**MANAGEMENT SCIENCE**
Vol. 00, No. 0, Xxxxx 0000, pp. 000–000
ISSN 0025-1909 | EISSN 1526-5501 | 00 | 0000 | 0001

**INFORMS**
DOI 10.1287/xxxx.0000.0000


# Emotional Engagement and Trading Performance

Peter Bossaerts
Faculty of Economics, University of Cambridge, Cambridge CB3 9DD, United Kingdom plb32@cam.ac.uk

Felix Fattinger
Institute for Finance, Banking and Insurance, Vienna University of Economics and Business, 1020 Vienna, Austria, felix.fattinger@wu.ac.at

Kristian Rotaru
Faculty of Business and Economics, Monash University, Caulfield East, VIC 3145, Australia, Kristian.Rotaru@monash.edu

Kaitong Xu
Centre for Brain, Mind and Markets, The University of Melbourne, Parkville, VIC 3010, Australia, kaitongxu@gmail.com

Emotional involvement is known to be necessary but not sufficient for good decision-making in the face of uncertainty. It has been conjectured that emotional engagement in anticipation of risky outcomes constitutes "good" emotions. We introduce a new methodology to determine whether anticipatory emotional engagement is beneficial in the context of trading in financial markets. We focus on heart rate changes because they occur at a sufficiently high frequency to discern timing relative to events in the marketplace. After conservatively adjusting for multiple hypothesis testing, we find that participants whose heart rate changes anticipate their order submissions at inflated prices earn significantly more, while participants whose heart rate responds to their trades earn significantly less. By investigating co-integration between skin conductance response and the dynamics of individual portfolio values, we confirm the importance of emotional involvement in determining who makes or loses money.

*Key words*: Financial Markets, Emotions, Decision-Making, Asset Pricing, SCR, ECG

## 1. Introduction

Since a seminal article in *Science* by Antoine Bechara and Antonio Damasio in 1997 (Bechara et al. 1997), it has generally been understood that emotional engagement is a necessary component of improved financial decision-making. At the same time, the behavioral economics literature abounds with claims that emotions may generally be bad, i.e., detrimental to decision quality (Loewenstein 2000, Shefrin 2002, Hirshleifer 2015). Consequently, while emotional engagement may be necessary, it is certainly not sufficient. This raises the issue: what are "good" emotions? Here, we explore the issue in the context of trading in a financial market.

To understand the neuroscientists, let us quote Bechara and Damasio (2005), who wrote (in their abstract):

*Emerging neuroscience evidence suggests that sound and rational decision making, in fact, depends on prior accurate emotional processing.* (Bechara and Damasio 2005, p. 336)

The distinction between good and bad emotions therefore boils down to *anticipatory vs. reactive* emotions. "Good" emotions are related to premonition and presentiment. At the level of the autonomous nervous system, good emotions are engaged by the parasympathetic system, to reach homeostasis (balance) in an ever-changing environment, as in traditional feedback and feed-forward control systems in engineering (Billman 2020).[1] Emotions play a positive role if they are engaged in feed-forward control (Wu et al. 2021, Wang et al. 2022).

The evidence that anticipatory emotions are "good" emerged in a financial decision-making task where participants were asked to play a multi-armed bandit gambling game, the Iowa Gambling Task. Bechara et al. (1997) observed how healthy controls, after a number of trials, avoided the inferior options, which had both lower expected reward and higher reward variability. Patients with lesions in the orbitofrontal cortex, however, continued to choose the disadvantageous options, despite awareness of their inferior payoff structure. The study documents one major differentiation in the emotional engagement of the two groups of participants. Healthy controls displayed significantly higher activation of eccrine glands (i.e., transpiration) in anticipation of uncertainty resolution when choosing *bad* options. The emotional engagement was measured through Skin Conductance Response (SCR).[2] Lesion patients, in contrast, showed no emotional reaction to making bad choices.[3] Further discussion of this and related research, mainly within the neuroscience domain, can be found in Naqvi et al. (2006) and Schiebener and Brand (2015). Most recently, Hamelin and Bonelli (2022) studied anticipatory SCR in a market-like guessing game and found that participants with higher anticipatory SCR scores performed significantly better than those with lower scores.[4]

In finance, the role of anticipatory emotions has been discussed in a recent cross-disciplinary review by Duxbury et al. (2020) who note the emerging evidence on the role of anticipatory emotions. Using the Behavioral Investment Allocation Strategy task, a stylized simulation of a real trading scenario, Cantarella et al. (2018) show that anticipatory physiological responses (heart

[1] This contrasts with the sympathetic system ("fight or flight system"), which is reactive to outside events.

[2] SCR is also known as electrodermal response or "emotional sweating" (Boucsein 2012, Bach 2016).

[3] Bechara et al. (1997) compare lesion patients with individuals with uninjured brains. The fact that it uses a clinical population does not make it less relevant for finance. The use of lesion patients is purely methodological: lesions can get at causality. Our study only reveals correlation, even if the temporal aspects of correlation are often referred to as (Granger) "causality." Likewise, imaging studies such as Smith et al. (2014) only reveal anticipatory brain activation that correlates with beneficial trading; it may not have caused it. For a detailed motivation of the use of lesion patients in the study of emotions, see Adolphs (2007).

[4] Hamelin and Bonelli (2022) refer to SCR responses as "feelings," while simultaneously describing them as "autonomic body responses." As discussed later, we will make a distinction between autonomic bodily responses, which we refer to as *emotions*, and feelings.

rate and skin conductance response), "primary inducers," contribute to optimal financial decision-making under conditions of uncertainty, while "secondary inducers," i.e., the physiological responses that are triggered by experience, moderate this effect. The detrimental effect of emotional reactivity to past events corroborates earlier reports in laboratory settings that mimic investment behavior (Shiv et al. 2005).

As we shall corroborate with our own data here, SCR tracks an emotional change that shows up in electrodermal recordings after a long, and potentially variable, delay, and whose response effect is too long-lasting to accurately determine whether the emotional change among better decision-makers was anticipatory rather than post-event. The claim that anticipation is key, while reactive emotions are "bad" is not settled with the data from the Bechara-Damasio lesion patient study. A more recent, albeit underpowered, study did produce supportive evidence (Cantarella et al. 2018).

As an (important) aside, we shall follow the neuroscience tradition and distinguish between *emotions* and *feelings*. The former are objectively measurable changes in bodily states (SCR, for instance). Emotions reflect reactions in the autonomous nervous system. As described in Damasio (2004, p. 51), "[E]motions are *bioregulatory reactions* that aim at promoting, directly or indirectly, the sort of physiological states that secure not just survival but survival regulate into the range that we, conscious and thinking creatures, identify with *well-being*." Feelings, in contrast, refer to subjective experience (for details, see Kringelbach 2005). Feelings are often referred to as "affect." Damasio (2004, p. 52) explains them as follows: "[F]eelings are the mental representation of the physiologic changes that occur during an emotion. (...) [F]eeling an emotion also includes the mapping of changes that occur in the cognitive processing style, as well as the evocation of thoughts that are congruent with the feeling state."[5]

Emotions are therefore a part of the autonomic nervous system. It deserves emphasis that traditional description of this system may make one believe that autonomic neural responses only concern biophysical interaction with the environment. This ignores the fact that many of those responses happen without physical cause. An example of this is breathlessness, which occurs even without, e.g., an abrupt drop in temperature (Marlow et al. 2019). The view that autonomic neural responses fit solely within the traditional stimulus-response model of the central nervous system is obviously wrong; such responses are also, sometimes even exclusively, impacted by predictions. It is these types of autonomic neural responses that we are studying here. We focus on heart rate changes, among others in anticipation of submitting an order to a financial market. The biophysics

[5] Fermin et al. (2022) expresses the merits of the distinction as follows: "In order to simplify our treatment, we will use here a general notion found in the literature that emotions are unconscious arousal states linked with visceral and physiological processes under reflexive control, and feelings are conscious representations of emotions." This quote takes a stance on whether emotions are experienced consciously, however. We discuss later on that our participants may have been aware of their emotional involvement in trading.

of this action (usually requiring merely a click of a computer mouse) hardly demands any change in heart rate. Yet, as we shall see, heart rate changes can be substantial, and therefore, without biophysical cause.[6]

The term "autonomic" neural response may lead one to believe that no consciousness is involved. We do not take a stance on whether the physiological changes (emotions) we report on here (heart rate; skin conductance) is experienced consciously. Hence, we will not investigate "feelings" (affect), which is the conscious experience of autonomic neural responses. Emotional involvement may be entirely subconscious. Professional traders do exhibit emotions during trading (Lo and Repin 2002, Lo et al. 2005, Singh et al. 2022), but may not be aware of it. They may even falsely believe themselves to be "cool," in the sense of "immune to feelings."

We leave it to future work to clarify the conscious experience of emotions and its effect on financial decision-making. This also means that we will not address the question as to where the resulting feelings are located in the traditional arousal/valence space. Nor will we address to what extent the emotions affect specific aspects of what is generally considered "cognition," such as deliberately avoiding risks (risk aversion). Our analysis is limited to emotions, as were the aforementioned analyses of emotional engagement of professional traders (Lo and Repin 2002, Lo et al. 2005, Singh et al. 2022) and its impact on earnings (Kandasamy et al. 2016, Hamelin and Bonelli 2022).[7]

In the primate (including human) brain, *anterior insula* is considered the key "conduit" that facilitates translation of emotions into feelings. As a result, activation in anterior insula can be used to track to what extent the subject is aware of changes in "bodily state." At the same time, anterior insula is a cortical structure known to be crucially involved in tracking uncertainty in financial decision-making (Williams et al. 2021), and as such, we shall refer to neuroimaging evidence several times in this paper.[8] To gauge feelings one can ask questions. To trace emotions, one uses established biomarkers that track the changes in the emotional state, such as SCR recording. But as mentioned before, SCR changes are too slow and long-lasting to settle the issue whether "good" emotions are anticipatory while "bad" emotions are reactive. ECG (Electrocardiography), which allows us to track changes in heart rate, have a far lower response delay (of the order of 5 seconds; we will be more specific later on), and effects of increases in heart rate last far less, though response delay and duration depend on the health of one's cardiovascular system. ECG has been reported as a

[6] The presence of autonomic neural responses that have no underlying biophysical cause is important from a clinical point of view. Breathlessness, for instance, cannot solely be addressed by avoiding potential biophysical causes. In addition, intervention is to be based on neuropharmacology and cognitive therapy to impact, e.g., anxiety and the faulty predictions it causes (see, e.g., Marlow et al. 2019).

[7] As explained before, Hamelin and Bonelli (2022) uses the term "autonomic bodily responses" (emotions) and "feelings" interchangeably, though focus on the former (in particular, SCR).

[8] For more recent evidence on the brain regions involved in anticipatory emotion regulation and benefits of anticipatory emotions in social contexts, see for example Allaert et al. (2022).

reliable biomarker of emotional flexibility, in particular, in measuring resilience to stress (Perna et al. 2020).

Here, we set out to explore the relation between emotions and performance during trading in a controlled markets experiment designed around the canonical paradigm from Smith et al. (1988). Eight participants (as in the original experiment) trade one asset for cash on an online trading platform organized as an anonymous, continuous open book system, Flex-E-Markets.[9] The asset pays a periodic, random dividend, 15 times, after which it expires worthless. The setting is known to reliably generate mispricing among participants, whether novices or professionals (Noussair 2017), who have never been exposed to this particular setting before. While trading, we tracked participants' change in skin conductance, through an Electrodermal Activity Assessment (EDA) device, as well as heart rate, through an ECG device. To ensure that our measurements were well calibrated, we used a three-minute "Emotional Calibration Protocol" (ECP) before trading in ten sessions. We implemented a sub-class of ECP, soft nature sounds, which is commonly used for the calibration of somatic response, to induce emotional relaxation (Thoma et al. 2013, Peng-Li et al. 2022).[10] Calibration is needed, especially for ECG,[11] since heart rate variability (i.e., the ability of the heart to change rates locally in time) decreases with heart rate level: if a participant starts the experiment with a relatively high heart rate baseline, heart rate changes are harder to detect. To avoid this, we used ECP. To assess the impact of our calibration procedure, we ran additional sessions *without* ECP.

Few prior studies have reported a relationship between heart rate and performance in a trading context, either directly (through ECG) or indirectly (through tracking of neural activation in anterior insula). Lo and Repin (2002) were the first to discover that finance professionals' heart rate correlates with market volatility in a naturalistic setting (see also Lo et al. 2005, and more recently, Singh et al. 2022). Kandasamy et al. (2016) discovered that professionals who are better able to sense (feel) their own heart rate perform better (higher profits) and tend to be longer on the job. The link between heart rate and anterior insula was first established via a positive association between heart rate sensing and size of grey matter within the latter (Critchley et al. 2004). Smith et al. (2014) report how activation in a small subregion of the anterior insula allows one to predict who gets out of a price bubble in time. The subregion was first identified in a study that localized brain activation with risk prediction errors, which can be thought of as the driving force behind a

[9] Ad Hoc Markets, Salt Lake City, Utah; see `http://quantahm.com`.

[10] ECP was referred to as MMIP (Music Mood Induction Protocol) in earlier versions of the paper. To avoid confounding our protocol with priming, we chose to use the term ECP rather than MMIP.

[11] Tan et al. (2015) report a statistically significant reduction in heart rate of patients who were exposed to relaxing music compared to those who were in the 'no music' (silence) group.

GARCH model, in a dynamic card game (Preuschoff et al. 2008). Finally, Fenton-O'Creevy et al. (2012) demonstrated that professionals' heart rate variability decreases with volatility while the sensitivity (between heart rate variability and volatility) increases with trading experience. Heart rate variability correlates negatively with heart rate levels, so the study effectively showed that professionals' heart rate increases with volatility, corroborating the earlier findings of Lo et al. (2005). There has been a growing consensus in neuroscience that the heart rate in combination with anterior insula constitute the core of a body-brain network engaged in introspection; see, e.g., Harrison et al. (2021).

Our setting is challenging methodologically because trading decisions are made continuously. This of course makes our study more relevant from an ethology point of view, which decision neuroscientists have been insisting on recently (Yoo et al. 2021). Particularly in the context of trading, stimuli, decisions, feedback, introspection and interoception occur in a continuous rather than in a staggered way. However, the absence of a clear trial structure,[12] with transparent timing of stimuli, choice windows, and responses, challenges the statistical analysis of the resulting data. In addition, unlike in functional magnetic resonance (fMRI) recording of brain activation, the precise response function of the heart rate is unknown.[13] On top, individual differences in cardiovascular resilience may lead to heterogeneity in response delays. When dealing with anticipatory emotions, further complications arise in that the event anticipated with an emotional reaction (change in heart rate) occurs at a random time in the future, or may not occur at all.[14]

As such, we may know the range of heart rate response delays, but cannot know the exact timing of the peak response for each participant and each occurrence. Therefore, our paper introduces novel methodology to establish whether, in the context of continuous trading, anticipatory emotions are "good," while reactive emotions are "bad." A type of emotion is considered to be "good" when it increases performance (overall earnings); it is bad if it decreases performance. Our statistical methodology is based on Vector Autoregressions (VAR) and Granger causality tests. The latter measures which out of multiple time series (in our case, local heart rate estimates and a market-based variable of interest, e.g., deviations of an individual's trading prices relative to equilibrium valuations), can be used to predict future realizations.

[12] Contrast with, e.g., Hamelin and Bonelli (2022), where the "trading" experiment had a clear trial structure: 8 or more trials of 20 seconds each, during which the data did not change – except whether choice was correct, which was revealed after 10 seconds.

