Behavioral Frictions in Predictive Signal Use: Mutual Funds and Trading Contests

Ahmed Guecioueur*

February 2, 2026

Abstract

This paper investigates why anomaly returns can persist after becoming public. Combining evidence from mutual fund portfolio holdings with laboratory-style trading contests in which participants receive common predictor sets, I document substantial under-adoption of even genuinely predictive signals. An investor's experience, rather than objective signal informativeness, is the primary determinant of signal adoption. Exploiting institutional features of the trading contests, I uncover a familiarity bias toward recognizable signals, together with an aversion to specifying complex predictive models that diminishes with experience. The benefits from predictive signals thus concentrate among experienced investors, who converge in the models they specify.

Keywords: Anomalies, Experience, Familiarity, Machine Learning, Complexity Aversion
JEL Codes: G11, G14, G41

*University of Washington, Foster School of Business, email: ahmedgc@uw.edu
I would like to thank participants at various conferences and doctoral consortia, including SFS Cavalcade North America, FIRS, the Cambridge Conference on Alternative Finance, the Chicago Booth Machine Learning in Economics Summer Institute, and the UT Austin PhD Symposium. I would also like to thank Simona Abis, Stefano Actis (Quantiacs), Frederico Belo, Olivier Dessaint, Jérôme Dugast, Bernard Dumas, Gleb Gertsman, Francisco Gomes, Naveen Gondhi, Georgina Hall, Xing Huang, Don Noh, Yukun Liu, Joël Peress, Rodolfo Prieto, Alexandra Roulet, Amit Seru, Stephan Siegel, Taisiya Sikorskaya, Yang Song, Ke Tang, Qingyuan Yang and Bart Zhou Yueshen for their helpful comments. I gratefully acknowledge the support of the Henry Grunfeld Foundation while I was a doctoral student. Earlier drafts were circulated and presented under the titles "How Do Investors Learn as Data Becomes Bigger? Evidence From a FinTech Platform" (October 2020) and "Machine Traders, Human Behavior, and Model (Mis)Specification."

1 Introduction

Data-driven prediction is increasingly central to economic decision-making, and nowhere more so than in finance. Academic researchers have uncovered an expanding set of return-predictive signals,[1] which practitioners may adopt once they are publicly disseminated. McLean and Pontiff (2016) show that when such signals are published, their implied portfolio returns subsequently decline by roughly one-half. But this also means that economically meaningful predictability persists, which raises the possibility of limited or delayed adoption, despite the strong economic incentives to do so.

This paper investigates the presence and nature of behavioral frictions in the usage of publicly available predictive signals. I study investors in two different yet complementary settings: mutual fund managers whose stock portfolios can incorporate publicly available return predictors, and participants in a series of tightly-controlled trading contests who are endowed with common sets of predictive signals. Across both settings, I document a notable reluctance to adopt even genuine predictive signals, as well as the key role played by human experience (as distinct from objective features of signal datasets such as informativeness). Exploiting various institutional features of the trading contests setting, I uncover a familiarity bias in signal adoption. Furthermore, I show that traders are averse to specifying complex predictive models (Oprea, 2020), although this aversion lessens with experience. The benefits of these signals thus concentrate among the experienced investors, who are most willing to adopt them.

Mutual Funds. In my first setting, I examine the common stock portfolios held by actively-managed mutual funds (and thus selected by their managers). Empirically, I measure signal adoption using quarterly portfolio holdings, building on a literature that infers characteristic tilts and portfolio preferences from disclosed positions.[2] Specifically, using the 212 firm-level predictors collected by Chen and Zimmermann (2022), I construct portfolio–signal–quarter *tilts* that capture whether a fund manager overweights stocks that a given signal ranks as having higher expected returns (net of the market's exposure). This yields a comparable, holdings-based proxy for implementing each public signal and allows a unified comparison across a large cross-section of predictors.

[1] Cochrane (2011) argues for the importance of understanding this "factor zoo," and various replication studies collect and compare such signals that predict "anomaly returns" (e.g., Hou, Xue, and Zhang, 2020; Chen and Zimmermann, 2022; Jensen, Kelly, and Pedersen, 2023).

[2] See, among others, Grinblatt, Titman, and Wermers (1995), Del Guercio (1996), Falkenstein (1996), Chan, Chen, and Lakonishok (2002), Bennett, Sias, and Starks (2003), Lewellen (2011), and Lettau, Ludvigson, and Manoel (2018). Holdings are the appropriate object of study because investors must actually hold the stocks implied by predictive signals in order to earn the next-period anomaly returns.

Two immediate facts emerge from analyzing portfolio–signal tilts. First, fund managers tilt more toward signals that have been public for longer: within a portfolio and quarter, signal tilts increase with time since publication. This gradual uptake is consistent with delayed adoption and, in particular, a preference for familiar predictors. Second, tilts are misaligned with predictability: I find a negative relationship between fund tilts and signals' subsequent (post-publication) performance, indicating that managers are *less likely* to adopt the most informative signals. Both patterns remain when controlling for fund characteristics and when using portfolio × quarter fixed effects that compare, within a given portfolio and quarter, which signals receive relatively greater weight. Only a small subset of 16 long/short quantitative funds tilt in the direction of predictability and place relatively greater weight on newer signals; the vast majority of funds (including most self-identified quant funds) do not.

Experience partly mitigates this misalignment: individual managers with more accumulated experience exhibit a weaker aversion to adopting genuine predictors; i.e., the negative association between portfolio tilt and signal informativeness is less pronounced. This experience effect is apparent both with manager fixed effects and manager × portfolio fixed effects that absorb manager–fund matching. Thus, while under-adoption of genuine public predictors is pervasive, experience partially attenuates this behavioral friction, though it is not eliminated.

Because mandates, career concerns, and organizational constraints may also vary with experience in the context of mutual fund management, I next turn to a complementary setting that holds fixed the incentives and information environment while allowing sharper tests of adoption mechanisms.

Trading Contests. My second setting exploits a series of online trading contests run by the Quantiacs platform. This setting offers a controlled yet realistic environment: 874 participants across 12 contests (launched during 2014–2019) implement futures trading strategies,[3] competing for profit-sharing contracts and assessed on both in-sample and out-of-sample performance. Participants thus face a prediction problem: to use the available signals on the Quantiacs platform to implement a trading strategy that translates signals to portfolio positions (which I do not observe). Critically, all contestants have access to identical sets of predictive variables – mostly, historical futures prices and, beginning in contest 8, 54 macroeconomic indicators – and cannot upload any external data to the platform. Contestants develop their strategies using a backtest (in-sample) period and are then evaluated on performance during

[3]Participants in the trading contests have access to machine learning libraries and optimization routines, and are thus fully able to exploit the high-dimensional set of predictive signals. Therefore, any behavioral frictions demonstrated by this group of subjects are informative about how machine learning tools are used by financial decision-makers more broadly (e.g., Sheng, Sun, Yang, and Zhang, forthcoming; Chen, Sialm, and Xu, 2025).

a subsequent live (out-of-sample) period. I observe not only contest-level performance metrics but also several decades' worth of daily portfolio positions and turnover per trading strategy, from which I infer signal usage.

The contest setting also allows sharper tests of a familiarity-based mechanism. Macro signals are updated on specific release dates, but the mix of indicators updated on a given day varies. I measure the familiarity of each indicator using past mentions in news and books; these measures are (strongly correlated) proxies for the familiarity of the signal to a layperson. I find that trading intensity is higher on days when the released signals are more familiar, revealing a preference for familiar and recognizable signals.

I next separate familiarity from objective signal quality by measuring the out-of-sample informativeness of released macro signals using Random Forest models (Gu, Kelly, and Xiu, 2020) and comparing informativeness to trading responses. While all strategies respond somewhat to informative signals in this setting, those of experienced traders respond significantly more strongly. This differential sensitivity to genuine predictive power, rather than mere familiarity, suggests that a willingness to adopt genuinely predictive signals is a dimension of investor skill, and one that accrues with experience. Consistent with this, experienced investors with access to the macroeconomic signals achieve higher out-of-sample Sharpe ratios compared to investors in earlier trading contests without access to these signals. I use Heckman's (1976) two-stage framework to rule out selection effects as a driver of this result, confirming that the benefits of predictive signals concentrate among experienced participants.[4]

I show that experience predicts investor decision-making along a number of additional dimensions in this pure prediction setting. Examining the cross-section of pairwise correlations in trading volume, I find that experienced investors act more similarly to one another that inexperienced investors do, suggesting a convergence in the predictive models (i.e., trading strategies) that they implement. Model disagreement (Cookson and Niessner, 2020) is thus partly driven by model mis-specification, and lessens among experienced investors.

Relatedly, I find that experienced investors tend to specify more complex predictive models that employ a greater number of predictive signals in their implementation. To arrive at this conclusion, I first extend Martin and Nagel's (2022) model of investor behavior when facing a prediction problem, allowing for the possibility that an investor expresses a bias against a

[4]This positive experience gradient contrasts with findings on AI adoption by Brynjolfsson, Li, and Raymond (2025), who determine that *in*experienced workers benefit most from chatbot-like tools, suggesting that such technologies substitute for experience. I find instead that human experience is a complement to the adoption of predictive signals in a financial markets setting.

valid predictive variable.[5] I show that, with appropriate assumptions, this framework admits a closed-form solution to the weights investors could assign to valid signals, and these weights are equivalent to those estimated by the well-known lasso machine learning technique (Tibshirani, 1996). I exploit this structural-like and behavioral microfoundation of the lasso to estimate signal usage by repeatedly fitting the lasso to individual strategy returns. I find a consistent pattern of experienced investors demonstrating a willingness to incorporate more signals into their models. Investors thus demonstrate an aversion to implementing complex predictive models, linking to findings by Oprea (2020) in a different (non-prediction-based) setting. Again, human experience is important in alleviating this behavioral friction.

Related Literature. This paper contributes to a growing economics literature emphasizing that the value of predictive information depends not only on statistical accuracy but also on how agents convert signals into decisions (Kleinberg, Ludwig, Mullainathan, and Obermeyer, 2015). Mechanisms highlighted by prior work include distraction and limited attention (DellaVigna and Pollet, 2009), imperfect understanding (Cohen and Frazzini, 2008), limits to learning driven by what agents notice (Hanna, Mullainathan, and Schwartzstein, 2014), theoretical models of sparse attention and selective information processing (Gabaix, 2014; Matějka and McKay, 2015), the psychic costs of following complex rules (Oprea, 2020), and more general processing costs (Blankespoor, deHaan, and Marinovic, 2020). I shed light on the adoption of predictive signals, including a setting in which subjects implement predictive models in the form of computer-coded trading strategies (with the ability to employ machine learning tools and optimization routines). I uncover pervasive under-adoption of genuinely informative signals, a bias toward familiar predictors – akin to familiarity effects in stock choice (Huberman, 2001; Seasholes and Zhu, 2010) – and experience gradients in adoption, performance, and model disagreement (Cookson and Niessner, 2020). I also contribute a behavioral microfoundation to the lasso and use it to empirically document an aversion to specifying rich, complex predictive models and a preference for simplicity, which also varies with experience.[6]

In asset pricing, this paper helps explain why return predictability can persist after signals become public. It relates to debates over whether documented predictors reflect data mining versus genuine regularities (Harvey, Liu, and Zhu, 2016; Linnainmaa and Roberts, 2018; Hou, Xue, and Zhang, 2020; Chen and Zimmermann, 2022; Jensen, Kelly, and Pedersen, 2023), and to whether anomaly returns reflect mispricing versus compensation for risk (Fama, 1998; Bar-

[5] A bias against unfamiliar signals is one such example. Inspired by Cao, Han, Hirshleifer, and Zhang's (2011) interpretation of unfamiliarity as a fear of worst-case outcomes, I model an investor's bias by injecting a fear of worst-case outcomes into her prediction problem.

