*	*Bust Dur*
GilStick	GS1	1.55	429.00	-31.17	0.49	3.10	145.65	5	3	7
	GS2	1.36	793.50	65.71	0.83	1.14	83.74	2	11	1
	GS3	0.32	60.00	-0.82	0.98	1.23	6.75	1	2	1
	GS4	2.76	595.00	53.40	0.37	1.85	66.93	3	6	2
	GS5	2.56	920.50	51.29	0.22	3.88	288.18	7	8	3
	GS7	1.96	924.50	75.38	0.65	2.66	109.57	2	11	1
	Median:	**1.75**	**694.25**	**52.35**	**0.57**	**2.25**	**96.66**	**2.50**	**7.00**	**1.50**

Note: Market B5 in the *Baseline* treatment, C5 in the *Carrot* treatment, and GS6 in the *GilStick* treatment are excluded because they contain subjects who participated in an earlier session of the experiment.

Appendix B: Participant instructions

The written instructions provided to participants are shown below (next page). These instructions relate to sessions where the market screen for asset X was displayed on the left-hand side of the screen – the instructions for sessions where Y was displayed on the left are qualitatively the same. Treatments vary according to how earnings are calculated, which is addressed in section 6 of the instructions – this section was unique to each treatment. Treatments also vary according the amount of relative performance feedback given, which is covered in Section 5.

1. General Instructions

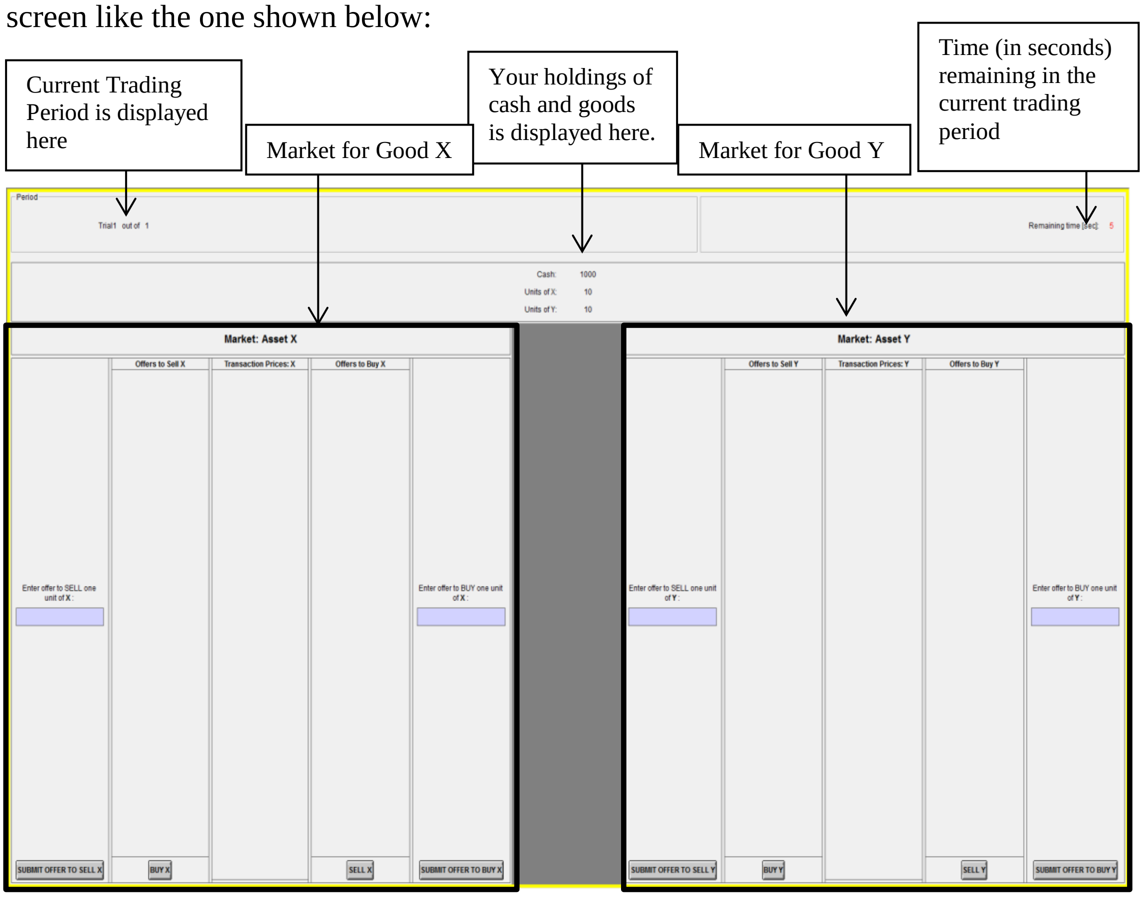
This is an experiment in the economics of market decision-making. The instructions are simple and if you follow them carefully and make good decisions, you may earn a considerable amount of money, which will be paid to you, in cash, at the end of the experiment. The experiment will consist of a sequence of trading periods in which you will have the opportunity to buy and sell in a market. All trading will be in terms of *francs*. The cash payment to you at the end of the experiment will be in Australian dollars, rounded up to the nearest 5 dollars. The conversion rate is ____ francs to 1 dollar.