[13] The peak response of the fMRI signal occurs approximately 4s after onset of neural activation; the response function – referred to as hemodynamic response function – is well known, even if the specifics differ marginally from one brain region to another.

[14] In paradigms where external stimuli are simple and tightly controlled, one can estimate participant-specific response delays prior to the experiment. Our external stimuli are both complex (order and trade flow, dividend payments) and mostly beyond our control (participants themselves generate orders and trade at times that they determine). Moreover, anticipatory heart rate changes are likely to be generated by introspection and interoception, unrelated to particular perceptual stimuli.

In an attempt to correctly synchronize local heart rate estimates with market-based variables of interest, we have to consider several possible heart rate response delays. In addition, absent *a priori* theoretical predictions as to which of several market-based variables of interest would correlate with heart rate changes, we have to run multiple tests. The obvious danger is that we would identify significant effects merely due to data mining. To appropriately address this concern, our statistical analysis builds on strict and very conservative corrections for multiple hypothesis testing (also known as family-wise error rate corrections, abbreviated "FWE"). To verify that our corrections work, we re-run our analysis on a sample where we expected no relationship between Granger causality tests and performance. There, indeed, we only discover one single case for which the null hypothesis (no effect) is rejected – at the margin – confirming the efficacy of our strict, conservative correction for multiple hypothesis testing. In all our tests, we use a $p$-level of 0.05 after correction for multiple hypothesis testing (which effectively implies that the actual test relies on a $p$-level as low as 0.0001; see below for details).

An analogous methodological issue emerges in the analysis of SCR data. In the presence of a clear trial structure, as in, e.g., Hamelin and Bonelli (2022), standard statistical analysis can be resorted to, which is to first filter out low-frequency components in SCR. These low-frequency effects (random drifts) are thought to originate in causes that are not of scientific interest.[15] Crucially, our setting, however, induces drifts in decision-relevant variables, such as the amount of cash or value of securities holdings. Thus, by filtering out random drifts from SCR recordings, there is a danger that potential correlation with these decision-relevant variables is lost. While abstaining from filtering out low-frequency components of SCR time series may lead to spurious correlations; doing so naturally amounts to loss of information: the researcher no longer has access to the signals that could causally correlate with economic variables of interest. However, statistical techniques have been developed, in particular co-integration analysis, that can separate spurious from genuine correlations. These are summarized, among others, in Phillips (1998). We apply them here.

Altogether, the statistical analyses performed in this paper may appear unusual for a study of emotions; but our unusual, yet etiologically relevant, experimental setting demanded that we resort to different methods. Those methods have been shown to work in other contexts.[16]

The remainder of the paper is organized as follows. In Section 2, we first describe the experimental setup, which is essentially a replication of the original Smith et al. (1988) experiment, except that

[15] This could include drift in measurements because the SCR patches become detached, but also because of secular changes in the mood of the participant as events evolve.

[16] Examples of the use of co-integration in asset pricing are Bossaerts (1988) and Bansal et al. (2009). Examples of the use of co-integration in SCR analysis include Liu and Palumbo (2013), Hajifar et al. (2021), and Han et al. (2022). Co-integration has also been applied in the analysis of EEG (Electro-Encephalo-Gram) analysis: e.g., Kammerdiner and Pardalos (2010).

we tracked participants' emotions using SCR and ECG devices. We then provide a descriptive analysis of the results in Section 3, to confirm that our data are in line with prior findings as far as mispricing and earnings distributions are concerned. Moreover, these descriptive results should give the reader a feeling for the nature of the SCR and local heart rate time series and their potential relation with the dynamics of individual market-relevant variables, such as marked-to-market value of asset holdings or the extent of mispricing of individuals' trades. The characteristics of the time series set the stage for a detailed explanation of our statistical analysis. Section 4 presents the main results, after re-capitulating the core hypotheses. Additionally, it re-produces statistical results for the case if one (incorrectly) abstains from emotional calibration via ECP. As expected, we only find weakly significant effects in the presence of artificial noise in such uncalibrated data. Finally, Section 5 summarizes and concludes.

## 2. Experimental Setting

Our experimental setup faithfully replicated Smith et al. (1988). Participants traded one security called *stock* with a finite life of 15 periods in a continuous double auction market. The security paid a random dividend equal to 0, 0.25, 0.5, or 1.25 experimental dollars, each with equal probability, at the end of every period. Thus, the expected dividend each period was 0.50 experimental dollars. The dividend was distributed to participants in the form of experimental cash and paid for each unit of stock owned at the end of the period and before entering a new period. After paying the last dividend in period 15, the stock expired worthless.

Trade took place on an online trading platform called "Flex-E-Markets" (developed by Ad Hoc Markets, Salt Lake City, UT, USA).[17] Flex-E-Markets uses a continuous, anonymous open-book system operating exclusively with limit orders (see Panel A in Figure EC.1 in the Electronic Companion). Unlike field markets such as Euronext, however, order submission is checked for feasibility (shortsales were not allowed, and cash had to remain non-negative), and settlement is instantaneous. As a consequence, the Flex-E-Markets interface gives participants continuous access to (settled) holdings (including periodic dividend payments), holdings available for order submission (holdings net of submitted but yet to execute orders), the book of limit orders, and the entire trade history (including their own).

As in Smith et al. (1988), eight participants were admitted to each experimental session. No participant took part in more than one session. We designed the initial allocations of securities to induce trading incentives: half of the participants started with 20 units of stock and 100 experimental dollars, the remaining participants began with 12 units of stock and 160 experimental dollars. If participants were risk averse, those with a high initial endowment of stock would have wanted to

[17] See http://quantahm.com.

sell their holdings to those with a low initial allocation of stock. Note that all participants started with an endowment of cash and stock with the same fundamental value: the total expected value of individual endowments was 350 experimental dollars regardless of initial allocations. Participants' goal was to maximize their performance, measured by their end-of-session holdings of cash. As such, they were compensated for dividends received from holding stocks, plus initial cash and cash accumulation or reduction through trading.

Accounting was done in experimental dollars, and converted to Australian Dollars (AUD) at a pre-annnounced simple exchange rate, namely 1:10. Additionally, all participants were paid an extra 10 AUD as a sign-up reward. In total, participants could have expected to earn 45 AUD (35 AUD from trading, plus 10 AUD as sign-up reward). Because earnings could fluctuate as a result of both trading and uncertain dividend payments, we imposed a cap at 55 AUD and a floor at 25 AUD.

Since, in each period, the expected dividend on one unit of stock was \$0.50 while there were 15 periods in total, the sum of all expected dividends as of the beginning of the session was \$7.50. At the beginning of the second period, the stock would have paid its first dividend. Therefore, only 14 more payments remained, and given that each dividend was worth \$0.50 in expectation, expected future payments decreased by \$0.50. So, the "fundamental value" of the security dropped to \$7.00, and so forth.

The distribution of dividends and the evolution of changes in the fundamental value of the asset were known to all traders. Participants were given comprehensive instruction in both trading using Flex-E-Markets and the experimental setting. Participants were approached individually, to further assure comprehension, and to check for understanding of the evolution of the fundamental value across periods (albeit emphasizing that actual trade prices may deviate from fundamental value since they were not determined by the experimenter but by the interaction of the eight participants in the room). Instructions, available in the Electronic Companion, followed the language first introduced in Kirchler et al. (2012) with the goal of minimizing confusion. The summary of these instructions (also used in the experiment) is reproduced in Panel A in Figure 1.

In total we ran 16 sessions. All sessions were administered at Monash Business Behavioural Laboratory (Monash University). In ten of those, and in line with standard protocol (Sutherland et al. 1982, Zentner et al. 2008), we administered an Emotional Calibration Protocol (ECP) using soft nature sounds (i.e., sound of rippling water, see Thoma et al. 2013),[18] to calibrate mood with the aim of lowering the baseline heart rate. As explained in the introduction, this allows for higher

[18] Thoma et al. (2013) test the effect of three different conditions, i.e., relaxing music, sound of rippling water, and rest without acoustic stimulation, on cortisol response in a sample of 60 healthy participants. The study reports a significantly lower cortisol concentration in the condition of rippling water, compared to the other two conditions.

A. Summary of experimental instructions

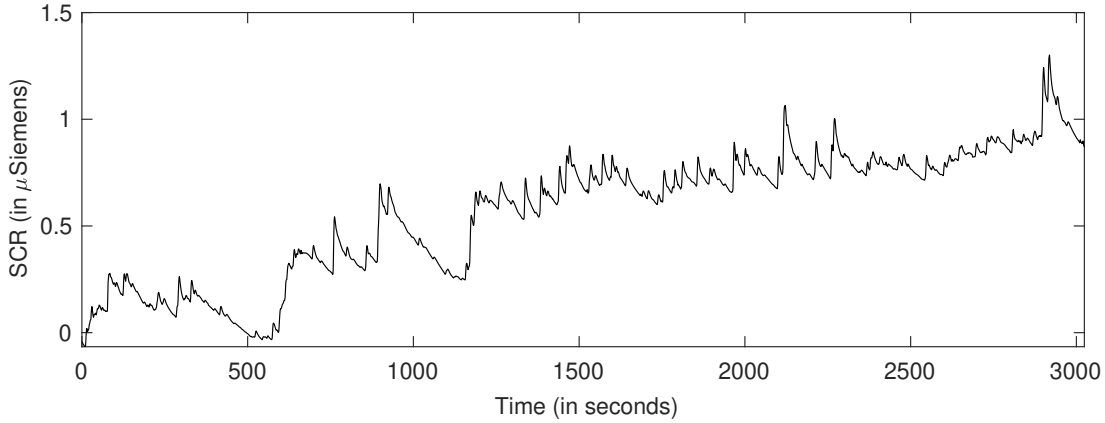
**Summary**

Your task will be to complete a trading game on our online trading platform. There will be a practice session, to provide ample opportunity to familiarise yourself with the online marketplace.

In the main trading session, which counts towards your earnings, you have the opportunity to trade a single, 15-period-lived security called "stock." The stock pays a random dividend at the end of each period. The expected dividend each period is 0.50 (experimental) dollar. You start with an endowment of cash and stock. You can sell stocks for cash, and/or use cash to buy stocks.

Your goal is to maximise your performance, measured by dividends received from holding stocks, plus cash accumulation through trading. You will be given an extra 10 (real) dollars as sign-up reward.

B. Exemplary SCR recording




C. Exemplary ECG recording

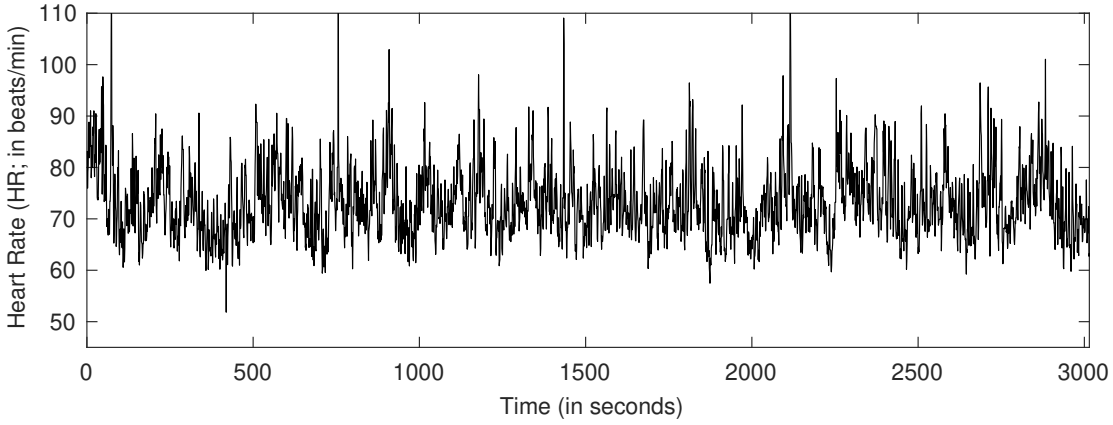



**Figure 1** A. Box with instructions summary. The same summary preceded the actual instructions used in the experiment (see e-companion). B. Evolution of SCR for one participant (participant 3 in session 3). C. Evolution of ECG recording for the same participant.

heart rate sensitivity to stimuli, in turn improving accuracy of local heart rate recordings and power of statistical tests. We ran six additional sessions without calibration. As we shall report, the noise from the non-calibrated sessions is sufficiently high for most tests to lose significance when bluntly pooling the two sets of sessions (ECP and non-ECP sessions). Formal model selection criteria, i.e., the Akaike Information Criterion (AIC) and Bayesian Information Criterion (BIC), mandate distinguishing between the two sets of sessions. Only one significant result emerges in the non-ECP sessions, but to increase accuracy of estimation, we nevertheless include them in the statistical analysis.[19]

During the entire experimental session, individual participant's physiological measures, electrodermal activity and heart electrical activity, were recorded through the ECG sensors and electrodermal activity assessment devices. We collected electrodermal activity (namely, SCR) data using a Shimmer3 GSR+ unit[20] with a sampling rate of 128Hz using Consensys multi sensor management software (see Panel B in Figure EC.1 in the Electronic Companion). To assure high quality of measurement, Velcro bands containing electrodes were placed on the volar surfaces of distal phalanges of the index and the middle fingers on each participant's non-dominant hand (see Panel B in Figure EC.1). Each participant's electrodermal activity was collected in kiloOhms ($k\Omega$) and then converted into MicroSiemens ($\mu$Siemens, a unit of electric conductance).[21] The relation is inverted: the higher is the MicroSiemens level, the lower is skin electric resistance due to transpiration, and hence the higher is the level of emotional arousal. SCR was measured at intervals of 1ms. We aggregated the raw data by taking averages over 1s. Panel B in Figure 1 displays the evolution of SCR for one participant (participant 3 in session 3).

The ECG data was collected using a Shimmer3 ECG unit[22] which records the pathway of electrical impulses through the heart muscle. In our experiment, Shimmer3 ECG recorded the participants' heart response (heart rate variability, HRV) with a sampling rate of 512Hz. Again, Consensys multi sensor management software was used to facilitate the recording. To ensure high quality of recording, the ECG was acquired using two biopotential electrodes connected to the device via biophysical leads and attached to the inner side of participants' right and left wrists (see Panel B in Figure EC.1 in the Electronic Companion). During the preparatory phase of the experiment, the research assistant helped the participants to apply abrasive skin preparation gel to the inner side of their wrists, to prepare the skin prior to attaching electrodes for biopotential

[19] Pooling allows one to more accurately estimate the time series dynamics that are common to the datasets even if these common dynamics are not germane to the issue at hand.

[20] See https://www.shimmersensing.com/products/shimmer3-wireless-gsr-sensor.

[21] 1 Siemens $= 1/\Omega$, so 1 $\mu$Siemens $= 1/M\Omega$.

[22] See https://www.shimmersensing.com/products/shimmer3-ecg-sensor.

measurements. The ECG provided recordings at 512Hz (one observation per approximately 2ms). We extracted instantaneous heartbeat rate estimates (HRV estimates) over periods of 0.5s using Matlab's *gqrs* function developed by Physionet (MIT, Cambridge, MA, USA). The typical output is as in Panel C in Figure 1 (participant 3 in session 3).

In total, 128 participants were recruited from Monash Business School, Monash University, through on-site advertisements. Eligible participants were healthy and willing people who were at least 18 years old. The recruited sample had an average age of 22.7 years (age range from 18 to 41 years, standard deviation of 3.1 years). Gender composition was 36% male and 64% female. Of all participants, 93% were enrolled in courses related to commerce (economics and business administration). The experiment was approved by the Human Research Ethics Committee at the University of Melbourne.[23] Rudimentary demographic information for each participant was collected, together with a written informed consent form, in accordance with ethics rules.