[6] Experience has been found to shape manual trading skill (Seru, Shumway, and Stoffman, 2010), and can attenuate behavioral biases in other contexts (List, 2003).

beris and Thaler, 2003). The post-publication attenuation documented by McLean and Pontiff (2016) is consistent with both data mining and partial correction of mispricing. Focusing on the post-publication period and on a comprehensive set of replicated predictors (Chen and Zimmermann, 2022), I show that behavioral frictions – in particular, familiarity bias and an experience-dependent aversion to complex predictive models – limit adoption even when predictors remain informative. These frictions therefore act as a psychological limit to arbitrage that may prevent the resolution of mispricing. Indeed, Martin and Nagel (2022) show that market efficiency depends on how agents process high-dimensional data.

Finally, this paper relates to multiple strands of the literature on mutual funds. First, a line of work studies fund managers' use of public information as a skill (Kacperczyk and Seru, 2007; Kacperczyk, Van Nieuwerburgh, and Veldkamp, 2016) and the related impact of novel data sources on managers themselves (Abis and Veldkamp, 2024; Bonelli and Foucault, 2023). Second, a number of studies examine funds' portfolio tilts (Lewellen, 2011) or trading activities (Edelen, Ince, and Kadlec, 2016; McLean, Pontiff, and Reilly, 2025) to quantify the extent to which they exploit return anomalies. My evidence complements these studies by measuring adoption within-portfolios and across dozens of published predictors, and by using a controlled trading environment to distinguish institutional constraints from behavioral mechanisms; namely, familiarity bias and complexity aversion. I also highlight the role of experience effects in signal adoption, and show their importance even in settings where fund managers face a variety of institutional constraints and influences.

2 Evidence From Mutual Fund Portfolio–Signal Tilts

The first setting examined in this paper is that of mutual funds' common stock portfolios and their tilts to characteristics that are known to predict returns.

2.1 Data

Mutual Fund Portfolios and Characteristics. I obtain mutual fund holdings data at a quarterly frequency from the CRSP Mutual Fund Database. I begin the sample in 2008, when CRSP holdings coverage and reliability improve (Schwarz and Potter, 2016); an additional advantage of CRSP over other data sources such as 13-F filings is that it records both long and short positions, which is essential for measuring active tilts. I focus on domestic equity funds, exclude incubating funds, and exclude index funds following standard practice (Ben-David, Li, Rossi, and Song, 2022). These filters yield a sample of actively managed mutual fund portfolios for which I can compute portfolio tilts.

In some analyses, I also use information from Form 485 and Form 497 disclosures to on whether a fund describes itself as following a quantitative strategy (Dyer, Guest, and Yu, 2025).

All other fund characteristics – including the names of their managers and their histories managing the current and past funds – are obtained from the CRSP Mutual Fund Database.

Stock Return-Predictive Signals. I obtain stock-level return predictors from the open-source database of Chen and Zimmermann (2022) (version 2.0). The database contains 212 firm-level characteristics that have been shown to predict returns, 208 of which I match to an academic publication date using a combination of the OpenAlex API and manual inspection.

Throughout the paper, I measure each signal's predictive power in the spirit of McLean and Pontiff (2016), using the *ex post* mean returns of long–short portfolios formed by sorting stocks on the signal. This provides a simple and comparable metric of post-publication signal performance across a broad set of predictors. The portfolio weighting scheme is the same as that used in the original academic article. Figure 1 summarizes the empirical distribution of these measures of signal predictive power throughout the common sample period.

[Insert Figure 1 here]

Portfolio–Signal Tilts. To measure whether a fund manager actively positions her portfolio in the direction implied by a given return-predictive characteristic, I construct a portfolio–signal tilt at a quarterly frequency. Let t index quarters, let c index a firm-level predictive characteristic, and let i index stocks. For each characteristic c and quarter t, I sort stocks into deciles using NYSE breakpoints and define $s_{c,i,t} \in \{1, \ldots, 10\}$ as stock i's decile rank. The decile ordering is chosen so that higher values of $s_{c,i,t}$ correspond to higher expected returns.

For a portfolio p, I define gross-normalized holdings weights

$$w_{p,i,t} = \frac{V_{p,i,t}}{\sum_i |V_{p,i,t}|},$$

where $V_{p,i,t}$ is the signed position value. Gross normalization makes exposures comparable across portfolios that differ in leverage and naturally accommodates short positions. I define the analogous market weight $w_{MKT,i,t}$ for the market portfolio.

Given these ingredients, I first measure a portfolio's raw exposure to characteristic c as the

holdings-weighted average of the decile ranks,

$$\text{Portfolio Exposure}_{p,c,t} \triangleq \sum_i w_{p,i,t}\, s_{c,i,t}, \tag{1}$$

$$\text{MKT Exposure}_{c,t} \triangleq \sum_i w_{MKT,i,t}\, s_{c,i,t}. \tag{2}$$

Because the market portfolio can have nontrivial exposure to many characteristics, raw exposure alone conflates active positioning with default market exposure. I therefore define a portfolio's *signal tilt* as its excess exposure relative to the market:

$$\text{Portfolio Tilt}_{p,c,t} \triangleq \sum_i w_{p,i,t}\Big(s_{c,i,t} - \text{MKT Exposure}_{c,t}\Big) \tag{3}$$

$$= \text{Portfolio Exposure}_{p,c,t} - \eta_{p,t}\,\text{MKT Exposure}_{c,t}, \tag{4}$$

where $\eta_{p,t} \triangleq \sum_i w_{p,i,t}$ is the portfolio's net exposure. For a long-only portfolio, which comprise the vast majority of mutual fund portfolios, $\eta_{p,t} = 1$. For a market-neutral long/short portfolio, $\eta_{p,t} = 0$. In this way, Portfolio Tilt$_{p,c,t}$ is a directional, relative proxy for active signal use: it is higher when the portfolio overweights stocks with higher signal ranks *beyond* any baseline exposure inherited from the market. This tilt measure is closely related to holdings-based characteristic-exposure measures used in Chan, Chen, and Lakonishok (2002), Lewellen (2011), and Lettau, Ludvigson, and Manoel (2018). Figure 2 summarizes the empirical distribution of these tilts across portfolios.

[Insert Figure 2 here]

2.2 Results

Table 1 relates portfolio–signal tilts to signal age and a direct measure of a signal's post-publication informativeness, defined as the signal's subsequent long–short mean return (from $t+1$ onward). Across specifications, the coefficient on time elapsed since signal publication is positive and highly significant, indicating that funds load more strongly on signals that have been public for longer; this fact suggests managers could be influenced by the familiarity or recognizability of the signal.

[Insert Table 1 here]

Additionally, tilts are *negatively* related to signal predictive power: managers are less likely to adopt – and more likely to tilt against – signals that are *more* informative out-of-sample. This negative association suggests that signal adoption is not simply a rational response to pre-

dictive performance. Both forms of misalignment are robust to adding standard fund controls and, most stringently, to portfolio × quarter fixed effects (column (3)), which compare signals within the same portfolio and quarter while absorbing time-varying portfolio-level determinants of holdings, such as mandates.

Table 2 shows that these patterns are mostly consistent across portfolio and fund types. The positive loading on older signals and the negative loading on informative signals are strongest among long-only portfolios and among the broad set of funds, including self-identified quant funds. One sub-sample of long/short quant strategies, exhibits the opposite pattern: these funds load positively on signal predictive power and (weakly) negatively on signal age, consistent with a greater willingness to adopt newer, more informative predictors. Nevertheless, this very small group consisting of 16 mutual funds is not representative of the broader empirical pattern of a mis-use of predictive signals.

[Insert Table 2 here]

Finally, Table 3 shows that experience partly mitigates the tendency to underweight informative signals. In the sample of single-manager funds, the interaction between an experienced-manager indicator and signal predictive power is positive and statistically significant across fixed-effect specifications. This implies that experienced managers tilt less strongly away from informative signals (i.e., they exhibit a weaker negative relation between tilts and predictability), even after controlling for manager heterogeneity and manager–portfolio matching.

[Insert Table 3 here]

Limitations. The mutual fund evidence reveals systematic patterns in signal adoption: behavior that suggests a preference for familiar over informative signals, and partial improvement with manager experience. However, isolating the underlying mechanisms is challenging, because portfolio tilts may be influenced by other manager-level characteristics that are correlated with experience, such as career concerns. Moreover, the causal direction between signal use and predictive power is unclear in a setting where adoption may causally impact predictability (McLean and Pontiff, 2016). Implementation lags and the gradual diffusion of information could also explain delayed adoption without invoking behavioral mechanisms. Finally, signal adoption in mutual funds is often a team effort involving analysts, portfolio managers, and risk officers, making it difficult to isolate individual-level biases. In acknowledgement of these limitations, I next turn to a complementary setting that holds fixed the incentives and information environment while allowing sharper tests of adoption mechanisms.

3 Trading Contests: Institutional Setting and Data

I now describe the controlled trading contests setting, my dataset of observable outcomes, and the institutional features I am able to exploit to uncover behavioral frictions in signal use. Further details are provided in Internet Appendix C.

3.1 The Quantiacs Trading Platform and Trading Contests

Quantiacs is a FinTech platform that runs simulated trading contests for investors with the inclination to design a systematic trading strategy. The business model of Quantiacs is to identify the best 3 contestants in each contest and allocate assets to them, thus building up a portfolio of multiple delegated systematic trading strategies.[7] Contestants upload code (in Python or Matlab) to implement a strategy that takes long or short positions in futures contracts, and each strategy's performance is assessed based on its in-sample ("backtest") Sharpe Ratio prior to the start of the contest, and its out-of-sample ("live") Sharpe Ratio during the contest period. The official scores assigned to entries incorporate the out-of-sample Sharpe Ratio, and so this incentivizes traders to perform well out-of-sample.[8] In-sample performance is determined from historical daily data, and this is visible to traders as they backtest and fine-tune their strategy ahead of a contest launch. The Sharpe Ratios reported by the trading platform include the effect of simulated transaction costs, which the investors also perceive. Out-of-sample performance is calculated using market data that arrives after the launch of a contest; i.e. during the Live period. During this Live period, contestants are unable to modify their trading strategy in any way whatsoever. Each trading strategy updates portfolio positions at a daily frequency. The Live period of each contest lasts for approximately 3 months, and the 12 contests in my panel thus cover a number of years in out-of-sample/live calendar time. Figure 3 illustrates the distinction between the in-sample/backtest and out-of-sample live periods of a contest.

[Insert Figure 3 here]

To make the institutional setting more concrete, Internet Appendix C presents screenshots of the Quantiacs platform that illustrate the steps taken by an investor to code up, backtest and then submit a trading strategy entry to a contest.

[7] Case studies by Fleiss, Kominers, and Ughetta (2017) and Zheng (2017) describe the business model further. I emphasize that my sample is limited to entries to the open-entry trading contests in which investors are on an equal footing, and does not include any observations from profit-sharing trades.

[8] An entry's official score is defined to be the minimum of its in- and out-of-sample Sharpe Ratios. Since in-sample Sharpe Ratios are known at the end of the Backtest period, the investor's problem at the beginning of the Live period is indeed to maximize the out-of-sample Sharpe Ratio. In Internet Appendix I, I check that results are robust to using the official score instead of the out-of-sample SR.