The experiment will last no more than 2.5 hours, and will include up to 30 minutes of instructions and practice. Please do not speak with any other participants during the experiment. Please also remember to switch off your mobile phone. Failure to comply with these rules will result in your exclusion from the experiment and the forfeiture of all payments.

2. How to Use the Computerised Market

Before proceeding, we introduce the market interface that you will be using for the remainder of the experiment. Please note that any actions you take during this demonstration will **not** count towards your earnings or influence your position later in the experiment.

In this experiment, you will have the opportunity to buy and sell two different goods, called X and Y, in separate markets. In each trading period, you will see a computer screen like the one shown below:



Market: Good X

The market for good X is displayed on the **left-hand** side of your screen. All activity in relation to good X is shown and conducted here.

When you would like to offer to sell a unit of X, use the text area entitled “Enter offer to sell one unit of X” in the first column on the left. In that text area you can enter the price at which you are offering to sell a unit of X, and then select “Submit Offer To Sell X”. Please do so now. Type in a number in the appropriate space, and then click on the button labelled “Submit Offer To Sell X”.

You will notice that 8 numbers, one submitted by each participant in your market, now appear in the second column from the left, entitled “Offers to Sell X”. Your offer is listed in blue. Submitting a new offer will replace your previous offer.

The lowest offer-to-sell price will always be on the top of that list and will, by default, be selected. You can select a different offer by clicking on it. It will then be highlighted. If you select “Buy X”, the button at the bottom of this column, you will buy one unit of X for the currently selected sell price. Please purchase a unit now by selecting an offer and clicking the “Buy” button. Since each of you had offered to sell a unit of X and attempted to buy a unit of X, if all were successful, you all have the same number of units of X you started out with. This is because you bought one unit of X and sold one unit of X.

You may make an offer to buy a unit of X by selecting “Submit Offer to Buy X.” Please do so now. Type a number in the text area “Enter offer to buy one unit of X”, then press the button labelled “Submit Offer To Buy X”. All offers to buy X appear under the column entitled “Offers to Buy X”. The highest offer-to-buy price will always be on top of that list and will, by default, be selected. You can accept any of the offers-to-buy by selecting the offer and then clicking on the “Sell X” button. Please do so now.

The middle column of the market, labelled “Transaction Prices: X”, shows the prices at which X has been bought and sold in this period. The most recent transaction will be listed at the top.

Market: Good Y

The market for good Y is displayed on the **right-hand** side of your screen. All activity in relation to good Y is shown and conducted here. The layout of this market is identical to the market for X. The trading rules and procedures for posting and accepting offers to buy and sell Y are also the same.