## 3. Descriptive Statistics

### 3.1. Trading and Earnings

Table 1 summarizes the parameters of each of the 16 sessions, numbered 1 to 12 and 21 to 24, respectively.[24] Panel A in Figure 2 displays, for all 16 sessions (color-coded), the evolution of the difference between transaction prices and the contemporaneous fundamental values, a standard way to display mispricing (Kirchler et al. 2012). Time (horizontal axis) is measured in seconds. Each dot represents a transaction. Inter-period intervals are clearly discernible because no trade occurred. In all sessions, trading was brisk, but decreased over time, which allowed us to reduce the length of periods from five minutes (300s) in the first period to two minutes (120s) in the last period. Typical mispricing patterns emerged, with substantial underpricing in early rounds, to substantial overpricing in later rounds, which eventually disappeared. In a few sessions, hardly any mispricing emerged.

The evolution of average mispricing is displayed in Panel B in Figure 2. For the *pooled* set of transaction prices, average mispricing is calculated as moving average over 320 "sequential" (in trading time) trades, which corresponds to the average number of trades per period (20 transactions) across all 16 trading sessions. Typically, overpricing emerges in period six, builds up continuously thereafter, before the bubble bursts eventually and mispricing converges rapidly to zero.

[23] Ethics ID: 1852054.1.

[24] Sessions 21 to 24 were run as part of responses to comments and queries from reviewers of this journal. The increased number of sessions added power. They also constitute a sort of out-of-sample "replication" test, to determine whether the results in the first version of the paper could be upheld.

| Session ID | Calibrated | Date (Yr-Mo-Day) | Session ID | Calibrated | Date (Yr-Mo-Day) |
|---|---|---|---|---|---|
| 1 | Yes | 2018-08-24 | 9 | Yes | 2019-03-06 |
| 2 | Yes | 2018-08-24 | 10 | No | 2019-03-29 |
| 3 | Yes | 2018-08-27 | 11 | Yes | 2019-03-29 |
| 4 | No | 2018-08-31 | 12 | No | 2019-03-29 |
| 5 | No | 2018-08-31 | 21 | Yes | 2021-05-20 |
| 6 | No | 2019-03-01 | 22 | Yes | 2021-05-26 |
| 7 | No | 2019-03-01 | 23 | Yes | 2021-05-27 |
| 8 | Yes | 2019-03-06 | 24 | Yes | 2021-06-15 |

**Table 1** Session IDs, indication of ECP calibration, and respective session dates.

Further investigation of mispricing reveals that there is no difference that could have been induced by ECP: for each period, the null that the difference in mispricing between non-calibrated and calibrated sessions equals zero can not be rejected (at $p = 0.05$, two-sided two-sample $t$-test with unequal variances).[25] In 11 out of 15 periods, mispricing is higher in non-calibrated sessions, but this difference is also insignificant ($p$-value 0.08, two-sided binomial test).

Figure 3 displays individual earnings (in Australian Dollars), stratified by session number. No calibration (ECP) was used in the sessions with earnings displayed as crosses. As is obvious from the plot, calibration did not create measurable effects on the cross-section of earnings.[26] The dashed line indicates average earnings across all sessions. Averages per session differed as a result of the randomness in the drawing of states.

### 3.2. SCR and ECG Recordings

As is evident from Panel B in Figure 1, the SCR time series are slow and persistent, consistent with our physiological understanding, as pointed out in the Introduction. Such time series are usually treated as unit root processes. Co-integration analysis allows us to investigate correlation with *non-stationary* holdings variables of interest, namely: (i) Asset Value Holdings (the current value of individual holdings of the bubble asset) and (ii) Cash Holdings (the current sum of initial cash, trading proceeds, and dividend payments). Note, co-integration analysis does not allow us to distinguish between *anticipatory* and *reactive* emotional responses, while this distinction is at the heart of our search for the "good" emotions (see below).

[25] $t$-statistics for the 15 periods ($p$-values in parentheses, two-sided test): 1.1341 (0.2802), 1.2887 (0.2224), 1.5828 (0.1370), 1.1591 (0.2686), 0.7693 (0.4560), 0.2353 (0.8180), -0.1441 (0.8879), 0.0402 (0.9687), 0.0424 (0.9672), 0.3056 (0.7675), 0.4855 (0.6375), -0.2121 (0.8351), 0.3029 (0.7669), -1.1389 (0.2745), -1.6364 (0.1347).

[26] The reader must not confuse ECP with priming (Andrade et al. 2016). The purpose of priming is to impact *choice* during a non-negligable time during the experiment. The purpose of ECP is to ensure calibration of psychophysical recordings (here: SCR and ECG) at the start of the experiment, and thus to enhance the accuracy of measurements.

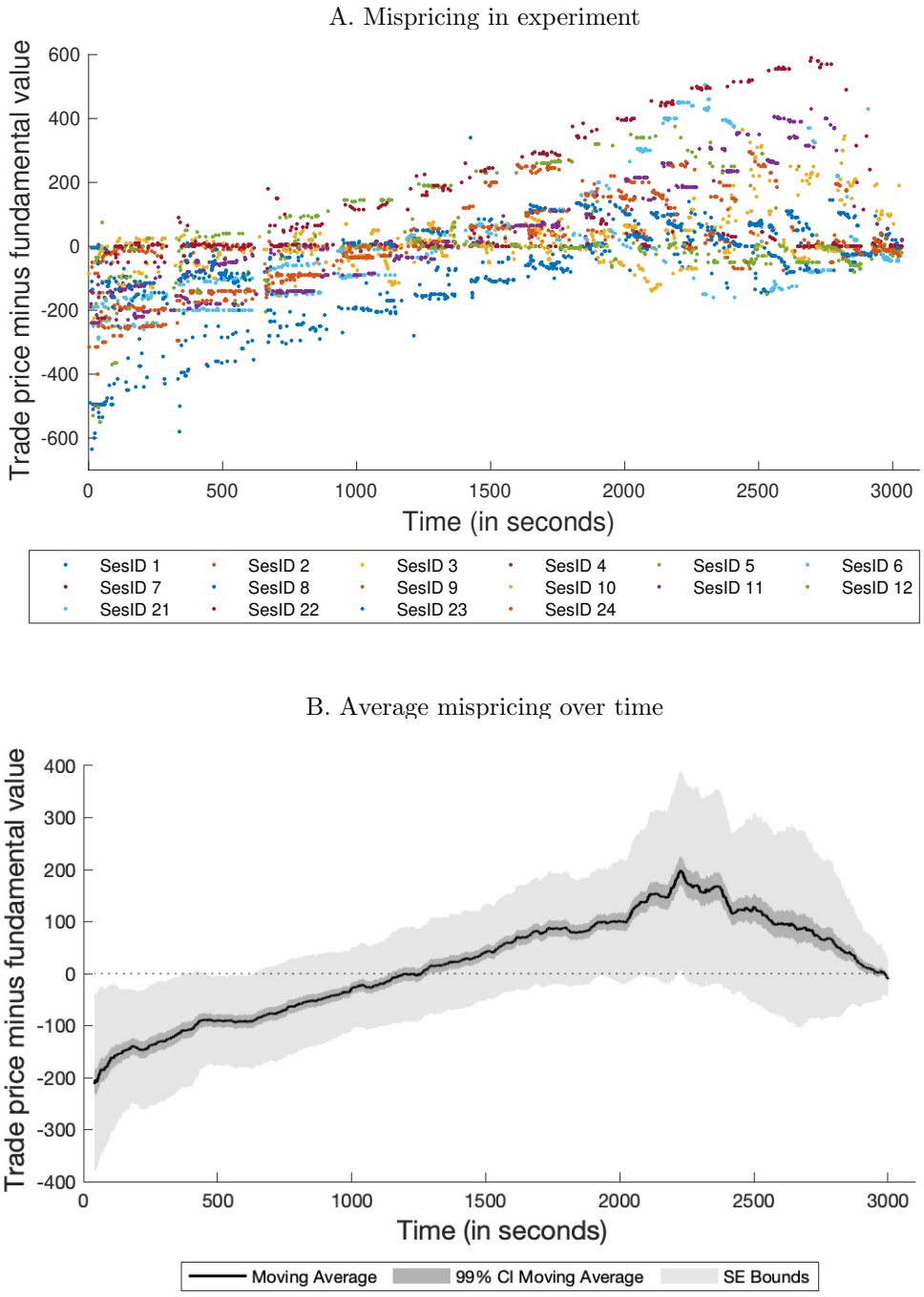



**Figure 2** A. Evolution of mispricing, in experimental cents (1/100th of one experimental dollar), for all sessions (see Table 1). B. Evolution of average mispricing over time. Mean, confidence interval (CI), and standard error (SE) bounds are calculated based on moving window of 320 trades, which corresponds to the average number of trades per period (20) across all 16 sessions.

However, SCR co-integration analysis *does* allow us to gauge to what extent there is any correlation between SCR and the above variables of interest. Figure 4 illustrates this: it displays barcharts

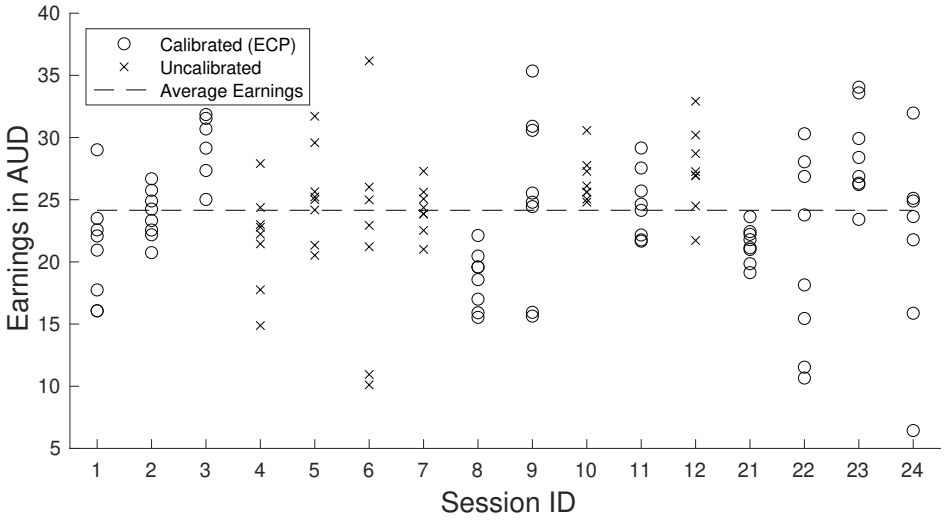



**Figure 3** Per-person earnings, in Australian Dollars (circles: sessions calibrated with ECP; crosses: uncalibrated sessions; 1 AUD = 10 experimental dollars).

of the fraction of participants whose SCR recordings exhibit significant ($p < 0.05$) co-integration with Asset Value Holdings (left plot) and Cash Holdings (right plot), stratified by earnings terciles. The barcharts show clearly how co-integration of SCR with Asset Value Holdings increases with earnings, while co-integration with Cash Holdings decreases with earnings. Both relations are significant at $p < 0.05$ (precise $p$-values of the standard chi-squared test for independence of fractions across categories are on top of each barchart).[27] The categorization of participants in earnings terciles may be considered arbitrary, but such evidence is only meant to be illustrative. The aim of our work is to *understand* the timing of emotions and market-relevant observations.

To this end, we have to turn to ECG. As the exemplar in Panel C in Figure 1 demonstrates, ECG generates time series that look stationary, with occasional but short episodes of elevated (estimated local) heart rates, lasting about 50s from beginning to end. With ECG, we have an opportunity to dissociate anticipatory and reactive heart rate responses to events in the marketplace and actions the individual trader is or plans taking.

The dissociation is valid, however, only if we manage to properly synchronize emotions behind ECG changes and the variables of interest. It is here that there is a problem, as pointed out before: ECG recordings are not synchronous with the true current emotional state. In particular, there is a response delay of about 5s or more before any emotional stimulation induces an increase in the heart rate; the heart rate change then reaches its peak after only 20s or even 30s (Hainsworth

[27] Figure 3 is based on data from all sessions, both calibrated and uncalibrated.

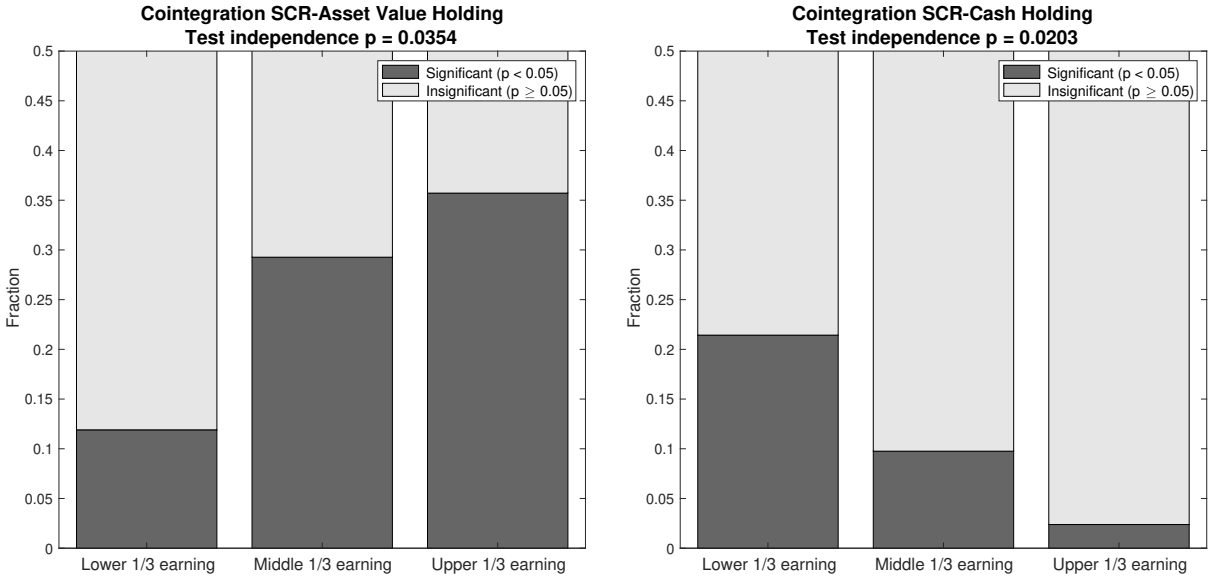



**Figure 4** Barcharts of the fraction of participants whose SCR recordings exhibit significant ($p < 0.05$) co-integration with Asset Value Holdings (left plot) or Cash Holdings (right plot), stratified by earnings terciles.

1995).[28] The actual response delay function depends not only on the nature of the stimulation but also on the physical fitness of the participant.

Consequently, we appeal to econometric techniques that allow for heterogeneous response dynamics. Specifically, we set out to estimate VARs (Vector Autoregressions) allowing for different levels of heart rate response delays (0s, 5s, 10s, ..., 30s). That is, we align the time series of local heart rates with market-relevant variables of interest leading the former by 0s (aligned in calendar time), 5s (putative heart rate response delay of 5s), up to 30s (putative heart rate response delay of 30s).[29]

Each of these alignments produces a separate VAR.[30] Each VAR produces two corresponding statistics of a Granger causality test,[31] one testing whether heart rate changes *anticipate* market-relevant observations ("FROM HR" test), and one testing whether market-relevant observations Granger-*cause* heart rate changes ("TO HR" test).