The Quantiacs setting is realistic in a number of ways that are difficult to achieve in a laboratory. First, the platform uses actual market data throughout; experimental studies tend to rely on simulations of simplified data-generating processes. Second, investors are granted access to a sophisticated backtesting facility and software packages, and have weeks to design and refine their trading strategies. Most importantly, participants are offered significant financial incentives: investors are incentivized by the offer of profit-sharing contracts of 10% of net profits on an allocation of $1 Mln, $0.75 Mln & $0.5 Mln to the 3 best-performing investors in each contest, respectively. These conditions make the trading contests setting a realistic one.

At the same time, because I use data on trading contests, participants are constrained in ways that ensure empirical comparisons between them are fair. First, the human investors are required to implement trading strategies in the form of computer code that runs remotely on the Quantiacs platform, and which they cannot modify during a contest. Beyond that, humans play no role in the actions of their machines; for example, they do not review or approve machine trades, nor can they make their own discretionary trades. Second, the contest setting assigns all investors a common, well-defined goal (incentivizing them to maximize their trading strategies' out-of-sample Sharpe Ratios) and endows them with a common dataset of predictive variables for use as inputs to the machine-based trading strategies; investors cannot upload their own datasets. Furthermore, the contest setting fixes investors' preferences (due to the common goal), horizons (contest period) and other well-known drivers of portfolio choice that might confound my analysis.[9]

It is possible for an investor to enter any contest that she pleases, and there is no entry fee. Contest live periods are non-overlapping. I am able to observe individual investors' performance outcomes at an individual trading strategy level. For confidentiality reasons, I am unable to observe the granular positions taken by trading strategies, or the code used to implement them. The platform did not collect identifying information from contestants, and did not require them to provide any demographic characteristics.

While investors share a common information set *within* a contest, the Quantiacs platform widened the set of available variables in between contests: macroeconomic indicators were added to investors' historical and live trading datasets after the end of the 7th contest and before the beginning of the 8th contest. Out of a total of 12 contests, investors in the first

[9]In addition to the commonly-enforced objective, investors are in a perfectly competitive setup because they cannot affect each others' payoffs through price impact or observability of actions, thus freeing them from the usual strategic considerations. Background risk or other wealth effects do not enter into investors' objectives, either, since no payment or stake is required to enter into a contest. Investors are fully committed to following their trading strategies during the out-of-sample phase of each contest, as it is no longer possible to modify a contest entry's computer code at that stage.

7 therefore had narrower datasets than investors in the final 5. I exploit this variation in the (common) information set across contests to measure the gains to investors from having access to the additional predictive variables. Sections 3.2.4 and 3.3 provide further details, and additionally describe the (irregular) release schedules of different macroeconomic indicators.

3.2 Data

3.2.1 Contest Leaderboard Panel

My leaderboard sample consists of 12 trading contests spanning a number of years. I can identify individual traders who may (and often do) take part in multiple contests over time in order to measure performance dynamics. I can also exploit the panel structure of the leaderboard dataset to incorporate fixed effects in my regressions to conduct within-investor analyses.

In reviewing the raw data, I identify as outliers two contestants who submitted an extremely high number of entries (over 100) to a contest, and so exclude these two contestants from the whole panel. Internet Appendix A contains details on frequency of participation.

3.2.2 Futures Contract Prices

The trading platform provides participants' trading strategies with access to actual market prices for the investable universe of 88 futures, throughout both the Backtest and Live periods of contests. These market prices are also used to calculate trading strategies' performance.

I assemble historical price data for the same universe of 88 futures, and later use these prices in two ways: firstly, to construct a benchmark portfolio that I will use to compare contestants' performance outcomes against, in order to enable valid comparisons between contestants at different time periods in situations where panel time effects cannot be employed. And second, I use signals derived from these prices to estimate a model of investor behavior, in order to measure the extent to which price-based predictive variables are ignored by investors.

Internet Appendix D provides full details of the universe of 88 futures contracts, the data download procedure, and my procedure for constructing a benchmark portfolio using the futures' historical prices.

3.2.3 Timeseries of Daily Returns and Volume for Trading Strategies

I supplement the comprehensive contest leaderboard panel with two additional datasets. Quantiacs separately makes available the daily returns of trading strategies, which I merge with the leaderboard panel. And, while granular portfolio positions remain confidential and thus

unobservable, I can observe two daily positioning indices per trading strategy: the aggregate long positioning $Long_{i,\tau}$ and the aggregate short positioning $Short_{i,\tau}$. These are computed by the platform based on portfolio weights $w_{i,j,\tau}$ as $|Long_{i,\tau}| + |Short_{i,\tau}| = \sum_j |w_{i,j,\tau}|$. I use these measures to indexes for daily trading volume (or turnover) as

$$Volume_{i,\tau} = \left|Long_{i,\tau} - Long_{i,\tau-1}\right| + \left|Short_{i,\tau} - Short_{i,\tau-1}\right|. \tag{5}$$

3.2.4 Macroeconomic Signals and Their Release Dates On the Trading Platform

The additional set of predictive signals added to the platform in between contests consists solely of macroeconomic indicators. Internet Appendix E provides a full listing.

I use the exact values of these macroeconomic variables as they were accessible to investors' trading strategies when running on the Quantiacs platform. I am also able to identify precisely which date each macroeconomic variable release was made available on the platform (for both backtesting & live trading periods); I can therefore identify exactly which trading date return is the first to incorporate the information contained in each set of macro variable updates.

[Insert Figure 4 here]

Figure 4 is a stylized illustration of the two dimensions of time (t, τ) in the Quantiacs institutional setting. Investors enter a trading contest, indexed by t. Macro indicators were only accessible on the trading platform to investors and their trading strategies from contest $t \geq 8$ onward. The trading platform feeds each trading strategy with a set of daily input predictive signals extending back to January 1990, and machine traders return a set of daily portfolio weights in response to each. This trading strategy calendar time is indexed by τ, and comprises both the backtest/in-sample and live/out-of-sample calendar periods.

I later use the exact macroeconomic variable release dates to analyze the responses of machine-driven trading strategies to this new information. Separately, I also estimate a model of investor learning to measure how many predictive variables are ignored entirely by investors and their trading strategies.

3.3 Institutional Variation in Macro Release Day Informativeness

When the Quantiacs platform updates the values of the macroeconomic variables that are available as inputs to trading strategy code, they are done so on a single calendar date per month; thus, these release dates do not necessarily correspond to macro variable announcements by statistical agencies. Not all variables are updated every month, and this is particularly

true for early periods in the backtest history, when a number of variables were not backfilled. Figure 5 illustrates the variation in the number of macroeconomic variables that are released by the Quantiacs platform to trading strategies running on the platform, in trading strategy calendar time τ.

[Insert Figure 5 here]

This institutional variation in the timing of macroeconomic variable releases allows me to measure how strongly machine-driven trading strategies react to informative variables; on some days, more predictive information is released, and on other days, less predictive information is released. If a trading strategy makes use of informative variables, it should turn over its positions more on more informative macro release days than on less informative release days.

Consistent with trading activity revealing machine traders' responses to information releases, Internet Appendix F shows that investors' trading strategies exhibit significantly different responses on macro release days when compared to non-macro release days. As a placebo, I repeat the analysis for strategies entered into early contests, when macro predictive variables were unavailable, and as expected find no significant response.

3.4 Model and Performance Heterogeneity

Differences in Performance. Table 4 summarizes investors' in-sample and out-of-sample performance during all twelve trading contests. When entering a trading strategy into a contest, most investors will have implemented a strategy that has performed well over the recent backtest period: the contest-level median in-sample Sharpe Ratio (SR) ranges from +0.38 to +1.19 units of SR. Most investors therefore *appear* to benefit from machine-driven trading strategies when their performance is assessed on past market data alone.

[Insert Table 4 here]

However, their performance degenerates out-of-sample – i.e. during contest Live periods – when trading strategies can no longer be modified by investors. During those periods, the contest-level median SR ranges from only -1.21 to +1.00 units of SR. Furthermore, the median is positive for only a third of all contests, suggesting that most investors typically do *not* benefit from adopting machine trading strategies. Better-performing investors can benefit considerably: the top quartile of performance in contest Live periods is typically positive, and this upper quartile level can range as high as 2.07 units of SR. Similarly, the top decile performers in each contest Live period always succeed in attaining a positive SR out-of-sample; in ten of

the contests, these top performers consistently exceed 1.4 units of SR.

Comparing investors over the same periods reveals that the dispersion in investor performance is substantial. For the out-of-sample SRs obtained during contest Live periods, the inter-quartile range for these is typically 2 to 3 units of SR. That is, the 75th-percentile investor typically obtains a SR that exceeds the 25th percentile investor's by over 2 units during the same contest Live period. Sharpe Ratios vary over time, being (scaled) returns, and yet this dispersion persists over much of the 2014-2019 period, in which contest Live periods run.

[Insert Figure 6 here]

Figure 6 shows that trading activity is also heterogeneous. I compute the pairwise correlation between the daily aggregate turnover series of every contestant's best live entry in every contest and the equivalent turnover series for every other contestant's best live entry in the same contest, excluding self-comparisons: I label this $\rho_{i,j,t}$ for trading strategies i, j in the same contest t. The figure displays the contest-level means of these pairwise correlations in trading activity. The yardstick of perfect agreement would correspond to a mean pairwise correlation of +1 in each contest; instead, the levels never exceed 0.4. This alternative characterization of disagreement based on trading activity also suggests that investors disagree strongly.

Differences of Opinion. The observed within-contest dispersion in performance has an important implication: trading strategies with access to precisely the same dataset still vary widely in their attained performance. When conditioning on any one contest index, investors have access to exactly the same set of predictive variables and their historical realizations; therefore, conditioning on the contest also conditions on investors' information sets. Since investors in the same contest are symmetrically informed, their disagreement is likely to stem from different interpretations of exactly the same data. Cookson and Niessner (2020) term this phenomenon "model disagreement," to distinguish it from disagreement that is due to heterogeneous information.

I now show that model disagreement is indeed considerable by measuring how investors react to identical predictive signals. My analysis takes advantage of institutional variation in the information content of macro release days.

3.4.1 Measuring Variable Informativeness

To make a comparison between turnover and macro variable informativeness, I require a measure of the predictive information content of different macro release days τ. To do so, I use a methodology from the literature on interpretable machine learning that allows individual

variable informativeness to be measured and then aggregated to groups of variables.

Benchmark Predictive Model. I begin by defining a benchmark machine learning model to detect out-of-sample return predictability. I use the Random Forest algorithm (Breiman, 2001) as my estimation technology; this algorithm has proven successful in detecting out-of-sample equity return predictability (Gu, Kelly, and Xiu, 2020), and is therefdore a suitable benchmark to compare human behavior to. As this empirical literature shows, a Random Forest can capture flexible and nonlinear predictive relationships, and is capable of incorporating the full set of high-dimensional macroeconomic predictive variables in my setting without overfitting. I set up the following cross-sectional return prediction problem, where i indexes a futures contract and τ a month:

$$r_{i,\tau+1} = \mathbb{E}_t[r_{i,\tau+1}] + \epsilon_{i,\tau+1}, \tag{6}$$

with the conditional prediction for each next-month return

$$\mathbb{E}_\tau[r_{i,\tau+1}] = f(\boldsymbol{x}_\tau) \tag{7}$$

given by a flexible nonparametric mapping f estimated by the Random Forest algorithm, and which uses a vector $\boldsymbol{x}_\tau$ of predictive variables available as of the end of month τ.