To post an offer to sell a unit of Y, use the text area entitled “Enter offer to sell one unit of Y” and then select “Submit Offer To Sell Y”. Please do so now.

You can purchase a unit of Y by clicking the button “Buy Y” at the bottom of the column called “Offers to Sell Y”. Once again, the lowest offer-to-sell price is listed at the top and is selected by default. You can accept any offer by selecting it before clicking “Buy Y”. Please purchase a unit of Y now.

To make an offer to buy a unit of Y, type a number into the text area entitled “Enter offer to buy one unit of Y” and then select “Submit Offer To Buy Y”. Please do so now.

These offers are listed in the column “Offers to Buy Y”. To accept an offer, click “Sell Y” at the bottom of this column. The highest offer-to-buy price is selected by default. You can accept any of the offers by selecting it before clicking “Sell Y”. Please do so now.

The middle column of the market, labelled “Transaction Prices: Y”, shows the prices at which Y has been bought and sold in this period. The most recent transaction will be listed at the top.

Other features of both markets:

When you buy a unit of a good (i.e. X or Y), your Cash balance decreases by the price of the purchase. Any other existing offer to buy that good submitted by you is also cancelled. When you sell a unit of a good, your Cash balance increases by the price of the sale, and any other existing offer to sell that good submitted by you is cancelled.

You can participate in both markets at the same time.

If you make offers to buy in both markets at the same time, and say your offer to buy X is accepted first, then your offer to buy Y remains standing as long as you have enough Cash after the purchase of X to honour it, and vice versa. If you do not have enough Cash, then your offer in the second market is cancelled. Similarly, if you have a standing offer to buy in one market, and accept another trader’s sell offer in the second market, then your offer to buy in the first market is cancelled if your remaining Cash balance is less than the amount of your offer.

You will now have about 10 minutes to buy and sell in both markets. This is a practice period. **Your actions in the practice period do not count toward your earnings and do not influence your position later in the experiment.** The only goal of the practice period is to master the use of the interface. Please be sure that you have successfully submitted offers to buy and offers to sell in both markets. Also be sure that you have accepted buy and sell offers in both markets. If you have any questions, please raise your hand and the experimenter will come by and assist you.

3. Specific Instructions for this Experiment

This experiment consists of you and 7 other traders. At the beginning of the experiment, all traders will be endowed with a portfolio consisting of 5 units each of two types of goods, called ‘X’ and ‘Y’, and 1950 francs in Cash.

The experiment consists of 12 periods, each lasting 3 minutes. In each period, two separate markets will operate in which you may buy and/or sell units of good X and Y respectively. Both goods can be considered assets with lives of 12 periods, and your inventory of X and Y carries over from one trading period to the next. Note that your cash balance and inventory of assets cannot fall below zero.

At the end of each trading period, each unit of X pays an identical dividend, which is randomly determined by the computer. The possible dividend values and the associated likelihoods are shown below:

Asset: X

Dividend		**Likelihood**
10	→	$^1/_2$
30	→	$^1/_2$

Since each dividend is equally likely, the average dividend per period for X is 20 francs.

Each unit of Y also pays an identical dividend at the end of each period, randomly determined by the computer. The possible dividend values and the associated likelihoods are shown below:

Asset: Y

Dividend		**Likelihood**
0	→	$^4/_5$
100	→	$^1/_5$

The average dividend per period for asset Y is 20 francs ($0 \times {}^4/_5 + 100 \times {}^1/_5 = 20$).

The dividend draws for X and Y are independent across trading periods. This means that for both assets, the likelihood of a particular dividend in a period is not affected by the dividends in previous periods. In addition, the dividend draws for X and Y are independent of each other. This means that the occurrence of a particular dividend for X does not affect the likelihood of a particular dividend for Y, and vice versa.

Each unit of X and Y expires worthless after the final dividend is paid at the end of period 12.