[28] A comprehensive, in-depth reference of cardiogram dynamics is Malik and Camm (2008).

[29] Formally, we let $H_{i,t-l}$ denote the locally estimated heart rate of participant $i$ at time $t$ minus $l$ (measured in seconds), where $l$ could be any value out of the set $\{0, 5, 10, ..., 30\}$, while $x_{i,t}$ denotes the value of the market-relevant variable under investigation. We run the following VAR:

$$\begin{bmatrix} H_{i,t-l} \\ x_{i,t} \end{bmatrix} = \Phi_0 + \Phi_1 \begin{bmatrix} H_{i,t-l-1} \\ x_{i,t-1} \end{bmatrix} + ... + \Phi_p \begin{bmatrix} H_{i,t-l-p} \\ x_{i,t-p} \end{bmatrix} + \epsilon_{i,t}.$$

The lag $p$ is determined as explained in the next footnote. We then compute two Granger causality statistics for each VAR (each participant $i$ and lag $l$), one testing whether heart rate $H_{i,t-l}$ Granger-causes $x_{i,t}$, denoted $G_{i,l}^{H\,\text{causes}}$, and the other testing the reverse Granger causality, denoted $G_{i,l}^{\text{causes H}}$.

[30] We set the order of the autoregression to the default setting of the *VAR.fit* method from Python's statsmodels library, i.e., $12(T/100)^{1/4}$, where $T$ denotes the number of observations.

[31] The output of Python's *VARResults.test_causality* method.

We then introduce the Granger-causality test statistics as regressors in a multivariate regression to explain participants' individual earnings. In particular, we run two separate regressions, one for the "FROM HR" Granger causality test statistics, and another one for the "TO HR" test statistics. In both cases we fit a generalized linear regression model with a linear link function and assuming Gaussian errors.[32] This procedure yields two tests:

**Test 1.** To determine *joint significance of the multiple heart rate response delays*, we use a likelihood ratio test, thereby correcting for common influence across the different heart rate response delays.[33]

**Test 2.** To determine *the influence at a specific heart rate response delay*, we use a standard $t$-test, but adjust the $p$-values for repeated testing using Holm-Bonferroni correction.[34]

Test 1 verifies the link between anticipatory or reactive heart rate changes when the econometrician wants to remain agnostic to the true heart rate response delay, or even wants to allow for heterogeneous response delays. In view of our discussion about the nature of heart rate response delays, Test 1 appears more appropriate. Nevertheless, we also report results for Test 2. *If* heart rate response delays are more homogeneous, Test 2 would lead us to identify the response function, in analogy with established practice in functional magnetic resonance imaging (see for example de Zwart et al. 2005).

Obviously, there is an identification problem for the Granger causality "FROM HR" tests with time series alignments where the putative response delay is 10s or larger: the heart rate changes could reflect, in some participants, stimuli that are as few as 5s in the past. When these heart rate changes are reactive, and one aligns the stimuli with heart rates 10s or more in the future, heart rate changes may spuriously Granger cause the stimuli (variables of interest). Because of this, we truncate time series alignments for the "FROM HR" tests at 15s (recall that we rely on 5s intervals).

[32] Formally, using the notation from the previous footnotes, and letting $E_i$ denote participant $i$'s earnings, we run the following generalized linear models with random intercepts:

$$E_i = \alpha_{s(i)} + \sum_{l\in\{0,5,10,15\}} \beta_l G_{i,l}^{H\ \text{causes}} + \eta_i,\ (i=1,2,...,I);$$

and

$$E_i = \alpha_{s(i)} + \sum_{l\in\{0,5,...,30\}} \beta_l G_{i,l}^{\text{causes H}} + \eta_i,\ (i=1,2,...,I).$$

Here, $s(i)$ determines the session number $s$ that $i \in \{1,2,...,I\}$ participated in, so we impute random effects at the session level.

[33] Referring to the notation in the previous footnote, we run a likelihood ratio test on the joint hypothesis that all $\beta_l$ are zero.

[34] We run separate $t$-tests for each $l$, on whether $\beta_l$ is significant, with $p$-values corrected for multiple hypothesis testing.

As mentioned before, we include six sessions without ECP calibration. To gauge the importance of calibration, we use model selection criteria (Akaike's Information Criterion AIC and the Bayesian Information Criterion BIC) to determine whether calibrated and non-calibrated sessions need to be distinguished in the multiple regressions of earnings onto Granger causality tests. Indeed, both AIC and BIC call for such a distinction, which we implement by adding an interaction dummy for ECP (calibrated) sessions (see Electronic Companion for detailed regression results). We first report results of the effect of Granger causality tests on earnings for the calibrated sessions only (i.e., focusing on the slope coefficients for the interaction effect). Second, to verify the effectiveness of our protocol that adjusts $p$-values for multiple hypothesis testing, we also report estimation results for all sessions pooled (i.e., based on the slope coefficients without accounting for potential interaction with calibration). For the latter case, if our $p$-value corrections are adequate, we expect to see a clear reduction in significant effects of Granger causality tests on earnings.

We study the following market-relevant variables of interest: (i) Individual Mispricing, (ii) Individual Holdings of the bubble asset, and (iii) Bid-Ask Spread. We define Individual Mispricing as the mispricing implied by the individual bids or asks submitted to the market, relative to the asset's current fundamental value. Moreover, the mispricing is *signed*, which means that it is higher when the order (bid, ask) is overpriced, while it is lower when the order is underpriced. Note this definition is supposed to solely differentiate an overpriced from an underpriced asset and is independent of whether one is selling or buying. This independence is important as the optimal strategy is generally unclear ex-ante: for instance, facing an overpriced asset, a trader could be worried about selling out too early or buying into a bubble too late. To reduce the chance of finding false positives, we limit our attention to the above three variables. We expect the first two market-relevant variables to be more powerful in predicting earnings because they entail direct involvement by the participant whose heart rate we tracked. While the participant controlled own order submission (variable (i)) or changes in holdings (variable (ii)), she may have had little impact on the aggregated bid-ask spread (variable (iii)).

Inspection of the evolution of these variables reveals that they are, with few exceptions, stationary. In the few cases where one of them (e.g., Holdings) exhibits non-stationarities (e.g., unit-root behavior), the VAR estimation produces insignificant Granger causality tests since the non-stationary variable at hand is used to predict a stationary heart rate. Hence, at times the Granger causality test may produce insignificant outcomes because no correction for non-stationarity is implemented, which adds noise to the explanatory variable (Granger causality tests) in the earnings regression. The associated cost is reduced power for discovering significant effects on earnings. The alternative would be to test and adjust for non-stationarities case-by-case, i.e., for each participant, session, and variable of interest separately. However, such fine-tuning of the analysis would clearly

raise the issue of over-fitting, and therefore increase the risk of inflated significance levels in the regressions of earnings on Granger causality tests.

Two more comments are in order. First, to adjust for differences in average earnings per session, we implement random-effects analysis in the regressions (one random intercept per session). Second, time is measured as the maximum of calendar time (measured in seconds) or business time (an uptick corresponds to a change in the market-relevant variable paired with the heart rate series).

For the reader to get a better feeling for the data, including heart rate response delays, and the nature of our ECG Granger causality tests, Figure 5 displays an exemplar of ECG time series and market-relevant data. Panel A in Figure 5 displays the calendar-time evolution of the locally estimated heart rate of participant 3 in session 3 (red) against the evolution of the Bid-Ask Spread in the marketplace (black). If one zooms in into the episode 600s-1100s (Panel B in Figure 5), it becomes clear that the Bid-Ask Spread (black) and the heart rate (red) are correlated. Once heart rate response delay is accounted for (Panel C in Figure 5, based on a putative heart rate response delay of 10s), the temporal correlation becomes visible: heart rate *anticipates* Bid-Ask Spread. The fact that we do find several cases of significant cross-autocorrelation of heart rate changes and the Bid-Ask Spread (as in Panel A in Figure 5) confirms earlier findings that professional traders' heart rate changes correlate with volatility (Lo and Repin 2002), as the Bid-Ask Spread generally correlates with market volatility.

## 4. Results

Before we show the main results, we re-cap the main hypotheses.

**Good Emotions.** When heart rate changes Granger-cause a market-relevant variable, the participant's earnings increase.

**Bad Emotions.** When heart rate changes are Granger-caused by a market-relevant variable, the participant's earnings decrease.

Figure 6 displays the main results.[35] The panel titles show the results from Test 1 with indication of FWE corrected $p$-values adjusted for testing across six panels. In each panel, the black dots depict the results from Test 2 and show the estimates of the slope coefficients when regressing individual earnings on the logarithm of Granger causality test statistics.[36] The test statistics account for different alignments of heart rate data, from assumed heart rate response delay of 0s up to 30s by increments of 5s. The dark gray area indicates the one standard-error interval. The light gray area (encompassing the dark gray interval) indicates the 95% confidence interval,

[35] The results are based on all ECG recordings save one participant in sessions 1, 2, and 6, and 2 participants in session 11. The recordings are missing because of technical issues with the ECG devices during the experiment.

[36] $10^{-4}$ is added to the test statistics to avoid numerical problems when taking logarithms of very small numbers.

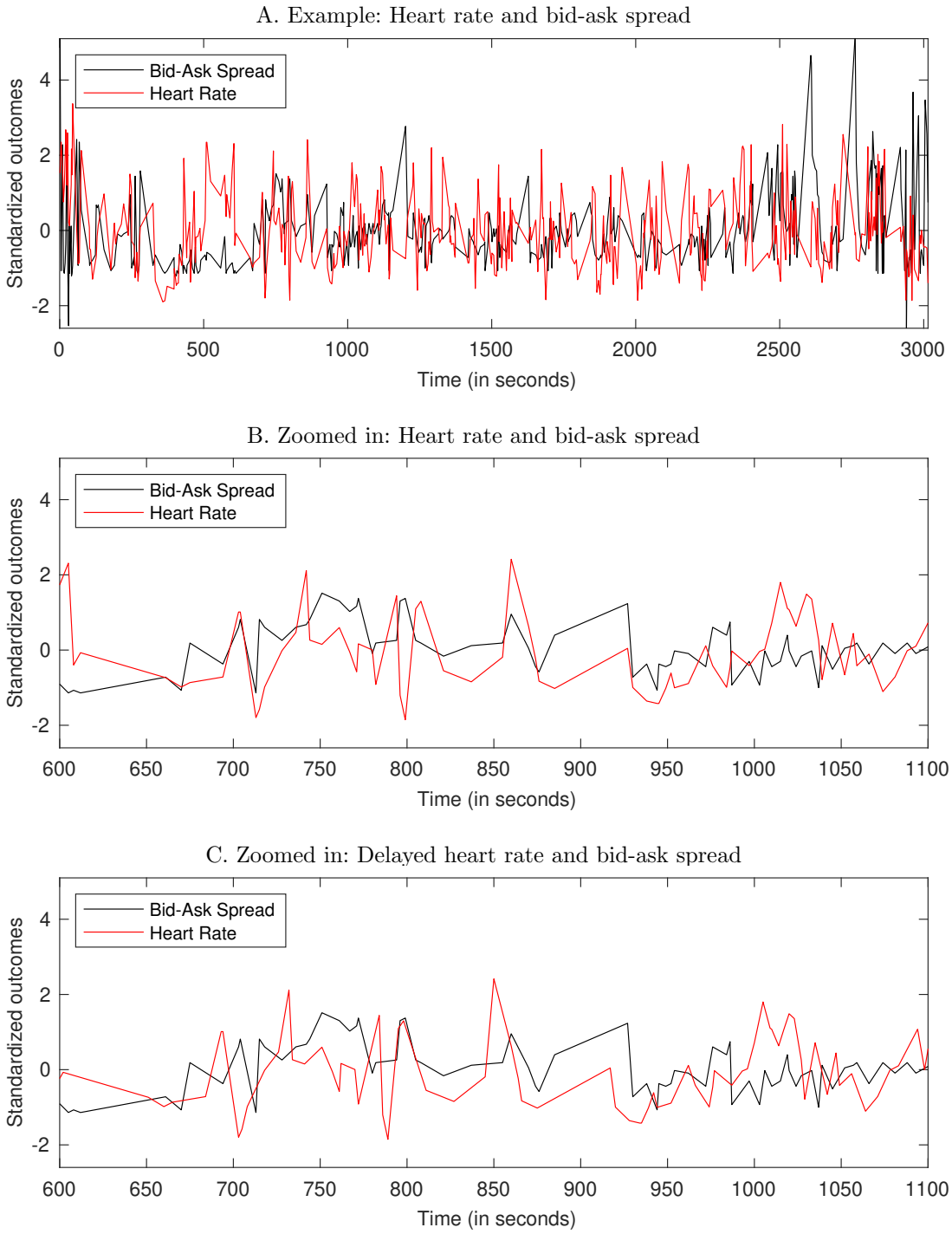



**Figure 5** A. Local heart rate of participant 3 in session 3 and concurrent Bid-Ask Spread assuming zero heart rate response delay (time series are aligned in calendar time). B. Zoomed-in snapshot of panel A. C. Same snapshot, but allowing for a heart rate response delay of 10s, which means that a Bid-Ask Spread observation at $t$ is paired with a local estimate of the heart rate at $t + 10$s. The emotional change detectable in the heart rate (after 10s) anticipates changes in the Bid-Ask Spread. The heart rate change is said to "Granger-cause" Bid-Ask Spread changes.