I define the predictive variables $\boldsymbol{x}_\tau$ as the last 3 lags (on a monthly basis) of all 54 macroeconomic predictive variables available on the Quantiacs platform. Missing values are imputed with the historical median and each series is standardized to have zero mean and unit standard deviation. Both imputation and standardization only use data up to 2013 to avoid introducing any look-ahead bias to the OOS (out-of-sample) predictions that are generated for 2014 onwards (the period in which Quantiacs contest Live periods fall). In addition, one dummy variable for each futures contract is included in the set of input variables $\boldsymbol{x}_\tau$. This is somewhat analogous to including fixed effects in a regression but allows much more flexibility: including these dummies allows an estimated Random Forest to nest up to one predictive sub-model per futures contract if necessary.

As I will be making implicit comparisons between the performance of machine traders and my benchmark algorithm, it is important to note that this comparison is a fair one, by design. First, the benchmark machine learning algorithm that I construct uses only data $\boldsymbol{x}_\tau$ that is accessible on the Quantiacs platform to contestants. Second, I use a software package that was available to Quantiacs participants to conduct my estimation (`scikit-learn` (Pedregosa et al., 2011)); in principle, therefore, participants could have conducted exactly the same analysis I do. Third, the Random Forest algorithm I use is well-known and appears in many

introductory articles and books on applied machine learning;[10] it is not an obscure algorithm that Quantiacs participants would be unlikely to use. Finally, I take care to avoid look-ahead bias, as I describe next.

My Random Forest estimation and tuning procedure is similar to that of Gu, Kelly, and Xiu (2020). I estimate predictive models at an annual frequency: for example, to predict futures returns for any month in 2014, I use a model that has been trained up to the end of 2013. The training procedure involves using a validation set for the 5 year period preceding the test set – for example, to produce 2014 forecasts, the validation set would run from 2009 to 2013. Multiple models are estimated for various sets of hyperparameters,[11] and evaluated on the validation set. After picking the optimal set of hyperparameters, the model is then re-estimated for the period 1990 up to the year before the test set – for example, to produce 2014 forecasts, the model has been estimated from 1990 to 2013. This ensures that the test set forecasts (for example, 2014) are true OOS predictions. For the next year of the test set (for example, 2015), the entire procedure is repeated with an expanded training set and a validation set that has rolled forward by one year.

In this manner, monthly OOS predictions are generated for the period 2014-2019. The Quantiacs platform prevents models from improperly using OOS data in order to avoid look-ahead bias. This careful procedure is therefore necessary to accurately capture genuine OOS predictability, which a Quantiacs contestant must try to exploit in order to achieve her objective of maximizing her OOS Sharpe Ratio.

My procedure for constructing a benchmark predictive model that uses all the macroeconomic predictors available on the Quantiacs platform successfully estimates a model that detects out-of-sample return predictability. As one example of how this predictability can translate into portfolio choice, I cross-sectionally rank predicted returns and form a monthly-rebalanced portfolio that goes long futures contracts in the highest quintile of predicted next-month returns and short the lowest quintile of predicted next-month returns, with equal weightings to futures contracts. Figure 7 plots the cumulative out-of-sample returns of following such a trading strategy. This example strategy earns a cumulative OOS return of 40.91% and an OOS Sharpe Ratio of 0.57 over the period (assuming a zero riskless rate for simplicity).[12]

[10] As one example out of many, an article on Coursera titled *"7 Machine Learning Algorithms to Know: A Beginner's Guide"* introduces the Random Forest algorithm.

[11] Refer to Gu, Kelly, and Xiu (2020, Internet Appendix Table A.7) for further details.

[12] It is possible that these results serve as a lower bound to the extent of macroeconomic variable-based predictability, since the input variables consisted solely of the (lags of) the levels of these variables, and so constructing further variables out of these may increase the level of measured OOS predictability. Similarly, more sophisticated machine learning methods could be employed for the task of measuring the extent of return predictability using these variables. It is not the goal of this study to examine OOS predictability among futures contracts, but my

[Insert Figure 7 here]

Informativeness of individual macro variables. I next measure the informativeness of individual variables in the predictive model input $\boldsymbol{x}_\tau$ using SHAP values (Lundberg and Lee, 2017).[13] These measures are computed at an individual variable level. Define $SHAP(k,m)$ as the SHAP value of individual predictive variable k for data sample m; a sample consists of the futures contract and prediction month. The SHAP value of a variable k measures the predictive power of variable k for the individual prediction indexed by m; like the prediction, the SHAP value is in units of return. Note that the SHAP value depends on the estimated model used, and these will be matched according to the contest taken part in, making sure to avoid any look-ahead bias.

Information content of macro release days. SHAP values have a crucial property for my purposes: they are additive (Lundberg and Lee, 2017). Recall that each macro release day τ is associated with a set of simultaneously-released variables $\mathcal{R}(\tau)$ (and a complementary set of variables that are *not* updated during that macro release day). Thanks to the additivity of SHAP values, a single SHAP value per macro release day τ and data sample m can be computed by summing up the SHAP values of the variables released that day:

$$SHAP(\tau,m) = \sum_{k\in\mathcal{R}(t)} SHAP(k,m). \tag{8}$$

I then compute an informativeness index for each macro release day τ as the mean absolute SHAP value across all M data samples:

$$\text{Informativeness}_\tau = \frac{1}{M}\sum_m \Big|SHAP(\tau,m)\Big|. \tag{9}$$

Equation (9) measures the extent to which the set of variables $\mathcal{R}(\tau)$ released on a macro release day τ have predictive power across the entire cross-section of futures contracts, irrespective of sign, and for all historical dates.

3.4.2 Explaining Trading Heterogeneity with Model Disagreement

Armed with my index (9) of the predictive informational content of macro release days, I measure the sensitivity of different trading strategies to identical predictive signals. Using

findings suggest such a study could be informative; prior work by Freyberger, Neuhierl, and Weber (2020), Gu, Kelly, and Xiu (2020), and others has mostly focused on the cross-section of US stocks.

[13] "SHAP" stands for "Shapley Additive Explanations." SHAP values are a popular technique for explaining predictive variable importance for estimated machine learning models, with a number of desirable properties. In the finance literature, Abis (2020) uses SHAP values to compare the relative importance of predictive variables.

Equation (5), I compute the daily trading volumes of individual strategies i on historical calendar days τ. I then standardize this volume within-strategy, so that these z-scores are in comparable units across strategies. Then, for each strategy i, I conduct the following individual time series regression over the subset of macro release days τ:

$$\text{Volume}_{i,\tau} = \alpha_i + \beta_i \times \text{Informativeness}_\tau + \epsilon_{i,\tau} \tag{10}$$

Each estimated coefficient $\widehat{\beta}_i$ captures the sensitivity of trading strategy i to the release of the predictive information contained in macroeconomic indicators.

Next, I conduct a cross-sectional analysis of the extent to which different interpretations of the same signals can explain heterogeneity in trading strategies' overall trading patterns. I conduct pairwise regressions that are variants of the specification

$$\underbrace{-\rho_{i,j}}_{\substack{\text{Trading} \\ \text{dissimilarity}}} = \gamma \times \underbrace{\delta(\widehat{\beta}_i, \widehat{\beta}_j)}_{\substack{\text{Disagreement} \\ \text{about macro} \\ \text{predictive info.}}} + \phi_t + \epsilon_{i,j}, \tag{11}$$

where $\rho_{i,j}$ measures the correlation between the daily turnover series of two trading strategies i, j that are matched based on being entered into the same trading contest t; self joins are excluded ($i \neq j$). The negative of this pairwise correlation therefore measures the *dis*similarity in trading activity between i, j. The trading response coefficients $\widehat{\beta}_i, \widehat{\beta}_j$ are estimated according to Equation (10).

It remains to measure the dissimilarity in trading response coefficients, $\delta(\widehat{\beta}_i, \widehat{\beta}_j)$, which captures how much a pair of trading strategies i, j disagree about the predictive content of the same macroeconomic indicators. I begin by computing the absolute distance $|\widehat{\beta}_i - \widehat{\beta}_j|$ between coefficient values. Since this produces a skewed measure, I specify three log-like transformations: the natural logarithm (of non-zero absolute values), the log of one plus the absolute value, and the inverse hyperbolic sine (arcsinh) of the absolute value.

[Insert Table 5 here]

Table 5 presents regression results for each of the variants of Eqn. (11). As might be expected, the $\widehat{\gamma}$ coefficient estimates are positive (and significant), suggesting that disagreement in trading activity tracks differential interpretations of the same information. Importantly, the fractions of variation explained by all specifications (i.e. R^2 values) are substantial. Specifications without contest fixed effects (in columns 1, 3 & 5) all explain just over a third of total variation in trading activity. The fraction increases even further when contest fixed effects are

added (in columns 2, 4 & 6), and within-R^2 values are of a similar magnitude to the overall fractions of variance explained.

These results confirm that a substantial portion of the heterogeneity in machine-initiated trading activity is due to different interpretations of the same information; i.e. model disagreement (Cookson and Niessner, 2020), or differences in opinion. The true extent to which model disagreement drives trading activity is most likely even higher, since this analysis excludes differential responses to price-based informational content (which is more challenging to measure). It follows that, in the setting studied by this paper, investors specify widely different predictive models of the world. I shortly examine the process by which they do so.

4 Signal Usage in Trading Contests

In this study's institutional setting, the role that human investors play is limited to designing machine traders.[14] To understand more deeply why experience drives performance, I examine how investors design machine trading strategies, and how this varies with investor experience. As all machine traders are constrained to use the same set of input predictive signals, a crucial margin of adjustment is along the weights assigned by the human designers to the different predictive signals ingested by their machine trading strategies. Investors have a lot of leeway in how they make use of the available data; for example, they may implement machine learning models as part of their trading strategies, or they might manually specify portfolio weights that depend on a subset of available variables.

In order to quantify how these input predictive signals are used, I examine how the observed actions taken by machine traders vary in response to changes to the input signals that the machines receive from the trading platform.[15] Using my index of the time-varying predictive power of macroeconomic indicators, I test whether these genuine predictors are used by both inexperienced and experienced investors when designing machine traders. Next, and in a similar manner, I test whether investors make use of variables that are familiar to them – independently of the variables' predictive power.

[14] As described in Section 3.1 and Internet Appendix C, human investors write computer code to design a systematic trading strategy, submit the strategy to enter into a trading contest, and then simply wait for the contest's Live period to end (with no further ability to affect the outcome).

[15] This approach is analogous to how one typically studies the behavior of human investors in response to market conditions or specific events. In empirical studies of investor trading or portfolio choice behavior, one typically cannot interrogate the investors in question on their trading motives (or peer into their thoughts) to understand the drivers of their actions. I face a similar situation in studying the operation of machine trading strategies, as I cannot observe the (confidential) computer code that drives their actions. In both cases, however, inferences can successfully be drawn from observed actions.

4.1 Usage of Genuine Predictive Signals

Given my measure of the information content of macro release days (9), I now analyze the responses of machine trading strategies to this predictive information content. Following Eqn. (5), I calculate daily turnover measures for investors' best live trading strategies. I standardize these volume values within the strategy level; i.e. each strategy i has a mean turnover of zero, and units in daily standard deviation. I then select the subset of daily turnovers that correspond to macro release days; i.e. I analyze days τ when the values of (one or more) macroeconomic predictive variables were updated on the Quantiacs platform. Using this sample, I conduct regressions of the following form:

$$\begin{aligned}\text{Volume}_{i,\tau} = \beta_1 \times \text{Experienced}_{i,t} + \beta_2 \times \text{Informativeness}_{\tau} \\ + \beta_3 \times \text{Experienced}_{i,t} \times \text{Informativeness}_{\tau} + \phi_t + \epsilon_{i,t,\tau}\end{aligned} \tag{12}$$

As the daily volume standardization step includes demeaning each strategy's daily turnover, this is similar to including strategy fixed effects that control for any trading strategy-level unobserved variables (including unobserved investor skill) that might confound an analysis. Another benefit to the standardization step is that turnover magnitudes are made to be comparable across strategies – in units of strategy turnover standard deviations (i.e. z-scores) – and thus data from multiple strategies can be pooled together in a single regression. The aggregate informativeness of macro release days, $\text{Informativeness}_{\tau}$, is calculated according to Eqn. (9). I take care to compute the informativeness index using only data up to the end of the year *before* the beginning of the live period of the current contest t. This ensures the informativeness measure does not contain information that was unavailable to Quantiacs participants at the time of the contest. Model year fixed effects ϕ_t control for different mean values of the informativeness measure that correspond to such different endpoints. A dummy for experience takes the value 1 if the human investor who implemented the strategy is taking part in her third (or higher) contest at that point.