4. Average Holding Value Table

You can use the table at the end of this document to help you make decisions. It calculates the average amount of dividends you will receive if you hold a unit of an asset in your inventory for the rest of the market, or equivalently, how much in dividends you give up, on average, when you sell a unit at any time. Each of the 5 columns of the table is described below:

1. *Ending Period*: indicates the last trading period of the market, period 12.
2. *Current Period*: indicates the period during which the average holding value is being calculated.
3. *Number of holding periods*: This is equivalent to the number of times a dividend can be received if a unit of an asset is held in your inventory from the current period to the end of the market.
4. *Average Dividend Per Period*: gives the average amount that the dividend will be in each period for each unit of the asset that is held in your inventory. The number in this column is 20. This is because the average dividend in each period for both X and Y is 20 francs. Since both types of assets have the same average dividend per period, you can use this table to determine the average holding value for both X and Y.
5. *Average Holding Value Per Unit of Inventory*: gives the expected total dividend for the remainder of the market for each unit of an asset that is held in your inventory for the rest of the market. That is, for each unit you hold in your inventory for the remainder of the market, you will receive on average the amount listed in column 5 in dividends. Equivalently, it tells you how much in future dividends you give up on average when you sell a unit in the current period. The number in column 5 is calculated by multiplying the numbers in columns 3 and 4.

<u>Example</u>: Suppose that there are 4 periods remaining. Since the dividend paid on a unit of X has a 50% chance of being 10 and a 50% chance of being 30, the dividend is in expectation 20 per period for each unit of X. Since the dividend paid on a unit of Y has an 80% chance of being 0 and a 20% chance of being 100, the dividend in expectation is also 20 per period for each unit of Y. If you hold a unit of X or Y for 4 periods, the total dividend paid on that unit over the 4 periods is in expectation $4 \times 20 = 80$.

5. Summary Screen

At the end of each trading period, a status report will appear on screen for 30 seconds. It displays the following information:

- Your Cash balance before the payment of dividends. This is calculated as:

  CASH BEFORE DIVIDENDS = BEGINNING OF PERIOD CASH

  + (PERIOD SALES REVENUE – PERIOD EXPENDITURE ON PURCHASES)

- The dividends paid by X and Y in this period.
- The number of units of X and Y in your inventory at the end of the period.
- The total amount of dividends you receive this period. This is calculated as:

  PERIOD TOTAL DIVIDEND = (END-OF-PERIOD UNITS OF X × DIVIDEND PER UNIT OF X FOR THE PERIOD)

  + (END-OF-PERIOD UNITS OF Y × DIVIDEND PER UNIT OF Y FOR THE PERIOD)

- Your Cash balance at the end of the period, which is calculated as follows:

  END-OF-PERIOD CASH = CASH BEFORE DIVIDENDS + PERIOD TOTAL DIVIDEND

- Your Account Total. This is equal to your end-of-period Cash plus the value of your holdings of X and Y.

  In periods 1 through to 11, your end-of-period holdings of X and Y are valued at their respective median traded price in that period. So, your Account Total at the end of period 1-11 is calculated as:

  ACCOUNT TOTAL = END-OF-PERIOD CASH

  + (END-OF-PERIOD UNITS OF X × MEDIAN TRADED PRICE OF X DURING PERIOD)

  + (END-OF-PERIOD UNITS OF Y × MEDIAN TRADED PRICE OF Y DURING PERIOD)

  Since all units of X and Y expire worthless after the final dividend payment at the end of period 12 (i.e. at the end of the market), your Account Total at the end of period 12 is equal to your end-of-period Cash balance:

  ACCOUNT TOTAL (end of period 12) = END-OF-PERIOD CASH

- The average Account Total in your market. **** ← this point does <u>not</u> appear in the *Baseline* treatment instructions, but does appear for all other treatments****
- Your rank out of the 8 participants in your market, based on your Account Total. A rank of 1 indicates the highest Account Total; a rank of 2 indicates the second-highest Account Total, and so on. **** ← this point <u>only</u> appears in the instructions for the *GilCarrot* and *GilStick* treatments****

After seeing the summary screen, press the “Continue” button to go to the next period. The next period will begin once everyone has pressed the “Continue” button, or once the 30 seconds have elapsed, whichever comes first.