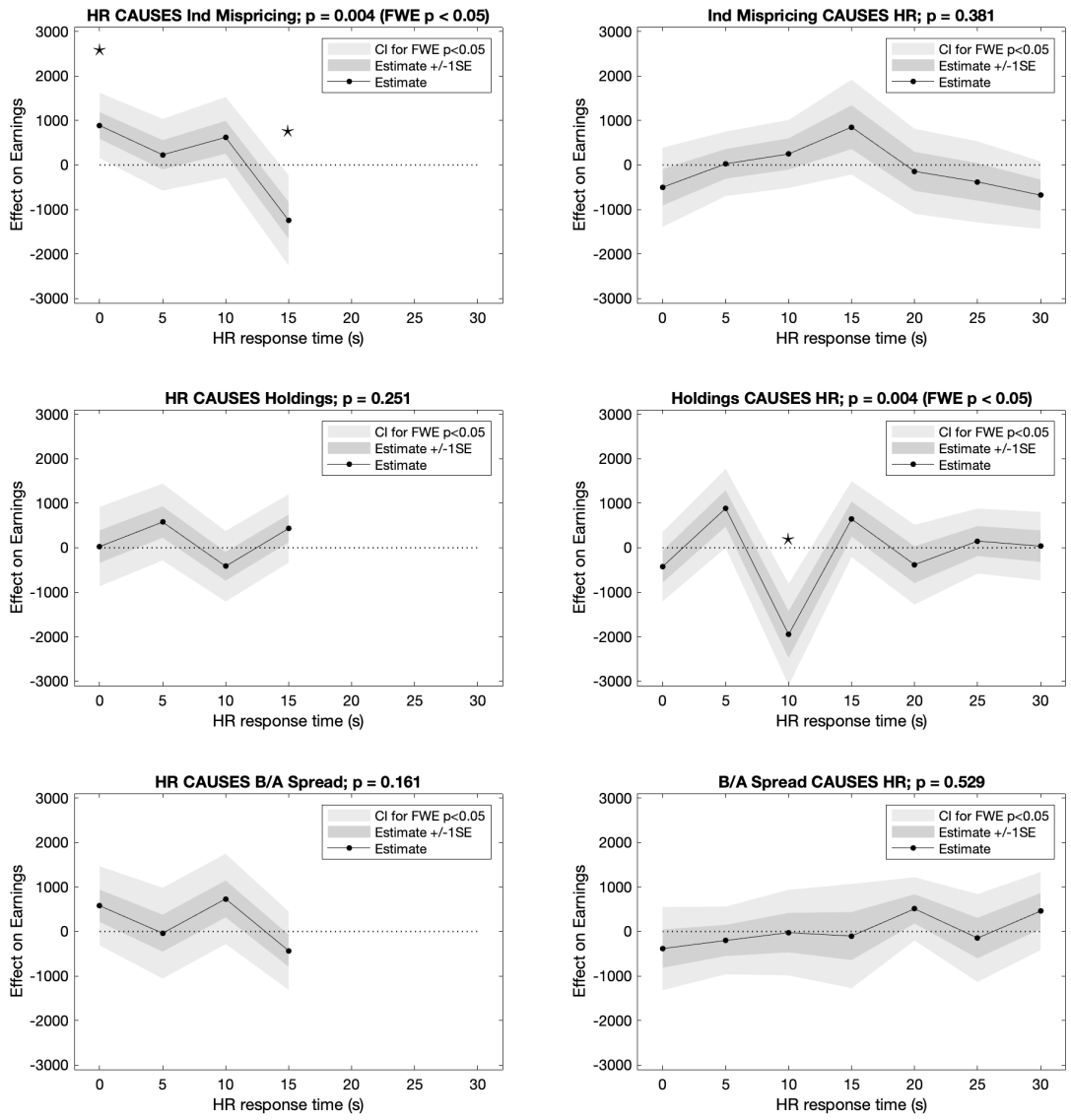



**Figure 6** The panel titles show the results from Test 1, with indication of FWE corrected $p$-values adjusted for testing across six panels. The black dots in each panel show the results from Test 2 and depict the estimates of the slope coefficients when regressing individual earnings on Granger causality test statistics for different alignments of heart rate data (from assumed heart rate response delay of 0s up to 30s by increments of 5s). The slope measures the effect on earnings (in experimental cents) of an increase in the Granger causality $F$-test statistic by 170% (natural logarithm scale). The dark gray area indicates the one standard-error interval. The light gray area (encompassing the dark gray interval) indicates the 95% confidence interval, FWE corrected for hypothesis testing across multiple alignment values. Stars indicate significance at $p < 0.05$, FWE corrected. HR response time equals the putative heart rate response delay, and hence, the offset in alignment of heart rate time series and time series for market-relevant variables (top row: Individual Mispricing; middle row: Holdings; bottom row: Bid-Ask Spread). Left panels test whether there are "Good Emotions," and which market-relevant variable they interact with. Right panels test whether there are "Bad Emotions," and which market-relevant variables they interact with.

which is FWE corrected for hypothesis testing across multiple alignment values. The left panels test "Good Emotions," i.e., whether Granger causality "FROM HR" increases earnings (positive effect on earnings). The right panels test "bad emotions," i.e., whether Granger causality "TO HR" causes decreased earnings (negative effect on earnings). Each panel row corresponds to one of the three market-relevant variables. Stars indicate slope coefficients of earnings onto Granger-causality statistics that produce significant effects at $p < 0.05$, adjusting for the fact that the figure effectively displays $4 \times 3 + 7 \times 3 = 33$ estimates and hence 33 tests.[37]

We find that, being agnostic as to the true heart rate response delay (Test 1), heart rate changes that *anticipate* Individual Mispricing events significantly increase earnings (FWE-corrected $p < 0.05$). Likewise, we find that heart rate changes *in response to* Holdings variations significantly decrease earnings (FWE-corrected $p < 0.05$). No other dynamic interactions between heart rate changes and market-relevant variables has any significant effect on earnings, correcting for testing this hypothesis six times (three market-relevant variables, two directions of Granger causality).

Assuming the heart rate response delay is unknown but fixed (Test 2), we find that concurrent heart rate changes that *anticipate* Individual Mispricing increase earnings (top left panel, effect estimate for heart rate response delay of 0s). By a putative heart rate response delay of 15s, this effect reverses sign. With such a high assumed response delay, however, the heart rate may in fact *follow* the Individual Mispricing series, and hence, in accordance with our definition, represent "Bad Emotions."[38]

For a heart rate response delay of 10s to the Holdings series, we find a significant negative effect on earnings when heart rate changes *follow* Holdings changes. No other effect reaches significance at $p$ equal to 5%, correcting for simultaneous testing of 33 hypotheses.

As pointed out above, throughout our analysis, we correct for multiple hypothesis testing, to ensure that our conclusions are not spurious because of data snooping. To verify that our FWE corrections are valid, we re-run all tests *without* distinguishing between calibrated (with ECP) and uncalibrated sessions. By doing so, we eliminate a possible interaction with calibration. The detailed regression results are collected in the Electronic Companion. As emphasized before, the

[37] Confidence intervals are based on separate Holmes-Bonferroni correction for each regression. Stars are shown only for results where the zero point is outside the corrected confidence interval *and* Test 1 produces a significant result at $p = 0.05$ (FWE-corrected across the $2 \times 3 = 6$ regressions). Correction of confidence intervals is based on toughest (smallest) $p$-values in the Holmes-Bonferroni procedure across four (left panels) and seven (right panels) regressors.

[38] A critical appraisal could qualify this reasoning as "having the cake and eat it too." Whether this is the case depends on how exactly the heart rate response delays should be modeled. In principle, and in analogy with fMRI analysis, one would want to model the response delay with a specific function (in fMRI analysis, this function is referred to as the "hemodynamic response function"). To do so is beyond the scope of this paper and would distract from its main goal. However, in fact, such a precise modeling exercise could constitute an important methodology study. In the absence of further guidance, one should focus on the results of Test 1, which are not affected by very conservative corrections for multiple hypothesis testing. Still, this biases one's interpretation towards concentrating on extreme effects only, disregarding marginal implications.

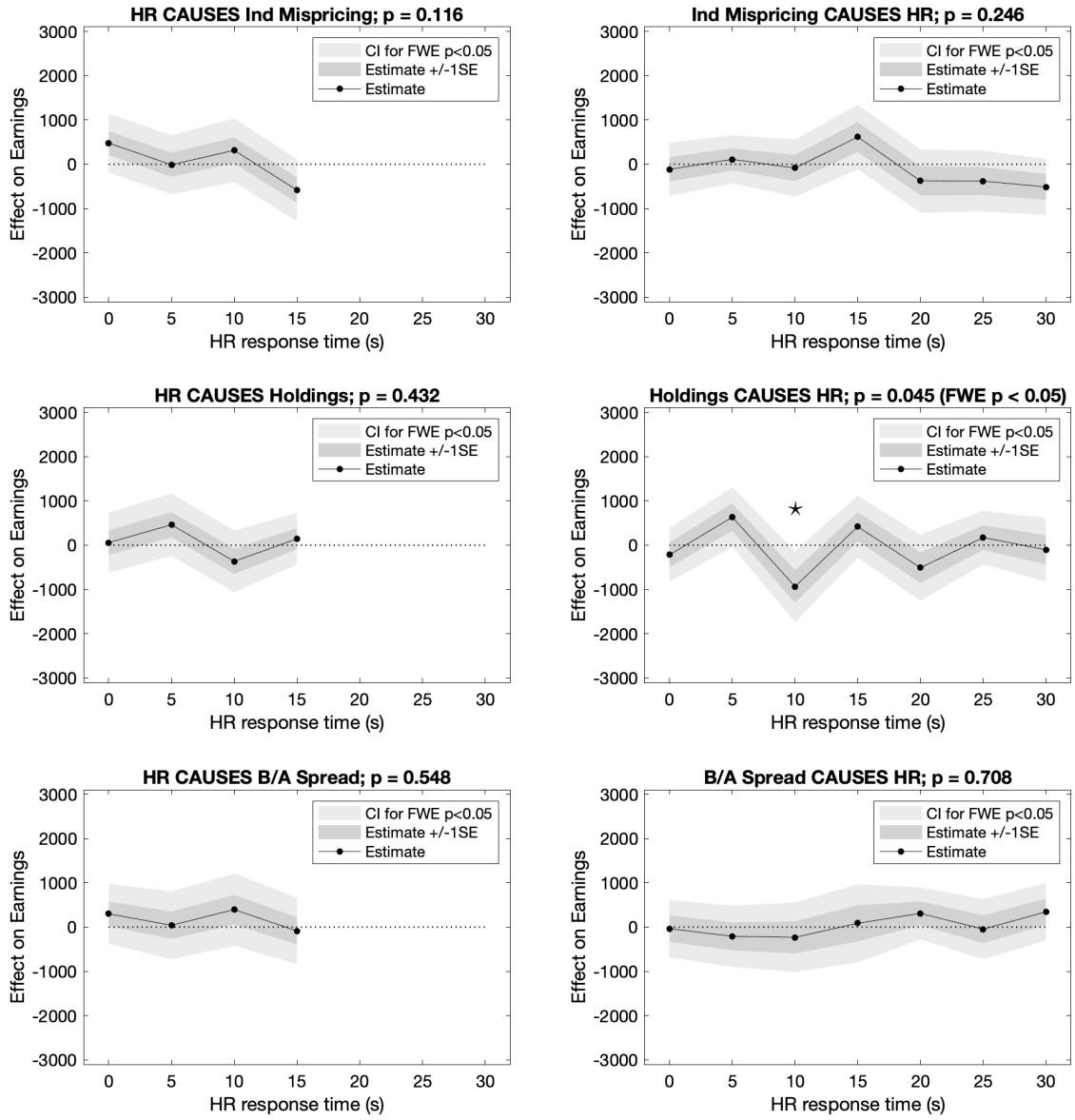



**Figure 7** The same results as in Figure 6 are shown, but for pooled data from calibrated and uncalibrated sessions.

application of formal information criteria (AIC and BIC) indicates that the uncalibrated sessions are distinct from the calibrated ones. Therefore, the two types of sessions must not be pooled, but distinguished by including interaction effect dummies. If we did not include such dummy variables, we expect the noise from uncalibrated sessions to overwhelm the signal from calibrated sessions, and to destroy (most) significance.

Indeed, this turns out to be true, as is evident from Figure 7. Relying on the pooled set of sessions, only one significant outcome emerges for either Test 1 (which tests with an unknown, potentially variable, heart rate response delay) or Test 2 (which tests for a particular heart rate response delay). Specifically, the negative effect on earnings when heart rate changes *follow* Holdings changes

remains (barely) significant according to Test 1 ($p = 0.045$). We conclude that our FWE correction works as intended.

## 5. Concluding Remarks

In the context of a standard, unstructured but ecologically relevant trading task, we set out to investigate the validity of the proposition that anticipatory heart rate changes are an example of "good" emotions, in the sense that they augment earnings, while reactive heart rate changes are generally detrimental to earnings. We focus on heart rate changes since their dynamics are far more short-lived and specific than, e.g., electrodermal effects (changes in skin conductance). Still, we also report significant relationships with earnings when skin conductance and market-relevant variables are co-integrated and identify a positive effect (i.e., increased earnings). This suggestive evidence already casts doubt on claims that emotional engagement is universally bad for financial performance: trading should not necessarily be done with "as cool a mind as possible."

Our study is challenging methodologically because the heart rate response function is not as specific as, e.g., the signal response function in functional magnetic resonance imaging (fMRI). Adapting techniques from fMRI analysis, we construct two types of tests. In Test 1, we remain agnostic as to response delays, and even allow the response delays to be heterogeneous across participants. The latter is important since it is generally acknowledged that there are major individual differences, among others related to individuals' physical condition. In Test 2, we try to identify the heart rate response function, assuming that there exists sufficient homogeneity to do so.

The two tests generate converging results, which are the following. (i) When heart rate responses predict the extent to which a participant's subsequent order exceeds the fundamental value, then her earnings increase. (ii) When the heart rate reacts to changes in the holdings of the risky security (as the result of trades), then earnings performance deteriorates. As such, we identify two clear-cut examples of "good" and "bad" emotions, while "good" emotions are tied to anticipation and "bad" emotions to reactive behavior. Importantly, these findings are not the result of data snooping. Throughout our analysis, we rigorously adjust for multiple hypothesis testing (of as much as 33 joint hypotheses). The validity of our (very conservative) adjustments is corroborated when we add noisy data from uncalibrated sessions. Formal model selection criteria call for separation of the calibrated from the uncalibrated sessions. Consistently, when adding uncalibrated data to our analysis, only one significant result survives, at the margin.

In sum, we find that electrodermal activity (skin conductance) and heart rate changes are engaged in financial decision-making, and ultimately impact earnings. In the case of heart rate changes, we can assess timing more precisely. There, we show that anticipatory heart rate changes increase earnings, while reactive heart rate changes in response to earnings-relevant events decrease earnings.

To make sense of these results, we propose that emotions are to be considered an integral part of cognition. In this view, emotions *reveal* cognition. Indeed, changes in heart rate could be an indication that the participant is "thinking." When thinking is in anticipation of actions, it is beneficial. When it is reactive to events, it may be detrimental.

We deliberately used the verb "may," since one cannot strictly map anticipatory and reactive emotions into "good" and "bad" categories. There exists a large literature linking emotion-related brain activation in response to outcomes as beneficial, since they constitute key inputs to a type of reinforcement learning that is known to eventually improve behavior (and incidentally, has inspired the core algorithms of machine learning). A review can be found in Etkin et al. (2015). The brain region mentioned before, anterior insula, features prominently in this framework. It is fair to say that, by now, it is generally accepted that emotional engagement is necessary (causal) for beneficial reinforcement learning. This is not to say that emotional engagement is universally positive: maladaptive choice often has its origins in excessive reactive emotions (see, e.g., Rutledge et al. 2017).

Claims have been made about the interaction between emotions and decision-making, but rarely did empirical tests verify whether emotions actually play the role the theory assigns to them. One example is loss aversion, which is explained with explicit reference to emotions, i.e., losses loom larger than gains. When the hypothesized link with emotions was put to an explicit test in a meta-analysis, it could not be found (Li et al. 2017). We propose that the putative links between emotions and decision-making be made explicit in future tests. This paper provides the methods to do so. Our approach builds on the hypothesis that there exist emotions that are beneficial to (financial) choice, while others are detrimental, and that the difference depends on timing.

## Acknowledgments

The authors gratefully acknowledge detailed comments and suggestions from the Editor, the Associate Editor, and the referees. The authors thank Fabienne Picard, Daniel Bennett, Pablo Franco, Lee Lawrence, and Odkhishig Ganbold for discussing issues that helped improve the manuscript. Financial support from ARC grant DP 180102284 (Bossaerts) and from a R@MAP Chair at the University of Melbourne (Bossaerts) is gratefully acknowledged. The authors also thank the Monash Business School for funds associated with the Monash Business Behavioural Laboratory facility and equipment used in this study. Some of the results reported here formed part of Katherine Xu's Honours Thesis at the University of Melbourne.

## References

Adolphs R (2007) Investigating human emotion with lesions and intracranial recording. *Handbook of Emotion Elicitation and Assessment* 426–439.

Allaert J, De Raedt R, Sanchez-Lopez A, Baeken C, Vanderhasselt MA (2022) Mind the social feedback: effects of tdcs applied to the left dlpfc on psychophysiological responses during the anticipation and reception of social evaluations. *Social Cognitive and Affective Neuroscience* 17(1):131–141.

Andrade EB, Odean T, Lin S (2016) Bubbling with excitement: an experiment. *Review of Finance* 20(2):447–466.