[Insert Table 6 here]

Table 6 presents regression results for (standardized) daily turnover on macro release days against the aggregate information content of these macro release days. The first two columns show baseline results. Column (1) shows a positive and significant value for the β_1 coefficient when it is estimated alone, indicating that experienced investors design trading strategies that trade more than usual on macro release days. Column (2) shows a positive and significant coefficient for the β_2 coefficient when it is estimated alone, indicating that investors design trading strategies that tend to respond more strongly to the release of more informative macro

variables; this estimate does not break out experienced and inexperienced investors.

The estimates from the full specification in column (3) show that experienced and inexperienced investors' trading strategies respond differently to the information content of macro release days. The estimate of $\widehat{\beta}_2 = 0.6348$ suggests that inexperienced investors design trading strategies that respond more to more informative macro release days, but the effect is only significantly different to zero at the 90% level. As for experienced investors, the estimates for the baseline coefficient of $\widehat{\beta}_1 = -0.3229$ and interaction effect of $\widehat{\beta}_3 = +1.025$ suggest that experienced investors respond less strongly to completely uninformative macro variable release days, and more strongly the more informative a macro release day is. That is, the trading volumes of machine-driven strategies designed by experienced investors are much more sensitive to the information content of macro release days than of those designed by inexperienced investors. The near-insignificance of the $\widehat{\beta}_2$ estimate for inexperienced investors even calls into question whether inexperienced investors respond at all to informative macro variable releases; there is no such ambiguity for the subset of experienced investors.

I conclude that experienced investors make better use of informative predictive variables than inexperienced investors do when specifying their models of the world. This occurs despite both groups of human designers sharing an identical opportunity to detect such informative predictive variables (thanks to the controlled Quantiacs setting).

4.2 Usage of Familiar Signals

If inexperienced investors do not use genuinely predictive macroeconomic indicators – or respond only weakly to their release – then what variables *do* they favor?

Among the prior studies on (in)experience and behavioral biases, Abreu, Mendes, and Santos (2011) find that inexperienced investors are prone to one form of familiarity bias: a home bias in their portfolio choices. It is possible that a familiarity bias such as this can influence how investors specify their models of the world. I therefore hypothesize that inexperienced investors use variables with which they are familiar – rather than determining which variables are more informative. If this familiarity bias holds investors back from maximizing their performance outcomes, and lessens with experience, this mechanism may explain why investor experience drives improved performance outcomes. I now test this hypothesis.

4.2.1 Measuring Signal Familiarity

To proxy for how familiar a macroeconomic indicator k is perceived to be by participants, I count how frequently it is mentioned in different forms of media. Specifically, I count its fre-

quency of appearance in (i) news articles indexed by the Factiva database, and (ii) published books indexed in the Google Books database. Both these databases allow me to search for occurrences of specific phrases in specific time periods, and I conduct annual searches for each of the 54 macroeconomic variables provided by the Quantiacs platform. These two comprehensive sources result in different (but correlated) counts for the number of mentions of each macroeconomic variable.

To calculate an aggregate familiarity index for each macroeconomic release date τ, fixing a data source and search period, I compute the aggregate familiarity of each macroeconomic indicator k among the set of macroeconomic indicators $\mathcal{R}(\tau)$ released together on date τ by the Quantiacs platform:

$$\text{Familiarity}_{\tau} = \frac{1}{\text{Total mentions}} \sum_{k \in \mathcal{R}(\tau)} \text{Mentions}(k). \tag{13}$$

Equation (13) measures the total share of media mentions of the subset of macroeconomic signals that is released by the Quantiacs platform at date τ. As described shortly, I match the annual period over which shares are computed to the corresponding trading contest when conducting my analyses.

Unlike my measure of macro release day informativeness, these aggregate familiarity indices are empirically concentrated, with the most familiar macroeconomic variables accounting for a large fraction of total media mentions; further details are provided in Internet Appendix A. Despite the distinct sources, the two familiarity indices that I construct using Equation (13) are highly correlated with one another, with a magnitude of 0.79. The familiarity indices covary much less strongly with the index of macro release day informativeness, which has a correlation of only 0.15 and 0.21 with the familiarity indices derived from news articles and books, respectively. I use both macro release day familiarity indices in the subsequent analyses to ensure my results are not sensitive to any one source.

4.2.2 Reliance on Familiar Signals

I now analyze the responses of machine trading strategies to investors' total familiarity with the variables released on macro release days. I prepare standardized daily trading strategy turnover values on macro release days, similar to Section 4.1. Where i denotes a trading strategy, τ a macro release day, and t a trading contest, I conduct regressions of the following

form, for each type of familiarity index:

$$\begin{aligned}\text{Volume}_{i,\tau} = \beta_1 \times \text{Experienced}_{i,t} + \beta_2 \times \text{Familiarity}_{\tau} \\ + \beta_3 \times \text{Experienced}_{i,t} \times \text{Familiarity}_{\tau} + \phi_t + \epsilon_{i,t,\tau}\end{aligned} \tag{14}$$

When calculating familiarity indices using Equation (13), I take care to restrict my search of news article or book mentions to the year in which the end of the backtest/in-sample period of the corresponding trading contest t falls; this ensures that I am measuring the familiarity of macro variables at the time investors are designing their trading strategies. Fixed effects for these benchmark periods ϕ_t absorb any variation in the annual means of familiarity indices. Note that daily volume is standardized (i.e. a z-score) as before, which once again implicitly controls for strategy-level (and investor-level) unobserved heterogeneity, and allows samples to be pooled across strategies.

[Insert Table 7 here]

Table 7 presents regression results of daily strategy turnover against the aggregate familiarity of those macroeconomic indicators being released to investors. The estimated coefficient $\widehat{\beta}_2$ takes approximately the same value across all specifications; that is, whether an interaction with investor experience is incorporated or not, and whether the measure of aggregate familiarity is computed based on news article or book mentions. All estimates for $\widehat{\beta}_2$ are significantly non-zero at the 99% level. By contrast, none of the estimates for $\widehat{\beta}_3$, the interaction effect between familiarity and experience, are significantly different to zero. Therefore, all trading strategies respond more strongly to an increase in the familiarity of the macroeconomic indicators being released. This result is consistent with my hypothesis that investors first reach for more familiar variables when specifying their models of the world.

Table IA.2 in Internet Appendix B augments the previous regressions to include the index of macro release day predictive informativeness, as well as an interaction with experience. All estimates are of a similar sign and significance to those in the separate informativeness regressions (Table 6) and familiarity regressions (Table 7). Neither index drives out the other, and results are similar across both familiarity measures.

5 Consequences of Experience Effects for Performance and Model Heterogeneity

Given that inexperienced investors do not fully adopt predictive signals, I now test whether the benefits of predictive signal available concentrate among experienced investors. I also test a related hypothesis of whether experienced investors specify convergent predictive models that disagree less.

5.1 Performance Outcomes and Experience

I begin by testing whether experience drives improved performance. I measure two types of performance outcome: the in-sample (backtest period) best Sharpe Ratios and the out-of-sample (live period) best Sharpe Ratios of the contestants i for each contest period t. An investor's experience is measured by the number of contests she has participated in so far. I conduct regressions that are variants of the specification

$$\text{SR}_{i,t}^{Best} = \beta \times \text{Contests experienced}_{i,t} + \pi_i + \phi_t + \epsilon_{i,t}, \tag{15}$$

where π_i denotes investor (contestant) fixed effects and ϕ_t denotes time (contest) fixed effects. Controlling for these two dimensions of unobserved heterogeneity allows me to exclude their effect on observable performance outcomes; for example, unobserved investor-specific attributes (including demographic characteristics or skill levels) are taken into account.

[Insert Table 8 here]

Table 8 displays the results of regressing performance outcomes against the experience levels of the contestants. In both OLS and panel specifications, the positive (and significantly non-zero) coefficient β on investor experience indicates investors perform better with experience.[16] OLS and panel regressions both detect this learning dynamic; furthermore, the observed magnitude of this effect strengthens once fixed effects are included in columns (2) & (4). The dependent variable in columns (1)-(2) is the in-sample (backtest period) performance outcome, so the positive estimates for β in those columns suggest that investors make use of the historical data that is available to them – they are more able to increase their observed performance outcomes before entering the Live period of each contest. The dependent variable

[16]Note that this channel of learning with experience is distinct from a statistical effect of using additional samples to form more accurate forecasts: while the number of contests a single investor participates in does increase with calendar time (i.e. length of historical samples), investors with equal levels of experience may have participated in different contests with different lengths of historical datasets available at the time the contests occurred. Time/contest fixed effects control for this statistical effect.

in columns (3)-(4) is the out-of-sample (Live period) performance outcome, so the positive estimates for β in those columns are consistent with previous studies that find a learning with experience effect for the (different) setting of retail investors who trade stocks (e.g. Nicolosi, Peng, and Zhu, 2009; Seru, Shumway, and Stoffman, 2010).

5.2 Model Disagreement and Experience

I now test whether model disagreement lessens with experience, concurrently with the improvement in investors' performance. For this, I rely on the pairwise correlations between the aggregate daily turnovers of trading strategies i, j matched to the same trading contest t (the means of which were shown in Figure 6). I perform an additional matching/filtering step: I retain only individual pairwise correlations $\rho_{i,j,t}$ for which contestants i and j have the same level of experience. I then conduct variants of the following panel regression:

$$\rho_{i,j,t} = \beta \times \text{Contests experienced}_{i,j,t} + \pi_i + \phi_t + \epsilon_{i,j,t}. \tag{16}$$

Since a linear regression produces a conditional mean, this simple procedure allows me to relate average changes to the pairwise correlations between entries (LHS) to the (equalized) changes in the experience levels of the contestants who implemented the trading strategies (RHS). The fixed effects allow me to progressively control for two of the three dimensions along which the panel varies: the contest index t and one of the contestant indices i.

[Insert Table 9 here]

Table 9 displays the regression results. The simplest specification in column (1) produces an easily-interpretable measure of the mean pairwise correlation between the (machine-initiated) daily trading volumes of all participants with the same experience level. This proxy for model agreement ranges from 0.3 to 0.4.

Across all specifications (1)-(3), an increase in the number of contests experienced by (both) contestants i and j is associated with an increase in the correlations of their trading strategies' daily turnover, of around 6 to 9 percentage points. Experience thus resolves model disagreement, albeit gradually. This is consistent with experienced investors converging on a more accurate model of the world.

6 Gains from Access to Predictive Signals

Inexperienced investors respond only weakly to informative macro predictive variables, and tend to rely on familiar macro predictive variables. On net, do they benefit from macro predic-

tive variables at all? I answer this question by comparing investors who have access to macro predictive variables to investors who do not. I find that inexperienced investors do *not* benefit, on net, from having access to macro predictors. This finding is consistent with the presence of a behavioral bias that afflicts these inexperienced investors.