6. Your Earnings

**** *Baseline* only: ****

Your earnings from this market will equal the balance of your Account Total at the end of the market. Remember that this is equal to your Cash balance at the end of the market.

Note that you do not have to calculate your earnings by yourself. The computer does all the work.

**** *Carrot* only: ****

Your earnings from this market will depend on your performance relative to the other traders in your market. Your performance is measured by comparing the balance of your Account Total at the end of the market (i.e. your final Cash balance) to the average end-of-market Account total/Cash balance in your market. Your payoff is calculated as follows:

$$Earnings_i = \begin{cases} 3000 & if \quad C_i < C^* \\ 3000 + 2(C_i - C^*) & if \quad C_i \geq C^* \end{cases}$$

where C_i is your final Account Total/Cash balance and C* is the average final Account total/Cash balance in your market.

Example: Suppose that the average end-of-market Cash balance in your market is 3500 francs. If your final Cash balance is say 3200 francs, you will earn 3000 francs. On the other hand, if your final Cash balance is say 4500 francs, you will earn 3000 + 2×(4500 – 3500) = 5000 francs.

Note that you do not have to calculate your earnings by yourself. The computer does all the work.

**** *Stick* only: ****

Your earnings from this market will depend on your performance relative to the other traders in your market. Your performance is measured by comparing the balance of your Account Total at the end of the market (i.e. your final Cash balance) to the average end-of-market Account total/Cash balance in your market. Your payoff is calculated as follows:

$$Earnings_i = \begin{cases} 0 & if \quad C_i < \frac{1}{2}C^* \\ 3000 & if \quad \frac{1}{2}C^* \leq C_i \leq C^* \\ 3000 + 2(C_i - C^*) & if \quad C_i > C^* \end{cases}$$

where C_i is your final Account Total/Cash balance and C* is the average final Account total/Cash balance in your market.

Example: Suppose that the average end-of-market Cash balance in your market is 3500 francs. If your final Cash balance is say 1000 francs, you will earn 0 francs from this market. If your final Cash balance is 3200 francs, you will earn 3000 francs. On the other hand, if your final Cash balance is say 4500 francs, you will earn 3000 + 2×(4500 – 3500) = 5000 francs.

Note that you do not have to calculate your earnings by yourself. The computer does all the work.

**** *GilCarrot* only: ****

Your earnings from this market will depend on your performance relative to the other traders in your market. The size of your payoff is determined by your rank at the **end** of the market (i.e. period 12), and is calculated as follows:

Rank	Your Earnings (francs)
1 ← largest final Account Total/Cash balance	10,000
2	4,000
3	4,000
4	4,000
5	4,000
6	4,000
7	4,000
8 ←smallest final Account Total/Cash balance	4,000

Note that you do not have to calculate your earnings by yourself. The computer does all the work.

**** *GilStick* only: ****

Your earnings from this market will depend on your performance relative to the other traders in your market. The size of your payoff is determined by your rank at the **end** of the market (i.e. period 12), and is calculated as follows:

Rank	Your Earnings (francs)
1 ← largest final Account Total/Cash balance	10,000
2	4,000
3	4,000
…	4,000
'Last' ←smallest final Account Total/Cash balance	0

Note that you do not have to calculate your earnings by yourself. The computer does all the work.

Average Holding Value Table

Ending Period	Current Period	Number of Holding Periods	×	Average Dividend Per Period	=	Average Holding Value Per Unit in Inventory
12	1	12		20		240
12	2	11		20		220
12	3	10		20		200
12	4	9		20		180
12	5	8		20		160
12	6	7		20		140
12	7	6		20		120
12	8	5		20		100
12	9	4		20		80
12	10	3		20		60
12	11	2		20		40
12	12	1		20		20