Bach DR (2016) Skin conductance measures in neuroeconomic research. *Neuroeconomics*, 345–357 (Springer).

Bansal R, Dittmar R, Kiku D (2009) Cointegration and consumption risks in asset returns. *The Review of Financial Studies* 22(3):1343–1375.

Bechara A, Damasio AR (2005) The somatic marker hypothesis: A neural theory of economic decision. *Games and Economic Behavior* 52(2):336–372.

Bechara A, Damasio H, Tranel D, Damasio AR (1997) Deciding advantageously before knowing the advantageous strategy. *Science* 275(5304):1293–1295.

Billman GE (2020) Homeostasis: the underappreciated and far too often ignored central organizing principle of physiology. *Frontiers in Physiology* 200.

Bossaerts P (1988) Common nonstationary components of asset prices. *Journal of Economic Dynamics and Control* 12(2-3):347–364.

Boucsein W (2012) *Electrodermal activity* (Springer Science & Business Media).

Cantarella S, Hillenbrand C, Aldridge-Waddon L, Puzzo I (2018) Preliminary evidence on the somatic marker hypothesis applied to investment choices. *Journal of Neuroscience, Psychology, and Economics* 11(4):228.

Critchley HD, Wiens S, Rotshtein P, Öhman A, Dolan RJ (2004) Neural systems supporting interoceptive awareness. *Nature Neuroscience* 7(2):189–195.

Damasio AR (2004) Emotions and feelings. *Feelings and emotions: The Amsterdam symposium*, 49–57 (Cambridge University Press Cambridge).

de Zwart JA, Silva AC, van Gelderen P, Kellman P, Fukunaga M, Chu R, Koretsky AP, Frank JA, Duyn JH (2005) Temporal dynamics of the bold fmri impulse response. *Neuroimage* 24(3):667–677.

Duxbury D, Gärling T, Gamble A, Klass V (2020) How emotions influence behavior in financial markets: a conceptual analysis and emotion-based account of buy-sell preferences. *The European Journal of Finance* 26(14):1417–1438.

Etkin A, Büchel C, Gross JJ (2015) The neural bases of emotion regulation. *Nature Reviews Neuroscience* 16(11):693–700.

Fenton-O'Creevy M, Lins JT, Vohra S, Richards DW, Davies G, Schaaff K (2012) Emotion regulation and trader expertise: Heart rate variability on the trading floor. *Journal of Neuroscience, Psychology, and Economics* 5(4):227.

Fermin AS, Friston K, Yamawaki S (2022) An insula hierarchical network architecture for active interoceptive inference. *Royal Society Open Science* 9(6):220226.

Hainsworth R (1995) *Heart Rate Variability*, chapter The control and physiological importance of heart rate, 3–19 (Armonk, NY: Futura Publishing Company, Inc.).

Hajifar S, Sun H, Megahed FM, Jones-Farmer LA, Rashedi E, Cavuoto LA (2021) A forecasting framework for predicting perceived fatigue: Using time series methods to forecast ratings of perceived exertion with features from wearable sensors. *Applied Ergonomics* 90:103262.

Hamelin N, Bonelli MI (2022) Traders' anticipatory feelings and traders' profitability: An exploratory study. *Journal of Behavioral and Experimental Finance* 100743.

Han Z, Kang J, Meng Q (2022) A time series analysis model of the relationship between psychoacoustic parameters of urban soundscape spatial sequences and emotional changes. *The Journal of the Acoustical Society of America* 152(4):2022–2037.

Harrison OK, Köchli L, Marino S, Luechinger R, Hennel F, Brand K, Hess AJ, Frässle S, Iglesias S, Vinckier F, et al. (2021) Interoception of breathing and its relationship with anxiety. *Neuron* 109(24):4080–4093.

Hirshleifer D (2015) Behavioral finance. *Annual Review of Financial Economics* 7:133–159.

Kammerdiner AR, Pardalos PM (2010) Analysis of multichannel eeg recordings based on generalized phase synchronization and cointegrated var. *Computational Neuroscience*, 317–339 (Springer).

Kandasamy N, Garfinkel SN, Page L, Hardy B, Critchley HD, Gurnell M, Coates JM (2016) Interoceptive ability predicts survival on a london trading floor. *Scientific Reports* 6(1):1–7.

Kirchler M, Huber J, Stöckl T (2012) Thar she bursts: Reducing confusion reduces bubbles. *American Economic Review* 102(2):865–83.

Kringelbach ML (2005) The human orbitofrontal cortex: linking reward to hedonic experience. *Nature Reviews Neuroscience* 6(9):691–702.

Li R, Smith DV, Clithero JA, Venkatraman V, Carter RM, Huettel SA (2017) Reason's enemy is not emotion: engagement of cognitive control networks explains biases in gain/loss framing. *Journal of Neuroscience* 37(13):3588–3598.

Liu S, Palumbo R (2013) Dynamic modeling of interpersonal electrodermal activity using cointegration methodology. *Psychophysiology*, volume 50, S19–S19 (Wiley-Blackwell, HOBOKEN, NJ USA).

Lo AW, Repin DV (2002) The psychophysiology of real-time financial risk processing. *Journal of Cognitive Neuroscience* 14(3):323–339.

Lo AW, Repin DV, Steenbarger BN (2005) Fear and greed in financial markets: A clinical study of day-traders. *American Economic Review* 95(2):352–359.

Loewenstein G (2000) Emotions in economic theory and economic behavior. *American Economic Review* 90(2):426–432.

Malik M, Camm AJ (2008) *Dynamic electrocardiography* (John Wiley & Sons).

Marlow LL, Faull OK, Finnegan SL, Pattinson KT (2019) Breathlessness and the brain: the role of expectation. *Current Opinion in Supportive and Palliative Care* 13(3):200.

Naqvi N, Shiv B, Bechara A (2006) The role of emotion in decision making: A cognitive neuroscience perspective. *Current Directions in Psychological Science* 15(5):260–264.

Noussair CN (2017) The discovery of bubbles and crashes in experimental asset markets and the contribution of vernon smith to experimental finance. *Southern Economic Journal* 83(3):644–648.

Peng-Li D, Andersen T, Finlayson G, Byrne DV, Wang QJ (2022) The impact of environmental sounds on food reward. *Physiology & Behavior* 245:113689.

Perna G, Riva A, Defillo A, Sangiorgio E, Nobile M, Caldirola D (2020) Heart rate variability: Can it serve as a marker of mental health resilience?: Special section on "translational and neuroscience studies in affective disorders" section editor, maria nobile md, phd. *Journal of Affective Disorders* 263:754–761.

Phillips PC (1998) New tools for understanding spurious regressions. *Econometrica* 1299–1325.

Preuschoff K, Quartz SR, Bossaerts P (2008) Human insula activation reflects risk prediction errors as well as risk. *Journal of Neuroscience* 28(11):2745–2752.

Rutledge RB, Moutoussis M, Smittenaar P, Zeidman P, Taylor T, Hrynkiewicz L, Lam J, Skandali N, Siegel JZ, Ousdal OT, et al. (2017) Association of neural and emotional impacts of reward prediction errors with major depression. *JAMA Psychiatry* 74(8):790–797.

Schiebener J, Brand M (2015) Decision making under objective risk conditions–a review of cognitive and emotional correlates, strategies, feedback processing, and external influences. *Neuropsychology Review* 25(2):171–198.

Shefrin H (2002) *Beyond greed and fear: Understanding behavioral finance and the psychology of investing* (Oxford University Press on Demand).

Shiv B, Loewenstein G, Bechara A, Damasio H, Damasio AR (2005) Investment behavior and the negative side of emotion. *Psychological Science* 16(6):435–439.

Singh M, Xu Q, Wang SJ, Hong T, Ghassemi MM, Lo AW (2022) Real-time extended psychophysiological analysis of financial risk processing. *PloS ONE* 17(7):e0269752.

Smith A, Lohrenz T, King J, Montague PR, Camerer CF (2014) Irrational exuberance and neural crash warning signals during endogenous experimental market bubbles. *Proceedings of the National Academy of Sciences* 111(29):10503–10508.

Smith VL, Suchanek GL, Williams AW (1988) Bubbles, crashes, and endogenous expectations in experimental spot asset markets. *Econometrica: Journal of the Econometric Society* 1119–1151.

Sutherland G, Newman B, Rachman S (1982) Experimental investigations of the relations between mood and intrusive unwanted cognitions. *British Journal of Medical Psychology* 55(2):127–138.

Tan YZ, Ozdemir S, Temiz A, Celik F (2015) The effect of relaxing music on heart rate and heart rate variability during ecg gated-myocardial perfusion scintigraphy. *Complementary Therapies in Clinical Practice* 21(2):137–140.

Thoma MV, La Marca R, Brönnimann R, Finkel L, Ehlert U, Nater UM (2013) The effect of music on the human stress response. *PloS ONE* 8(8):e70156.

Wang X, Wang P, Lu J, Zhou J, Li G, Garelik S (2022) Episodic future thinking and anticipatory emotions: Effects on delay discounting and preventive behaviors during covid-19. *Applied Psychology: Health and Well-Being* .

Williams TB, Burke CJ, Nebe S, Preuschoff K, Fehr E, Tobler PN (2021) Testing models at the neural level reveals how the brain computes subjective value. *Proceedings of the National Academy of Sciences* 118(43).

Wu S, Sun S, Camilleri JA, Eickhoff SB, Yu R (2021) Better the devil you know than the devil you don't: Neural processing of risk and ambiguity. *NeuroImage* 236:118109.

Yoo SBM, Hayden BY, Pearson JM (2021) Continuous decisions. *Philosophical Transactions of the Royal Society B* 376(1819):20190664.

Zentner M, Grandjean D, Scherer KR (2008) Emotions evoked by the sound of music: characterization, classification, and measurement. *Emotion* 8(4):494.

# Electronic Companion

## EC.1. Experimental Instructions

THE UNIVERSITY OF MELBOURNE

BRAIN MIND AND MARKETS LABORATORY

**INSTRUCTIONS**

**Summary**

Your task will be to complete a trading game on our online trading platform. There will be a practice session, to provide ample opportunity to familiarise yourself with the online marketplace.

In the main trading session, which counts towards your earnings, you have the opportunity to trade a single, 15-period-lived security called "stock." The stock pays a random dividend at the end of each period. The expected dividend each period is 0.50 (experimental) dollar. You start with an endowment of cash and stock. You can sell stocks for cash, and/or use cash to buy stocks.

Your goal is to maximise your performance, measured by dividends received from holding stocks, plus cash accumulation through trading. You will be given an extra 10 (real) dollars as sign-up reward.

**Online platforms**

The trading games take place in an online trading platform called Flex-E-Markets. This online marketplace can be accessed through the following link: http://flexemarkets.com. Log on to the account and with the email and password given to you.

### *A. Practice Session*

**A1. Setting**

During the practice session, participants will be given the opportunity to trade a single asset called 'Apple', in the marketplace is named '*Monash Apple Market*'. At the end of the session, all available apples will be bought back by us at a random price. The price will be one of 0, 4, 6 and 10 (experimental) dollars, with equal chance. This means that the expected payoff for each apple is (0 + 4 + 6 + 10)/4 = 5 dollars.

**A2. Trading**

The trading platform is organised as a continuous open book system, in which you can submit orders to buy (bids) and orders to sell (asks) at any time when the market is open. When your bid is at a price higher than the best standing ask, you will trade immediately with the originator of that ask, at the ask price. When you submit an ask at a price below that of the best standing bid, you trade immediately with the bid originator, at the bid price. If there are more than one order at the same price, earlier orders get executed first. Order submission and transactions are anonymous.


BRAIN MIND AND MARKETS LABORATORY

The order submission area is located at the left-hand-side of the trading screen. There, you can switch between buying and selling using the "Buy/Sell" button, specify the quantity under "Units" and set your price using the slider below "Price". At any moment while the market is open, you can see outstanding bids and asks as blue (bids) and red (asks) entries. By toggling the "All/Mine" button, you can switch between viewing all the orders in market or the orders originated by you. You will be able to cancel your own orders if you need to do so. You will also be able to observe the list of prior transactions including information on time, price and (cumulative) quantity.

## *B. Main Trading Session*

In the main trading session, you will be trading a single security called 'Stock' against cash, in the marketplace named '*Monash Stock Market*', while sensors attached to your fingers and wrist track your heartbeat and transpiration. There will be fifteen (15) trading periods. After each of them, the stock pays a random dividend. After paying the last dividend in period 15, the stock expires worthless. Periods will last 5 minutes or less.

At the beginning of the first period, about half of the participants start with 20 units of stock and 100 experimental dollars. The remaining participants will start with 12 units of stock and 160 experimental dollars.

During each period, the market will be open and you will be able to trade. When the market closes we use random number generator to determine the dividend. We generate a uniformly distributed random integer from 1 to 100. If the generated number is in the range of [1 25], the dividend will be $0. If the generated random number falls in [26 50], the dividend will be $0.25. If the generated number falls in [51 75], the dividend will be $0.50, and if the generated number falls in [76 100], the dividend will be $1.25. As such, the expected period dividend equals $0.50 (=1/4*(0+0.25+0.5+1.25)).

The dividend will be distributed to you in the form of experimental cash and paid for each unit of stock you own at the end of the period. For example, if in period 1 you end with 25 units of stock and the dividend is $0.25, then $6.25 (=25*0.25) will be added to your experimental cash. In period 2 you will thus start with the same 25 units of the stock, but your cash will have increased relative to the end of period 1 by $6.25.

Each period, the dividend on each stock is $0.50 on average. Since there are 15 periods, the sum of all expected dividends in period 1 is $7.50 (=15*0.50) for each stock. In the second period the stock will have paid its first dividend, so there are only 14 more payments left, each worth $0.50 in

THE UNIVERSITY OF MELBOURNE
BRAIN MIND AND MARKETS LABORATORY

expectation. So, the sum of all remaining expected dividends is $7 (=14*0.50). In the third period there will be 13 more payments left, with an expected total dividend of $6.50, etc.

How much you are willing to buy or sell your stock for in any given period will depend on the horizon you plan to hold this stock for and what you foresee other participants to be willing to buy and sell the stock for.

## *C. Compensation*

The cash you are holding at the end of the 15th period is yours to take home. It will be converted to real dollars at a 10:1 exchange rate (meaning that 10 experimental dollars convert into 1 real dollar). In addition, you will be given 10 (real) dollars as sign-up reward. The expected amount of compensation for each participant in this experiment is $35. The actual amount can be higher or lower depending on your individual performance but will not be above $55 or below $25.