6.1 Performance Gains Due to Macro Signals

To cleanly measure the benefits that investors derive from data, I define two groups. The Control group consists of investors during contests 1-7, when the additional macro predictive variables were not available to them. The Treatment group consists of investors who only traded from contest 8 onward; that is, in the environment when the additional macro predictive variables were available to them. To be able to include fine-grained experience level dummies, I focus on investors who have experienced 1-4 contests.

[Insert Table 10 here]

The Quantiacs institutional setting implicitly controls for a variety of potential confounders by enforcing identical goals, horizons, and other common conditions upon the investors in my sample. In addition, Table 10 verifies that contestants are balanced across the Treatment and Control groups with respect to their observable characteristics: Panel A displays the means of a contestant's experience level, and a contestant's relative ranking in the previous contest she took part in (for contestants who participate in multiple contests). The differences are not statistically different to zero (even at a low significance level of, say, 90%). Similarly, Panel B compares contest-level participation attributes; namely, the average level of experience of contestants per contest, the fraction of first-time participants, and the fraction of last-time participants. The relevant differences are also not statistically different to zero (including at low significance levels such as 90%). The combination of implicit controls by the institutional setting and groups that are well-balanced across observables allows me to attribute changes to investors' performance outcomes to the availability of the macroeconomic predictive variables.

[Insert Table 11 here]

The regressions in Table 11 analyze the relationship between investor experience and the availability of additional predictive signals (as covariates) and the out-of-sample performance outcomes (as the response) using various specifications. The dummy variable named "New variables available$_t$" indicates whether investors compete after the introduction of the additional macroeconomic indicators to the common parts of investors' information sets; i.e. whether investors belong to the Treatment group. All regression specifications in Table 11 take

performance outcomes to be the excess of the out-of-sample SR over the SR from holding the benchmark index (equally-weighted buy-and-hold positions in contestants' tradeable universe of futures contracts) in order to permit valid comparisons across time periods.[17]

The regression results in Panel A of Table 11 are variants of the following linear specification:

$$\begin{aligned}&\text{Excess Live SR}_{i,t}^{Best}\\ &= \beta_1 + \beta_2 \times \text{Contests experienced}_{i,t} + \beta_3 \times \text{New variables available}_t\\ &\quad + \beta_4 \times \text{Contests experienced}_{i,t} \times \text{New variables available}_t + \epsilon_{i,t}.\end{aligned} \tag{17}$$

The estimate $\widehat{\beta}_4 = 0.908$ of the incremental effect of data availability on performance is positive and significant. This indicates that experience and deriving a benefit from Bigger Data are complements. The availability of the additional macro predictive variables is associated with a steepening in investors' performance dynamics, over and above the positive slope captured by the estimate $\widehat{\beta}_2 = 0.589$ for the baseline; in this case, the baseline represents the performance dynamics of investors without access to the new data.

To allay any concerns about assuming a linear relationship between experience and performance, I also relax the functional form by incorporating a dummy for every experience level, and find the effect continues to hold. The regression results in Panel B of Table 11 are variants of the following specification:

$$\begin{aligned}&\text{Excess Live SR}_{i,t}^{Best}\\ &= \beta_1 + \sum_{k=2}^{4}\left[\beta_k \times \mathbb{1}\{\text{Contests experienced}_{i,t} = k\}\right] + \gamma_1 \times \text{New variables available}_t\\ &\quad + \sum_{k=2}^{4}\left[\gamma_k \times \mathbb{1}\{\text{Contests experienced}_{i,t} = k\} \times \text{New variables available}_t\right] + \epsilon_{i,t}.\end{aligned} \tag{18}$$

In Panel B column (3), the treatment indicator is interacted with dummies for investor experience to capture the incremental effects γ_k of the availability of the new predictive variables on the performance outcomes of investors with comparable levels of experience k between groups. In column (3), the interaction between the dummy variable for treatment and dummies for investor experience have positive (and significantly non-zero) coefficients $\widehat{\gamma}_3$ and $\widehat{\gamma}_4$ for higher levels of experience, in particular. For lower levels of experience, however, $\widehat{\gamma}_1$ and $\widehat{\gamma}_2$ are not statistically significantly different to zero, confirming that inexperienced investors

[17] Note that since contestants are divided into two groups, I cannot identify individual-level fixed effects. Similarly, since the dummy for the availability of new predictors is time-invariant on a per-contest level, I can no longer identify contest-level fixed effects either; I therefore control for market conditions using the benchmark portfolio's performance.

fail to derive a net benefit from gaining access to the set of macro predictive variables.

To sum up, across both functional forms I find evidence that more experienced investors are successful in making use of the additional macroeconomic indicators to better attain their objective. By contrast, inexperienced investors do not derive any incremental benefit from having access to these newly-added macro variables.

6.2 Potential Alternative Explanations for the Effect

My tightly controlled institutional setting is the primary means by which I identify investor performance dynamics. In this section, I address potential alternative explanations so as to rule them out as the main drivers of my observed results.

6.2.1 Participation Bias

Having found that experienced individual investors derive a benefit from the availability of the macro predictive variables, I now examine the robustness of this finding to a possible selection bias between investors' experience levels and their performance outcomes. Unlike Seru, Shumway, and Stoffman (2010) and Linnainmaa (2011), modeling participation and selection is not a primary object of my study; however, attrition exists in my panel of investors, and it is therefore important to understand what direction any bias might introduce to OLS regression specifications. I therefore re-estimate the magnitude of experience effects using a Heckman (1976) two-stage model, with standard errors computed according to Greene (1981).

[Insert Table 12 here]

The first stage of the procedure consists of a probit model for the participation of the contestants in the next contest, based upon a number of covariates that I specify. I argue that two of the covariates that I make use of in the first-stage selection equation satisfy an exclusion restriction: the mean of the Google Search Index for "Quantopian"[18] and the *relative* number of entries that the contestant has made in the *prior* contest compared to the mean. The first-stage coefficients are shown in the top section of Table 12, and the estimates indicate that contestants are less likely to participate in the next contest as they gain in experience and as

[18] Quantopian was one of Quantiacs' rival FinTech platforms for running trading contests (Fleiss, Kominers, and Ughetta, 2017; Zheng, 2017). The logic of my exclusion restriction is that the contestants on each platform may exhibit substitutability in which platform they participate in. It seems implausible to argue that search interest in and attention to a rival platform (Quantopian) could affect performance outcomes in my focal setting (Quantiacs) through any channel other than an effect on participation. Note that the rival Quantopian platform is now defunct (as announced on 29 October 2020) but was active throughout my sample period.

the rival trading platform (Quantopian) gains in attention. Contestants are also more likely to participate again if they have recently participated more intensively compared to others.

The second-stage coefficients represent the outcome equation of interest, and these are shown in the middle section of Table 12. The coefficient on experience is positive throughout, and statistically significant both overall (in column 1), and when conditioning on the availability of the new predictive variables (column 3). These findings are reassuring that the earlier results on learning with experience are not in fact driven by selection/attrition bias or, as Seru, Shumway, and Stoffman (2010) framed it, an investor learning about her type.

More importantly, when conditioning on the availability of the new predictive variables in column (3), the estimate of the ρ parameter in the second stage of the Heckman (1976) model is negative, and therefore so is the corresponding coefficient on the Inverse Mills Ratio, which is also statistically significant. This indicates the presence of *negative selection* in an OLS estimate of the magnitude of the effect of learning with experience. To put it differently, uncorrected OLS estimates of the magnitude of the experience effect would be biased *downwards* for the investors with access to the additional predictive variables. The direction of this effect is in agreement with the intuition of Linnainmaa (2011). Furthermore, since the negative selection is concentrated among the investors who had access to the additional predictive variables, this suggests that the incremental steepening in performance dynamics due to additional data availability is likely to be even steeper than that detected by the earlier OLS regressions (Table 11 Panel A).

6.2.2 Competition Effects

I also test whether competition (to win one of the top 3 places in a contest) can explain my empirical findings on how investors make use of additional predictive variables.[19] Dugast and Foucault (2021) highlight that, while data abundance may benefit investors, the effect of competition may work in the opposite direction. In Internet Appendix H, I measure competitive intensity and analyze to what extend it impedes investors' ability to benefit from the additional macro predictive variables. The results of this analysis suggest that competition effects do not explain my findings.

[19] One advantage of the Quantiacs institutional setting is that investors have no price impact, which eliminates other channels for them to engage in strategic behavior.

7 Predictive Model Complexity

This section introduces a model of investor behavior that allows me to measure investors' usage of individual predictive signals. The framework that I introduce allows investors to express a bias against predictive variables, and shows that this bias can lead investors to completely ignore a subset of signals that are available to them. I also estimate this behavioral model: the results reveal an aversion to specifying complex predictive models, which lessens with experience.

7.1 Investor Behavior When Solving a Prediction Problem

7.1.1 Portfolio Choice with Historical Data

Since investors are incentivized by the contest setup to maximize their single-period out-of-sample Sharpe Ratios, the investor's objective for a particular contest is defined as

$$\max_{w} \frac{\boldsymbol{\mu}^T \boldsymbol{w}}{\sqrt{\boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}}}, \tag{19}$$

that is, to maximize the Sharpe Ratio of a portfolio of weights $\boldsymbol{w}$ on all tradable futures contracts given a vector $\boldsymbol{\mu}$ of expected returns of those futures contracts and a variance-covariance matrix $\boldsymbol{\Sigma}$. There are no shorting constraints, and I omit any discussion of transaction costs.

In my institutional setting, investors all aim to maximize the Sharpe Ratio of their portfolios during the Live period that the trading contest is active, so their common objective can be analyzed by appealing to the single-period mean-variance framework of Markowitz (1952).[20] To solve this portfolio choice problem and attain her objective, the investor needs to know the first two moments of the returns of her tradable assets during that future period.

In reality, a prediction problem arises because the out-of-sample moments of asset returns are not known. To make progress, let us assume that $\boldsymbol{\Sigma}$ is known and so the investor must therefore estimate the unknown parameter vector $\boldsymbol{\mu}$ in order to achieve her objective.[21] Investors in the Quantiacs setting behave consistently with this assumption: the results in Internet Ap-

[20] Note that, under relatively mild assumptions on feasible portfolios and constraint sets, the solution to the Sharpe Ratio maximization problem is simply the tangency portfolio solution to the classical Markowitz (1952) mean-variance portfolio choice problem (Cornuéjols, Peña, and Tütüncü, 2018, pp. 102-103), so there is a direct analytical link between the solutions of the Sharpe Ratio maximization problem and of the mean-variance portfolio choice problem.

[21] Empirically, it is well known that the first moment of asset returns is more difficult to estimate than the second moment. Theoretically, an asset's volatility can even be calculated exactly from continuously-observed returns. Merton (1980) discusses these empirical and theoretical considerations in detail.

pendix G suggest they draw their performance improvements primarily from improved mean returns, rather than decreases in their strategies' return volatilities.

Considering just one of those futures contracts, the investor must predict the (unknown) expected return μ in order to solve her portfolio choice problem (and to repeat the exercise for all other tradable futures contracts). She receives m signals $s_1, s_2, \ldots, s_m$ that can be used to produce a prediction $\hat{\mu}$. I begin by assuming that she is aware of the true functional form of the relationship and that it is some linear combination of the predictive signal values,

$$\mu = \sum_{i=1}^{m} b_i s_i = \boldsymbol{s}\,\boldsymbol{b}, \tag{20}$$

collecting the predictive signal values in a row vector $\boldsymbol{s}$ and linear coefficients in a column vector $\boldsymbol{b}$.