## EC.2. Experimental Setup: Trading Software and Physiological Measurement Devices

A. Screenshot of trading software B. Measurement devices

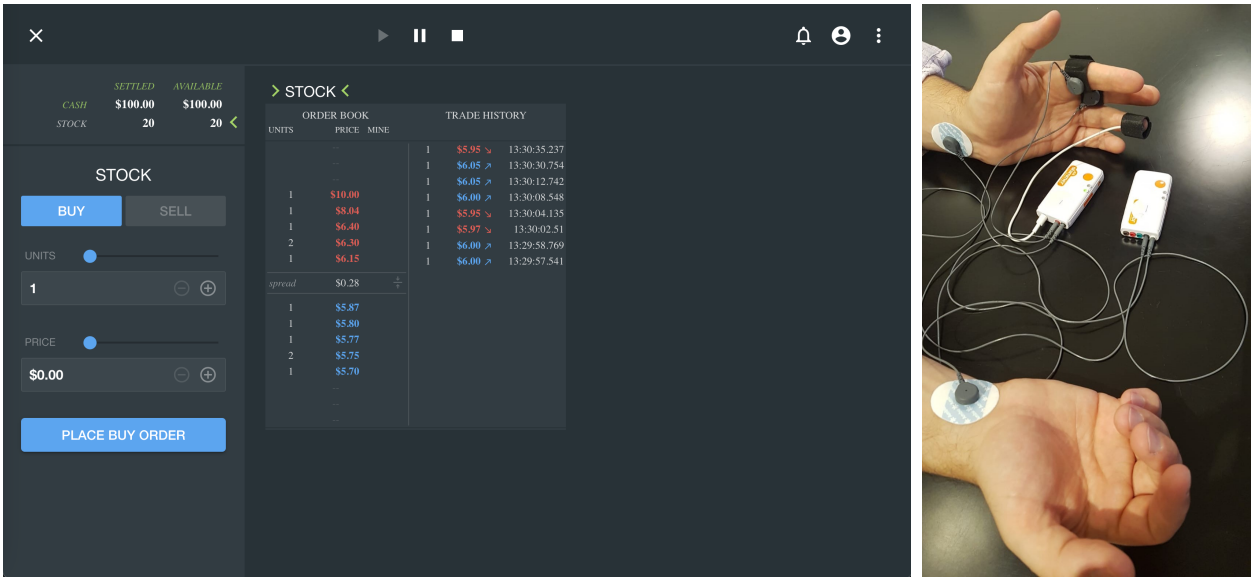



**Figure EC.1** A. Limit order book of continuous double auction implemented with Flex-E-Markets. B. ECG sensors on wrists, SCR sensors on fingers of non-dominant hand.

## EC.3. Data and Code

The data and code for all results reported in the paper can be found in the Github repository bmmlab/EmotionsInSSW. Here, we provide the output of the general linear modeling used to produce Figure 6 and Figure 7. Meaning of the variables can found in the README file and Matlab code in the Github repository.

### EC.3.1. Regression of Earnings onto Logarithm of Granger Causality Test Statistics, Interaction with ECP Calibration ONLY

Test 1 results are listed under the heading "Theoretical Likelihood Ratio Test." Individual slope coefficients (dots in Figure 6) are listed under "Fixed Effects."

# 1. HR Granger-Causes Ind Mispricing

mdl =
**Generalized linear mixed-effects model fit by PL**

**Model information:**

| | |
|---|---|
| Number of observations | 123 |
| Fixed effects coefficients | 5 |
| Random effects coefficients | 16 |
| Covariance parameters | 2 |
| Distribution | Normal |
| Link | Identity |
| FitMethod | MPL |

**Formula:**
Earnings ~ 1 + Granger_0:Treatment + Granger_5:Treatment + Granger_10:Treatment + Granger_15:Treatment + (1 | Session)

**Model fit statistics:**

| AIC | BIC | LogLikelihood | Deviance |
|---|---|---|---|
| 2512.4 | 2532.1 | -1249.2 | 2498.4 |

**Fixed effects coefficients (95% CIs):**

| Name | Estimate | SE | tStat | DF | pValue | Lower | Upper |
|---|---|---|---|---|---|---|---|
| {'(Intercept)' } | 24751 | 753.38 | 32.854 | 118 | 3.2581e-61 | 23259 | 26243 |
| {'Granger_0:Treatment_1' } | 886.9 | 299.2 | 2.9643 | 118 | 0.0036714 | 294.41 | 1479.4 |
| {'Granger_5:Treatment_1' } | 229.32 | 327.95 | 0.69928 | 118 | 0.48576 | -420.1 | 878.74 |
| {'Granger_10:Treatment_1'} | 621.07 | 368.78 | 1.6841 | 118 | 0.094799 | -109.21 | 1351.4 |
| {'Granger_15:Treatment_1'} | -1244.1 | 411.97 | -3.0198 | 118 | 0.0031013 | -2059.9 | -428.24 |

**Random effects covariance parameters:**
Group: Session (16 Levels)

| Name1 | Name2 | Type | Estimate |
|---|---|---|---|
| {'(Intercept)'} | {'(Intercept)'} | {'std'} | 1329.7 |

Group: Error

| Name | Estimate |
|---|---|
| {'sqrt(Dispersion)'} | 6106 |

LRstat =
**Theoretical Likelihood Ratio Test**

| Model | DF | AIC | BIC | LogLik | LRStat | deltaDF | pValue |
|---|---|---|---|---|---|---|---|
| mdlR | 3 | 2519.6 | 2528 | -1256.8 | | | |
| mdl | 7 | 2512.4 | 2532.1 | -1249.2 | 15.144 | 4 | 0.0044115 |

## 2. Ind Mispricing Granger-Causes HR

mdl =
**Generalized linear mixed-effects model fit by PL**

**Model information:**

| | |
|---|---|
| Number of observations | 123 |
| Fixed effects coefficients | 8 |
| Random effects coefficients | 16 |
| Covariance parameters | 2 |
| Distribution | Normal |
| Link | Identity |
| FitMethod | MPL |

**Formula:**

Earnings ~ 1 + Granger_0:Treatment + Granger_5:Treatment + Granger_10:Treatment + Granger_15:Treatment + Granger_20:Treatment + Granger_25:Treatment + Granger_30:Treatment + (1 | Session)

**Model fit statistics:**

| AIC | BIC | LogLikelihood | Deviance |
|---|---|---|---|
| 2526.1 | 2554.2 | -1253 | 2506.1 |

**Fixed effects coefficients (95% CIs):**

| Name | Estimate | SE | tStat | DF | pValue | Lower | Upper |
|---|---|---|---|---|---|---|---|
| {'(Intercept)' } | 24151 | 639.97 | 37.738 | 115 | 1.2855e-66 | 22883 | 25419 |
| {'Granger_0:Treatment_1' } | -499.86 | 409.78 | -1.2198 | 115 | 0.22503 | -1311.6 | 311.83 |
| {'Granger_5:Treatment_1' } | 25.454 | 332.77 | 0.076492 | 115 | 0.93916 | -633.69 | 684.6 |
| {'Granger_10:Treatment_1'} | 247.76 | 351.11 | 0.70565 | 115 | 0.48183 | -447.72 | 943.24 |
| {'Granger_15:Treatment_1'} | 850.5 | 490.61 | 1.7335 | 115 | 0.085679 | -121.31 | 1822.3 |
| {'Granger_20:Treatment_1'} | -142.78 | 439.11 | -0.32515 | 115 | 0.74565 | -1012.6 | 727.02 |
| {'Granger_25:Treatment_1'} | -379.38 | 419.17 | -0.90507 | 115 | 0.36732 | -1209.7 | 450.92 |
| {'Granger_30:Treatment_1'} | -677.59 | 348.62 | -1.9436 | 115 | 0.054384 | -1368.1 | 12.964 |

**Random effects covariance parameters:**

Group: Session (16 Levels)

| Name1 | Name2 | Type | Estimate |
|---|---|---|---|
| {'(Intercept)'} | {'(Intercept)'} | {'std'} | 679.7 |

Group: Error

| Name | Estimate |
|---|---|
| {'sqrt(Dispersion)'} | 6393.1 |

LRstat =
**Theoretical Likelihood Ratio Test**

| Model | DF | AIC | BIC | LogLik | LRStat | deltaDF | pValue |
|---|---|---|---|---|---|---|---|
| mdlR | 3 | 2519.6 | 2528 | -1256.8 | | | |
| mdl | 10 | 2526.1 | 2554.2 | -1253 | 7.4777 | 7 | 0.38089 |

## 3. HR Granger-Causes Holdings

mdl =
**Generalized linear mixed-effects model fit by PL**

**Model information:**

| | |
|---|---|
| Number of observations | 123 |
| Fixed effects coefficients | 5 |
| Random effects coefficients | 16 |
| Covariance parameters | 2 |
| Distribution | Normal |
| Link | Identity |
| FitMethod | MPL |

**Formula:**
Earnings ~ 1 + Granger_0:Treatment + Granger_5:Treatment + Granger_10:Treatment + Granger_15:Treatment + (1 | Session)

**Model fit statistics:**

| AIC | BIC | LogLikelihood | Deviance |
|---|---|---|---|
| 2522.2 | 2541.9 | -1254.1 | 2508.2 |

**Fixed effects coefficients (95% CIs):**

| Name | Estimate | SE | tStat | DF | pValue | Lower | Upper |
|---|---|---|---|---|---|---|---|
| {'(Intercept)' } | 24411 | 752.91 | 32.422 | 118 | 1.329e-60 | 22920 | 25902 |
| {'Granger_0:Treatment_1' } | 26.283 | 362.87 | 0.07243 | 118 | 0.94238 | -692.31 | 744.87 |
| {'Granger_5:Treatment_1' } | 577.8 | 351.21 | 1.6451 | 118 | 0.1026 | -117.7 | 1273.3 |
| {'Granger_10:Treatment_1'} | -414.88 | 322.7 | -1.2856 | 118 | 0.20109 | -1053.9 | 224.16 |
| {'Granger_15:Treatment_1'} | 433.23 | 312.67 | 1.3856 | 118 | 0.16849 | -185.95 | 1052.4 |

**Random effects covariance parameters:**
Group: Session (16 Levels)

| Name1 | Name2 | Type | Estimate |
|---|---|---|---|
| {'(Intercept)'} | {'(Intercept)'} | {'std'} | 1381.3 |

Group: Error

| Name | Estimate |
|---|---|
| {'sqrt(Dispersion)'} | 6353.6 |

LRstat =
**Theoretical Likelihood Ratio Test**

| Model | DF | AIC | BIC | LogLik | LRStat | deltaDF | pValue |
|---|---|---|---|---|---|---|---|
| mdlR | 3 | 2519.6 | 2528 | -1256.8 | | | |
| mdl | 7 | 2522.2 | 2541.9 | -1254.1 | 5.3794 | 4 | 0.25054 |

# 4. Holdings Granger-Cause HR

mdl =
**Generalized linear mixed-effects model fit by PL**

**Model information:**

| | |
|---|---|
| Number of observations | 123 |
| Fixed effects coefficients | 8 |
| Random effects coefficients | 16 |
| Covariance parameters | 2 |
| Distribution | Normal |
| Link | Identity |
| FitMethod | MPL |

**Formula:**
Earnings ~ 1 + Granger_0:Treatment + Granger_5:Treatment + Granger_10:Treatment + Granger_15:Treatment + Granger_20:Treatment + Granger_25:Treatment + Granger_30:Treatment + (1 | Session)

**Model fit statistics:**

| AIC | BIC | LogLikelihood | Deviance |
|---|---|---|---|
| 2512.4 | 2540.5 | -1246.2 | 2492.4 |

**Fixed effects coefficients (95% CIs):**

| Name | Estimate | SE | tStat | DF | pValue | Lower | Upper |
|---|---|---|---|---|---|---|---|
| {'(Intercept)' } | 24565 | 666.29 | 36.869 | 115 | 1.5272e-65 | 23246 | 25885 |
| {'Granger_0:Treatment_1' } | -421.02 | 358.24 | -1.1753 | 115 | 0.24232 | -1130.6 | 288.58 |
| {'Granger_5:Treatment_1' } | 884.5 | 407.47 | 2.1707 | 115 | 0.032006 | 77.392 | 1691.6 |
| {'Granger_10:Treatment_1'} | -1948.4 | 522.82 | -3.7267 | 115 | 0.00030242 | -2984 | -912.79 |
| {'Granger_15:Treatment_1'} | 644.17 | 391.12 | 1.647 | 115 | 0.10229 | -130.57 | 1418.9 |
| {'Granger_20:Treatment_1'} | -379.74 | 411.78 | -0.92218 | 115 | 0.35837 | -1195.4 | 435.93 |
| {'Granger_25:Treatment_1'} | 150.79 | 335.95 | 0.44883 | 115 | 0.6544 | -514.67 | 816.24 |
| {'Granger_30:Treatment_1'} | 36.288 | 353.68 | 0.1026 | 115 | 0.91846 | -664.29 | 736.86 |

**Random effects covariance parameters:**
Group: Session (16 Levels)

| Name1 | Name2 | Type | Estimate |
|---|---|---|---|
| {'(Intercept)'} | {'(Intercept)'} | {'std'} | 1272.3 |

Group: Error

| Name | Estimate |
|---|---|
| {'sqrt(Dispersion)'} | 5963.3 |

LRstat =
**Theoretical Likelihood Ratio Test**

| Model | DF | AIC | BIC | LogLik | LRStat | deltaDF | pValue |
|---|---|---|---|---|---|---|---|
| mdlR | 3 | 2519.6 | 2528 | -1256.8 | | | |
| mdl | 10 | 2512.4 | 2540.5 | -1246.2 | 21.134 | 7 | 0.0035765 |

## 5. HR Granger-Causes B/A Spread

```
mdl =
Generalized linear mixed-effects model fit by PL

Model information:
    Number of observations             123
    Fixed effects coefficients           5
    Random effects coefficients         16
    Covariance parameters                2
    Distribution                    Normal
    Link                            Identity
    FitMethod                       MPL

Formula:
    Earnings ~ 1 + Granger_0:Treatment + Granger_5:Treatment + Granger_10:Treatment + Granger_15:Treatment + (1 | Session)

Model fit statistics:
    AIC     BIC       LogLikelihood    Deviance
    2521    2540.7    -1253.5          2507

Fixed effects coefficients (95% CIs):
    Name                             Estimate    SE        tStat        DF     pValue         Lower      Upper
    {'(Intercept)'           }         24535     633.13       38.751    118    5.8343e-69      23281     25789
    {'Granger_0:Treatment_1' }         578.56     361.56       1.6002    118       0.11223    -137.43    1294.5
    {'Granger_5:Treatment_1' }        -37.248     415.26    -0.089697    118       0.92868    -859.58    785.08
    {'Granger_10:Treatment_1'}         732.75     412.42       1.7767    118      0.078193    -83.954    1549.5
    {'Granger_15:Treatment_1'}        -430.38      358.2      -1.2015    118       0.23196    -1139.7    278.96

Random effects covariance parameters:
Group: Session (16 Levels)
    Name1                Name2                Type           Estimate
    {'(Intercept)'}      {'(Intercept)'}      {'std'}        522.49

Group: Error
    Name                        Estimate
    {'sqrt(Dispersion)'}        6431.2

LRstat =
    Theoretical Likelihood Ratio Test

    Model    DF    AIC       BIC       LogLik     LRStat    deltaDF    pValue
    mdlR     3     2519.6      2528    -1256.8
    mdl      7       2521    2540.7    -1253.5    6.5582    4          0.16116
```

## 6. B/A Spread Granger-Causes HR

mdl =
**Generalized linear mixed-effects model fit by PL**

**Model information:**

| | |
|---|---|
| Number of observations | 123 |
| Fixed effects coefficients | 8 |
| Random effects coefficients | 16 |
| Covariance parameters | 2 |
| Distribution | Normal |
| Link | Identity |
| FitMethod | MPL |

**Formula:**
Earnings ~ 1 + Granger_0:Treatment + Granger_5:Treatment + Granger_10:Treatment + Granger_15:Treatment + Granger_20:Treatment + Granger_25:Treatment + Granger_30:Treatment + (1 | Session)

**Model fit statistics:**

| AIC | BIC | LogLikelihood | Deviance |
|---|---|---|---|
| 2527.5 | 2555.6 | -1253.7 | 2507.5 |

**Fixed effects coefficients (95% CIs):**

| Name | Estimate | SE | tStat | DF | pValue | Lower | Upper |
|---|---|---|---|---|---|---|---|
| {'(Intercept)' } | 24558 | 736.68 | 33.336 | 115 | 6.1272e-61 | 23099 | 26017 |
| {'Granger_0:Treatment_1' } | -383.21 | 429.62 | -0.89197 | 115 | 0.37427 | -1234.2 | 467.79 |
| {'Granger_5:Treatment_1' } | -200.83 | 349.42 | -0.57474 | 115 | 0.56659 | -892.96 | 491.31 |
| {'Granger_10:Treatment_1'} | -24.739 | 442.46 | -0.055912 | 115 | 0.95551 | -901.16 | 851.68 |
| {'Granger_15:Treatment_1'} | -101.55 | 538.95 | -0.18842 | 115 | 0.85088 | -1169.1 | 966 |
| {'Granger_20:Treatment_1'} | 509.28 | 326.28 | 1.5608 | 115 | 0.12131 | -137.03 | 1155.6 |
| {'Granger_25:Treatment_1'} | -146.55 | 453.95 | -0.32283 | 115 | 0.74741 | -1045.7 | 752.64 |
| {'Granger_30:Treatment_1'} | 458.57 | 402.89 | 1.1382 | 115 | 0.2574 | -339.48 | 1256.6 |

**Random effects covariance parameters:**
Group: Session (16 Levels)

| Name1 | Name2 | Type | Estimate |
|---|---|---|---|
| {'(Intercept)'} | {'(Intercept)'} | {'std'} | 1653.6 |

Group: Error

| Name | Estimate |
|---|---|
| {'sqrt(Dispersion)'} | 6287.6 |

LRstat =
**Theoretical Likelihood Ratio Test**

| Model | DF | AIC | BIC | LogLik | LRStat | deltaDF | pValue |
|---|---|---|---|---|---|---|---|
| mdlR | 3 | 2519.6 | 2528 | -1256.8 | | | |
| mdl | 10 | 2527.5 | 2555.6 | -1253.7 | 6.0892 | 7 | 0.52938 |

### EC.3.2. Regression of Earnings onto Logarithm of Granger Causality Test Statistics, All Sessions

Test 1 results are listed under the heading "Theoretical Likelihood Ratio Test." Individual slope coefficients (dots in Figure 7) are listed under "Fixed Effects."

# 1. HR Granger-Causes Ind Mispricing

mdl =
**Generalized linear mixed-effects model fit by PL**

**Model information:**

| | |
|---|---|
| Number of observations | 123 |
| Fixed effects coefficients | 5 |
| Random effects coefficients | 16 |
| Covariance parameters | 2 |
| Distribution | Normal |
| Link | Identity |
| FitMethod | MPL |

**Formula:**
Earnings ~ 1 + Granger_0 + Granger_5 + Granger_10 + Granger_15 + (1 | Session)

**Model fit statistics:**

| AIC | BIC | LogLikelihood | Deviance |
|---|---|---|---|
| 2520.2 | 2539.9 | -1253.1 | 2506.2 |

**Fixed effects coefficients (95% CIs):**

| Name | Estimate | SE | tStat | DF | pValue | Lower | Upper |
|---|---|---|---|---|---|---|---|
| {'(Intercept)'} | 24381 | 845.56 | 28.834 | 118 | 2.9158e-55 | 22707 | 26055 |
| {'Granger_0' } | 477.15 | 269.27 | 1.7721 | 118 | 0.078966 | -56.065 | 1010.4 |
| {'Granger_5' } | -10.289 | 271.12 | -0.037949 | 118 | 0.96979 | -547.18 | 526.6 |
| {'Granger_10' } | 315.02 | 292.57 | 1.0767 | 118 | 0.2838 | -264.36 | 894.4 |
| {'Granger_15' } | -584.46 | 285.24 | -2.049 | 118 | 0.042675 | -1149.3 | -19.616 |

**Random effects covariance parameters:**
Group: Session (16 Levels)

| Name1 | Name2 | Type | Estimate |
|---|---|---|---|
| {'(Intercept)'} | {'(Intercept)'} | {'std'} | 1243 |

Group: Error

| Name | Estimate |
|---|---|
| {'sqrt(Dispersion)'} | 6322.2 |

LRstat =
**Theoretical Likelihood Ratio Test**

| Model | DF | AIC | BIC | LogLik | LRStat | deltaDF | pValue |
|---|---|---|---|---|---|---|---|
| mdlR | 3 | 2519.6 | 2528 | -1256.8 | | | |
| mdl | 7 | 2520.2 | 2539.9 | -1253.1 | 7.3954 | 4 | 0.11641 |

## 2. Ind Mispricing Granger-Causes HR

mdl =
**Generalized linear mixed-effects model fit by PL**

**Model information:**

| | |
|---|---|
| Number of observations | 123 |
| Fixed effects coefficients | 8 |
| Random effects coefficients | 16 |
| Covariance parameters | 2 |
| Distribution | Normal |
| Link | Identity |
| FitMethod | MPL |

**Formula:**
Earnings ~ 1 + Granger_0 + Granger_5 + Granger_10 + Granger_15 + Granger_20 + Granger_25 + Granger_30 + (1 | Session)

**Model fit statistics:**

| AIC | BIC | LogLikelihood | Deviance |
|---|---|---|---|
| 2524.5 | 2552.6 | -1252.2 | 2504.5 |

**Fixed effects coefficients (95% CIs):**

| Name | Estimate | SE | tStat | DF | pValue | Lower | Upper |
|---|---|---|---|---|---|---|---|
| {'(Intercept)'} | 23961 | 664.31 | 36.07 | 115 | 1.5575e-64 | 22646 | 25277 |
| {'Granger_0' } | -113.69 | 276.04 | -0.41187 | 115 | 0.6812 | -660.48 | 433.09 |
| {'Granger_5' } | 111.68 | 249.42 | 0.44776 | 115 | 0.65517 | -382.37 | 605.72 |
| {'Granger_10' } | -85.489 | 296.92 | -0.28792 | 115 | 0.77393 | -673.64 | 502.66 |
| {'Granger_15' } | 616.85 | 335.6 | 1.838 | 115 | 0.068636 | -47.91 | 1281.6 |
| {'Granger_20' } | -376.28 | 328.8 | -1.1444 | 115 | 0.25484 | -1027.6 | 275.02 |
| {'Granger_25' } | -379.39 | 314.68 | -1.2056 | 115 | 0.23043 | -1002.7 | 243.93 |
| {'Granger_30' } | -511.44 | 290.96 | -1.7577 | 115 | 0.081451 | -1087.8 | 64.902 |

**Random effects covariance parameters:**
Group: Session (16 Levels)

| Name1 | Name2 | Type | Estimate |
|---|---|---|---|
| {'(Intercept)'} | {'(Intercept)'} | {'std'} | 677.78 |

Group: Error

| Name | Estimate |
|---|---|
| {'sqrt(Dispersion)'} | 6351.2 |

LRstat =
**Theoretical Likelihood Ratio Test**

| Model | DF | AIC | BIC | LogLik | LRStat | deltaDF | pValue |
|---|---|---|---|---|---|---|---|
| mdlR | 3 | 2519.6 | 2528 | -1256.8 | | | |
| mdl | 10 | 2524.5 | 2552.6 | -1252.2 | 9.0867 | 7 | 0.24649 |

## 3. HR Granger-Causes Holdings

```
mdl =
Generalized linear mixed-effects model fit by PL

Model information:
    Number of observations             123
    Fixed effects coefficients           5
    Random effects coefficients         16
    Covariance parameters                2
    Distribution                    Normal
    Link                            Identity
    FitMethod                       MPL

Formula:
    Earnings ~ 1 + Granger_0 + Granger_5 + Granger_10 + Granger_15 + (1 | Session)

Model fit statistics:
    AIC       BIC       LogLikelihood    Deviance
    2523.7    2543.4    -1254.9          2509.7

Fixed effects coefficients (95% CIs):
    Name                  Estimate    SE        tStat      DF     pValue         Lower      Upper
    {'(Intercept)'}         24330     764.47     31.826    118    9.5022e-60      22816      25844
    {'Granger_0'  }         58.631     276.5    0.21205    118       0.83244    -488.92    606.18
    {'Granger_5'  }         462.49     286.21    1.6159    118       0.10878    -104.28    1029.3
    {'Granger_10' }        -364.96     282.2    -1.2933    118       0.19845    -923.81    193.88
    {'Granger_15' }         145.07     237.85    0.60994    118       0.54308    -325.93    616.07

Random effects covariance parameters:
Group: Session (16 Levels)
    Name1                 Name2                 Type           Estimate
    {'(Intercept)'}       {'(Intercept)'}       {'std'}        1290.6

Group: Error
    Name                        Estimate
    {'sqrt(Dispersion)'}        6410.2

LRstat =
    Theoretical Likelihood Ratio Test

    Model    DF    AIC       BIC       LogLik     LRStat    deltaDF    pValue
    mdlR     3     2519.6      2528    -1256.8
    mdl      7     2523.7    2543.4    -1254.9    3.8158    4          0.43151
```

## 4. Holdings Granger-Cause HR

mdl =
**Generalized linear mixed-effects model fit by PL**

**Model information:**

| | |
|---|---|
| Number of observations | 123 |
| Fixed effects coefficients | 8 |
| Random effects coefficients | 16 |
| Covariance parameters | 2 |
| Distribution | Normal |
| Link | Identity |
| FitMethod | MPL |

**Formula:**
Earnings ~ 1 + Granger_0 + Granger_5 + Granger_10 + Granger_15 + Granger_20 + Granger_25 + Granger_30 + (1 | Session)

**Model fit statistics:**

| AIC | BIC | LogLikelihood | Deviance |
|---|---|---|---|
| 2519.2 | 2547.3 | -1249.6 | 2499.2 |

**Fixed effects coefficients (95% CIs):**

| Name | Estimate | SE | tStat | DF | pValue | Lower | Upper |
|---|---|---|---|---|---|---|---|
| {'(Intercept)'} | 24329 | 741.31 | 32.818 | 115 | 3.1303e-60 | 22860 | 25797 |
| {'Granger_0' } | -215.99 | 278.78 | -0.77476 | 115 | 0.44007 | -768.2 | 336.22 |
| {'Granger_5' } | 633.49 | 310 | 2.0435 | 115 | 0.043284 | 19.445 | 1247.5 |
| {'Granger_10' } | -932.09 | 366.59 | -2.5426 | 115 | 0.012332 | -1658.2 | -205.95 |
| {'Granger_15' } | 417.21 | 323.25 | 1.2907 | 115 | 0.19941 | -223.08 | 1057.5 |
| {'Granger_20' } | -507.62 | 342.4 | -1.4825 | 115 | 0.14093 | -1185.8 | 170.6 |
| {'Granger_25' } | 173.79 | 278.83 | 0.62329 | 115 | 0.53433 | -378.51 | 726.09 |
| {'Granger_30' } | -104.22 | 329.31 | -0.31647 | 115 | 0.75222 | -756.51 | 548.07 |

**Random effects covariance parameters:**
Group: Session (16 Levels)

| Name1 | Name2 | Type | Estimate |
|---|---|---|---|
| {'(Intercept)'} | {'(Intercept)'} | {'std'} | 1681.5 |

Group: Error

| Name | Estimate |
|---|---|
| {'sqrt(Dispersion)'} | 6064.2 |

LRstat =
**Theoretical Likelihood Ratio Test**

| Model | DF | AIC | BIC | LogLik | LRStat | deltaDF | pValue |
|---|---|---|---|---|---|---|---|
| mdlR | 3 | 2519.6 | 2528 | -1256.8 | | | |
| mdl | 10 | 2519.2 | 2547.3 | -1249.6 | 14.382 | 7 | 0.044783 |

## 5. HR Granger-Causes B/A Spread

```
mdl =
Generalized linear mixed-effects model fit by PL

Model information:
    Number of observations             123
    Fixed effects coefficients           5
    Random effects coefficients         16
    Covariance parameters                2
    Distribution                    Normal
    Link                            Identity
    FitMethod                       MPL

Formula:
    Earnings ~ 1 + Granger_0 + Granger_5 + Granger_10 + Granger_15 + (1 | Session)

Model fit statistics:
    AIC       BIC       LogLikelihood    Deviance
    2524.5    2544.2    -1255.3          2510.5

Fixed effects coefficients (95% CIs):
    Name                  Estimate    SE        tStat       DF     pValue        Lower      Upper
    {'(Intercept)'}         24726     776.79      31.831    118    9.3406e-60      23188      26264
    {'Granger_0'  }        299.15      275.2       1.087    118       0.27925    -245.83    844.12
    {'Granger_5'  }        39.605      310.28     0.12764    118       0.89865    -574.84    654.05
    {'Granger_10' }        394.78      332.44      1.1875    118       0.23741    -263.54    1053.1
    {'Granger_15' }        -88.63      304.89    -0.29069    118        0.7718     -692.4    515.14

Random effects covariance parameters:
Group: Session (16 Levels)
    Name1                 Name2                 Type           Estimate
    {'(Intercept)'}       {'(Intercept)'}       {'std'}        1438.4

Group: Error
    Name                        Estimate
    {'sqrt(Dispersion)'}        6406.6

LRstat =
    Theoretical Likelihood Ratio Test

    Model    DF    AIC       BIC       LogLik     LRStat    deltaDF    pValue
    mdlR     3     2519.6      2528    -1256.8
    mdl      7     2524.5    2544.2    -1255.3    3.0578    4          0.54819
```

## 6. B/A Spread Granger-Causes HR

```
mdl =
Generalized linear mixed-effects model fit by PL

Model information:
    Number of observations             123
    Fixed effects coefficients           8
    Random effects coefficients         16
    Covariance parameters                2
    Distribution                    Normal
    Link                            Identity
    FitMethod                       MPL

Formula:
    Earnings ~ 1 + Granger_0 + Granger_5 + Granger_10 + Granger_15 + Granger_20 + Granger_25 + Granger_30 + (1 | Session)

Model fit statistics:
    AIC     BIC       LogLikelihood    Deviance
    2529    2557.1    -1254.5          2509

Fixed effects coefficients (95% CIs):
    Name                  Estimate    SE        tStat       DF     pValue        Lower      Upper
    {'(Intercept)'}         24523     755.62      32.454    115    9.9793e-60      23026     26020
    {'Granger_0'  }        -35.877    298.64    -0.12013    115       0.90459    -627.43    555.67
    {'Granger_5'  }         -208.5    316.43    -0.65889    115       0.51128    -835.29     418.3
    {'Granger_10' }        -232.59    360.72     -0.6448    115       0.52034    -947.12    481.93
    {'Granger_15' }         85.933    406.24     0.21153    115       0.83285    -718.75    890.62
    {'Granger_20' }         309.37    268.21      1.1535    115       0.25112    -221.91    840.64
    {'Granger_25' }         -47.52    312.26    -0.15218    115       0.87931    -666.06    571.02
    {'Granger_30' }         343.42    297.86       1.153    115       0.25132    -246.58    933.42

Random effects covariance parameters:
Group: Session (16 Levels)
    Name1                 Name2                 Type           Estimate
    {'(Intercept)'}       {'(Intercept)'}       {'std'}        1470.6

Group: Error
    Name                        Estimate
    {'sqrt(Dispersion)'}        6359.4

LRstat =
    Theoretical Likelihood Ratio Test

    Model    DF    AIC       BIC       LogLik     LRStat    deltaDF    pValue
    mdlR      3    2519.6      2528    -1256.8
    mdl      10      2529    2557.1    -1254.5    4.6072    7          0.70777
```