Despite knowing the true functional form, the parameters $\boldsymbol{b}$ themselves are unknown to the investor, so she must proceed by learning or estimating them based on the historical data of similar futures contracts. Recall that all futures contracts mature at predefined dates; for example, there are four S&P500 E-mini futures contracts traded on the CME that mature each year (at predefined dates in March, June, September and December). Therefore an investor who wishes to predict the expected return μ of a specific futures contract before it matures can make use of historical return data for similar futures contracts that have matured in the past.

More formally, to estimate these unknown parameters $\boldsymbol{b}$, the investor collects t previous realizations, each of which relates to the realized return v of some past futures contract and the corresponding prior signals available $s_1, s_2, \ldots, s_m$ at the time. Let us arrange these t sets of historical samples in t-dimensional column vectors $\boldsymbol{v}$ and $\boldsymbol{s}_1, \boldsymbol{s}_2, \ldots, \boldsymbol{s}_m$, respectively. For convenience, define the $t \times m$ data matrix $\boldsymbol{S} := [\boldsymbol{s}_1 \;\; \boldsymbol{s}_2 \;\; \ldots \;\; \boldsymbol{s}_m]$. Then the investor's prediction problem requires her to determine the values of these unknown parameters based on the historical data that she observes. Defining $\mathrm{Err} : \mathbb{R} \times \mathbb{R} \to [0, +\infty)$ as some measure of error or deviance between its scalar arguments, her objective during the model estimation process will be to minimize the error between the observed historical realized returns $\boldsymbol{v}$ and her predictions based on the historical signal values $\boldsymbol{S}$:

$$\min_{\boldsymbol{b} \in \mathbb{R}^m} \mathrm{Err}(\boldsymbol{v}, \boldsymbol{S}\boldsymbol{b}) \tag{21}$$

Minimizing the errors in the predictions of these historical realized returns (21) directly improves the investor's ability to attain her objective: in fact, Best and Grauer (1991) showed analytically that both over- or under-estimating assets' expected returns can lead to large changes

in the mean or variance of the solution to the closely-related Markowitz (1952) portfolio choice problem. In my setting, then, both over- and under-estimates of the expected returns may lead to sub-optimal Sharpe Ratios. Knowing this, an investor who is learning to predict expected returns should seek out a functional form for the error/deviance function in (21) that penalizes both over- and under-estimates. Minimizing the root-mean-square error fulfills this goal. I therefore define this to be the (unbiased) investor's baseline prediction model. In what follows, $||\cdot||_1$ and $||\cdot||_2$ denote the ℓ_1/taxicab and ℓ_2/Euclidean norms of a vector, respectively.

Definition 1. *An unbiased investor determines the unknown parameters $\boldsymbol{b}$ by minimizing the historical prediction error*

$$\min_{\boldsymbol{b}\in\mathbb{R}^m} ||\boldsymbol{v} - \boldsymbol{S}\boldsymbol{b}||_2. \tag{22}$$

The (well-known) solution to the above problem is for the investor to run an OLS regression on historical values to compute $\widehat{\boldsymbol{b}}$, and then predict the expected return according to Eqn. (20).

7.1.2 Incorporating Irrationality Into the Investor's Prediction Problem

In my characterization, the investor has well-defined preferences relating to Sharpe Ratio maximization, as required by the Quantiacs institutional setting. Much like the scenario analyzed by Martin and Nagel (2022), she faces a prediction problem, in predicting futures contracts' expected returns from historical data. I now show how notions of irrationality can be incorporated into this scenario by adding a key deviation from the rational baseline.

I continue to maintain the baseline functional form for the futures contract's expected return (20), and the assumption that the investor has access to a dataset of historical signals and their realizations for similar futures contracts. I now assume, in addition, that the investor falls victim to a form of familiarity bias by expressing an aversion to unfamiliar variables. In a similar spirit to Cao, Han, Hirshleifer, and Zhang (2011), I capture this aversion to the unfamiliar as a fear of worst-case outcomes.

Definition 2. *A biased investor determines the unknown parameters $\boldsymbol{b}$ by considering the worst-case error that may result for any given choice of parameter values,*

$$\min_{\boldsymbol{b}\in\mathbb{R}^m} \max_{\boldsymbol{U}\in\mathcal{U}} ||\boldsymbol{v} - (\boldsymbol{S}+\boldsymbol{U})\boldsymbol{b}||_2, \tag{23}$$

where $\boldsymbol{U}$ is a matrix of signal-wise perturbations that maximizes the ℓ_2 norm-based error for any choice of $\boldsymbol{b}$ and is constrained by an uncertainty set

$$\mathcal{U} := \left\{[\boldsymbol{u}_1\ \boldsymbol{u}_2\ \ldots\ \boldsymbol{u}_m] : ||\boldsymbol{u}_i||_2 \leq \delta_i\ \forall\ i = 1,\ldots,m\right\} \tag{24}$$

that is characterized by a set of variable-specific upper bounds $\delta_i \geq 0$, perceived by the investor, on the ℓ_2 norm of each possible signal-wise disturbance $\boldsymbol{u}_i$.

Definition 2 frames the biased investor's prediction problem as her attempting to minimize the worst-case prediction error while a malevolent opponent (Nature) conspires to maximize it up to the constraints permitted by $\mathcal{U}$. The higher the variable-specific bounds δ_i perceived by the investor, the worse the worst-case error and the more conservative she will be. An investor who is averse to some unfamiliar variable indexed by x will act as if she expects Nature to introduce disturbances up to magnitude δ_x to the historical values $\boldsymbol{s}_x$ of this variable in exactly such a way that it would maximize the investor's overall prediction error.

I will shortly describe a simplified version of this setup that delivers a closed-form solution and the intuition for why the biased investor will underweight – and may even ignore – current signal values, even though they determine the asset's expected return by definition (Eqn. 20). An investor who behaves according to Definition 2 is thus consistent with this paper's empirical findings that investors may choose to underweight even genuine predictive variables.

Incorporating worst-case scenarios into the investor's objective function (23) has a parallel with the setup of a zero-sum game, or indeed the well-established literature on ambiguity aversion. Notably, Epstein and Schneider (2008) and Illeditsch (2011) model the behavior of an investor responding to a (single) signal of unknown precision. Another related work by Garlappi, Uppal, and Wang (2007) models ambiguity aversion in portfolio choice as a constraint on the investor's perceived (squared) confidence interval of the expected return of an asset. There are also important differences between my approach and this literature. First, the investor's prediction problem in my case specifically incorporates a (crucial) role for historical data. Second, I do not rely on any probabilistic assumptions – not even the existence of a random variable. Third, as I will show shortly, investor behavior in my setup can easily be taken to the data thanks to strong connections with the machine learning literature.

I now show that, given an investor-specific penalization parameter, the biased investor's problem is analogous to that of running the well-known Lasso machine learning technique on the historical data that she possesses.

Assumption 1. *Assume that $\delta_i = \delta \ \forall \ i = 1, \ldots, m$.*

Proposition 1. *Under Assumption 1, a biased investor solves her original problem* (23) *by solving an equivalent formulation*

$$\min_{\boldsymbol{b} \in \mathbb{R}^m} \frac{1}{2} ||\boldsymbol{v} - \boldsymbol{S}\boldsymbol{b}||_2^2 + \lambda ||\boldsymbol{b}||_1, \tag{25}$$

where $\lambda \geq 0$ *is a scaling of* δ *in* (24).

The Lasso prediction problem (25) has been studied extensively in the machine learning literature. As a model of investor behavior, it connects the magnitude of the investor's familiarity bias λ with the extent to which she underweights genuine predictive variables. I now make a further assumption that delivers closed-form predictive variable weights to make this point clearly.

Assumption 2. *Assume that the data matrix* $\boldsymbol{S}$ *is orthonormal:* $\boldsymbol{S}^T\boldsymbol{S} = \boldsymbol{I}$.

Proposition 2. *Under Assumption 2, the biased investor will solve her prediction problem* (25) *by predicting*

$$\widehat{\mu} = \boldsymbol{s}\,\widehat{\boldsymbol{b}}, \tag{26}$$

where $\boldsymbol{s}$ *is an m-dimensional row vector consisting of the m predictive signals to the currently traded asset's expected return* μ*, and* $\widehat{\boldsymbol{b}}$ *is an m-dimensional column vector of weights whose elements are defined by*

$$\widehat{b}_k = \text{sign}(\boldsymbol{s}_k^T\boldsymbol{\nu})\max\left\{|\boldsymbol{s}_k^T\boldsymbol{\nu}| - \lambda, 0\right\}. \tag{27}$$

Figure 8 illustrates the functional form given by Equation (27). Over the region $|\boldsymbol{s}_k^T\boldsymbol{\nu}| \leq \lambda$ of the input domain, a biased investor will entirely ignore a predictive variable k by choosing the weighting parameter $\widehat{b}_k = 0$. The larger the magnitude of the investor's bias λ, the more likely that she will ignore any predictive signals. This simplified setting corresponds to an investor who applies her familiarity bias equally to all predictive variables, although all variables k will differ in their historical predictive contributions $\boldsymbol{s}_k^T\boldsymbol{\nu}$ and so will still be treated differently by the investor.

[Insert Figure 8 here]

7.1.3 Consequences of Investor Bias

Under-reaction & model mis-specification. One interpretation of the biased investor's model of the world (27) is to note that $\boldsymbol{s}_k^T\boldsymbol{\nu}$ would simply be the parameter value that the decision-maker would have specified *if* she had used OLS regression to solve her problem, under Assumption 2. Therefore, the investor specifies parameter values $\widehat{\boldsymbol{b}}$ that are "shrunken" in comparison to what they would have been had she used OLS regression; this is an example of "soft-thresholding," as illustrated in Figure 8. Even when predictive variables are not ignored entirely, the soft-thresholding effect reduces their magnitudes in comparison to the correctly-specified model of the world (Definition 1) in which the investor would have simply run an OLS

regression on her historical dataset. The investor's familiarity bias thus leads her to under-react to predictive signals.

Market (in)efficiency & return predictability. The familiarity bias that I investigate in this section has implications for market efficiency. In the most striking case, Proposition 2 implies that any signal s_k to the asset's expected return may not find its way into the price if the historical signals $\boldsymbol{s}_k$ contributed to predicting (historical) expected returns by an amount that falls below some threshold. Even when all variables are incorporated, market inefficiency may result — to see this, consider the case of a single representative investor with a λ parameter that captures her bias. This representative agent setup corresponds closely to Martin and Nagel (2022)'s model of a Bayesian learner who has a prior that is parameterized by an equivalent λ parameter. Martin and Nagel (2022, Proposition 6) argue that a prior that is incorrectly specified – i.e. with a value of λ that is excessively high – "should show up in the data as out-of-sample predictable returns" to an econometrician, since the representative agent under-weights predictive information. A familiarity bias in how investors make use of data can therefore impede market efficiency, and is a potential explanation for the existence of out-of-sample return predictability in the empirical asset pricing literature.[22] According to my empirical findings in Section 6 and in my next analysis, predictive signal availability does not necessarily resolve this inefficiency.

Sparsity & overconfidence. Since a biased learner may ignore valid predictive signals (Proposition 2), she effectively adopts a simplified and sparse model of the world. This lends itself to an alternative behavioral interpretation: Montiel Olea, Ortoleva, Pai, and Prat (2022) argue that economic agents who ignore valid predictive variables exhibit a form of overconfidence in how they specify their model of the world.

7.2 Empirically Measuring the Extent of Investor Bias

I now take the predictive model (25) to the data. By optimizing for the best fit at the individual investor-contest level, I estimate which predictive variables are used by each investor in each contest. The objectives of this analysis are threefold. First, it allows me to detect whether investors underweight predictive variables severely enough that they are ignored altogether. Second, it allows me to measure individuals' usage of *all* predictive variables, beyond the subset of macroeconomic predictive variables analyzed in Section 4. And third, counting the estimated number of predictive variables employed produces a simple measure of how strongly investors are biased; I can therefore test whether investors' familiarity bias lessens

[22] A variety of studies including Freyberger, Neuhierl, and Weber (2020) and Gu, Kelly, and Xiu (2020) find evidence of out-of-sample return predictability for US stocks. I find similar evidence for futures contracts.

with experience, as hypothesized earlier.

7.2.1 Estimation Procedure

My estimation procedure is built upon that of Friedman, Hastie, Höfling, and Tibshirani (2007), who provide a coordinate-wise descent algorithm that numerically solves the investor's prediction problem (25) without the need for an orthonormality assumption: given a fixed bias parameter value λ, their algorithm estimates the optimal value of $\boldsymbol{b}$ in Eqn. (25).

It remains to estimate the optimal value of λ. To achieve this, I define the following procedure. First, prepare a grid of possible values for λ, lower-bounded at zero, and estimate $\widehat{\boldsymbol{b}}$ for each of these possible values $\widehat{\lambda}$. Secondly, for each pair $(\widehat{\boldsymbol{b}}, \widehat{\lambda})$, predict expected returns for each of the historical returns in the investor's dataset as $\widehat{\boldsymbol{v}} = \boldsymbol{S}\widehat{\boldsymbol{b}}$. Compare these predicted values $\widehat{\boldsymbol{v}}$ to the actually observed historical returns $\boldsymbol{v}$ by calculating the element-wise mean-squared error (MSE) between these two vectors. Finally, pick the optimal pair $(\widehat{\boldsymbol{b}}^*, \widehat{\lambda}^*)$ to be the pair $(\widehat{\boldsymbol{b}}, \widehat{\lambda})$ that minimizes this MSE.

Operationalizing this procedure requires me to supply empirical values for $\boldsymbol{v}$ and $\boldsymbol{S}$. I therefore apply my estimation procedure to the timeseries of daily portfolio returns earned by the investor in place of $\boldsymbol{v}$,[23] and define a corresponding matrix of predictive signals $\boldsymbol{S}$ that the investor had access to at the time of forming her daily portfolio. For contests in which macroeconomic predictive variables were not available, the 880 columns of $\boldsymbol{S}$ are defined to contain the latest available 5 daily returns (i.e. lags 1 to 5) and their squares (proxying for a historical volatility signal), for each of the 88 futures contracts in the investor's tradeable universe. For contests in which macroeconomic predictive variables were available, the 934 columns of $\boldsymbol{S}$ include, in addition, the latest available values (i.e. lag 1) for each of the 54 macroeconomic predictive variables. Predictive variables are standardized to a common scale. The high dimensionality of the signals matrix $\boldsymbol{S}$ illustrates the challenges investors face when making predictions in a Big Data environment (Martin and Nagel, 2022) but does not pose any practical difficulties to the procedure of Friedman, Hastie, Höfling, and Tibshirani (2007).

I perform this estimation procedure for all contest entries of all investors who take part in more than one contest. To enable valid comparisons,[24] I report the estimated number of predictive variables that each investor entry uses; that is, the number of non-zero elements in

[23] In this regard, I follow a very large literature in empirical asset pricing that uses *ex post* returns to proxy for *ex ante* return expectations; for example, when testing factor models or other asset pricing models.

[24] As Lemma 2 in the Appendix makes clear, estimates for λ are only comparable if the dataset used is fixed throughout, which is not the case. The estimated number of predictive variables, on the other hand, is always comparable.

the corresponding vector of parameter solutions $\widehat{\boldsymbol{b}}^*$.

7.2.2 Results and Discussion

Figure 9 displays the mean number of estimated variables used by an investor according to experience level (i.e. number of contests participated in so far), split into the Treatment and Control groups that I defined in Section 6.1.

[Insert Figure 9 here]

These estimates of $\widehat{\boldsymbol{b}}^*$ indicate that investors make use of more predictive variables in the information environment in which more were available (Treatment) compared to the baseline environment (Control). However, investors of all experience levels ignore a subset of the predictive variables that are available to them. The ignored variables include both price-based and (in the Treatment group) macroeconomic predictive variables, suggesting that investor underweighting of predictive variables is a general phenomenon that extends beyond the set of macroeconomic indicators (Section 4).

Within each of the information environments, my estimates of $\widehat{\boldsymbol{b}}^*$ indicate that less experienced investors ignore more predictive variables than more experienced investors. This is consistent with the core of my hypothesized mechanism: that, as investors gain in experience, they learn to overcome their familiarity bias, and they therefore make better use of the predictive variables that are available to them. According to these results, one way in which investors make better use of predictive variables is by ignoring fewer of them: Figure 9 illustrates that each contest's worth of experience is associated with an additional 25-50 predictive variables being incorporated into investors' models of the world.

8 Conclusion

Across two different, yet related settings, I find that human experience is a crucial determinant of predictive signal adoption. Moreover, investors demonstrate a preference for familiar and recognizable signals, as well as an aversion to specifying complex predictive models. I document a positive experience gradient for performance and the appropriate adoption of valid predictive signals; this implies an important role for human skill, as well as a complementarity between the human decision-maker and the data-driven signal.

These findings have implications for our understanding of market efficiency and the persistence of return predictability. The behavioral frictions I document – familiarity bias and experience-dependent complexity aversion – act as psychological limits to arbitrage that slow

the incorporation of public information into prices. Even when signals are freely available and genuinely informative, investors may fail to exploit them due to behavioral barriers rather than rational constraints. This mechanism helps explain why the post-publication decline in signal informativeness documented by McLean and Pontiff (2016) is only partial: behavioral frictions prevent complete arbitrage even after information becomes public.

More broadly, this paper contributes to understanding how economic agents use predictive information and machine learning tools in practice. The value of predictive signals depends not only on their statistical properties but also on whether and how decision-makers incorporate them into their models. Familiarity bias, complexity aversion, and the gradual accumulation of experience shape this incorporation process, with important consequences for market efficiency and the distribution of investment returns. As predictive technologies become increasingly sophisticated and accessible, understanding these behavioral mechanisms becomes ever more critical for explaining observed patterns in financial markets and beyond.

References

Abis, S. (2020). Man vs. machine: Quantitative and discretionary equity management. *Available at SSRN*, *3717371*.

Abis, S., & Veldkamp, L. (2024). The changing economics of knowledge production. *Review of Financial Studies*, *37*(1), 89–118.

Abreu, M., Mendes, V., & Santos, J. A. (2011). Home country bias: Does domestic experience help investors enter foreign markets? *Journal of Banking & Finance*, *35*(9), 2330–2340.

Barberis, N., & Thaler, R. (2003). A survey of behavioral finance. In *Handbook of the Economics of Finance* (pp. 1053–1128, Vol. 1B). Elsevier.

Ben-David, I., Li, J., Rossi, A., & Song, Y. (2022). What do mutual fund investors really care about? *Review of Financial Studies*, *35*(4), 1723–1774.

Bennett, J. A., Sias, R. W., & Starks, L. T. (2003). Greener pastures and the impact of dynamic institutional preferences. *Review of Financial Studies*, *16*(4), 1203–1238.

Best, M. J., & Grauer, R. R. (1991). On the sensitivity of mean-variance-efficient portfolios to changes in asset means: Some analytical and computational results. *Review of Financial Studies*, *4*(2), 315–342.

Blankespoor, E., deHaan, E., & Marinovic, I. (2020). Disclosure processing costs, investors' information choice, and equity market outcomes: A review. *Journal of Accounting and Economics*, *70*(2-3), 101344.

Bonelli, M., & Foucault, T. (2023). Displaced by big data: Evidence from active fund managers. *Available at SSRN*, *4527672*.

Breiman, L. (2001). Random forests. *Machine Learning*, *45*(1), 5–32.

Brynjolfsson, E., Li, D., & Raymond, L. (2025). Generative ai at work. *Quarterly Journal of Economics*, *140*(2), 889–942.

Cao, H. H., Han, B., Hirshleifer, D., & Zhang, H. H. (2011). Fear of the unknown: Familiarity and economic decisions. *Review of Finance*, *15*(1), 173–206.

Chan, L. K., Chen, H.-L., & Lakonishok, J. (2002). On mutual fund investment styles. *Review of Financial Studies*, *15*(5), 1407–1437.

Chen, A. Y., & Zimmermann, T. (2022). Open source cross-sectional asset pricing. *Critical Finance Review*, *11*(02), 207–264.

Chen, S., Sialm, C., & Xu, D. X. (2025). The growth and performance of artificial intelligence in asset management. *Available at SSRN*, *5638612*.

Cochrane, J. H. (2011). Presidential address: Discount rates. *Journal of Finance*, *66*(4), 1047–1108.

Cohen, L., & Frazzini, A. (2008). Economic links and predictable returns. *Journal of Finance*, *63*(4), 1977–2011.

Cookson, J. A., & Niessner, M. (2020). Why don't we agree? Evidence from a social network of investors. *Journal of Finance*, *75*(1), 173–228.

Cornuéjols, G., Peña, J. F., & Tütüncü, R. (2018). *Optimization Methods in Finance* (2nd ed.). Cambridge University Press.

Del Guercio, D. (1996). The distorting effect of the prudent-man laws on institutional equity investments. *Journal of Financial Economics*, *40*(1), 31–62.

DellaVigna, S., & Pollet, J. M. (2009). Investor inattention and friday earnings announcements. *Journal of Finance*, *64*(2), 709–749.

Dugast, J., & Foucault, T. (2021). Equilibrium data mining and data abundance. *HEC Paris Research Paper No. FIN-2020-1393, Université Paris-Dauphine Research Paper*, *3710495*.

Dyer, T., Guest, N., & Yu, E. (2025). New accounting standards and the performance of quantitative investors. *Journal of Accounting and Economics*, *79*(2-3), 101740.

Edelen, R. M., Ince, O. S., & Kadlec, G. B. (2016). Institutional investors and stock return anomalies. *Journal of Financial Economics*, *119*(3), 472–488.

Epstein, L. G., & Schneider, M. (2008). Ambiguity, information quality, and asset pricing. *Journal of Finance*, *63*(1), 197–228.

Falkenstein, E. G. (1996). Preferences for stock characteristics as revealed by mutual fund portfolio holdings. *Journal of Finance*, *51*(1), 111–135.

Fama, E. F. (1998). Market efficiency, long-term returns, and behavioral finance. *Journal of financial economics*, *49*(3), 283–306.

Fleiss, S., Kominers, S. D., & Ughetta, C. B. (2017). Quantopian: A new model for active management. *HBS Case Study 9-218-046*.

Freyberger, J., Neuhierl, A., & Weber, M. (2020). Dissecting characteristics nonparametrically. *Review of Financial Studies*, *33*(5), 2326–2377.

Friedman, J., Hastie, T., Höfling, H., & Tibshirani, R. (2007). Pathwise coordinate optimization. *Annals of Applied Statistics*, *1*(2), 302–332.

Gabaix, X. (2014). A sparsity-based model of bounded rationality. *Quarterly Journal of Economics*, *129*(4), 1661–1710.

Garlappi, L., Uppal, R., & Wang, T. (2007). Portfolio selection with parameter and model uncertainty: A multi-prior approach. *Review of Financial Studies*, *20*(1), 41–81.

Greene, W. H. (1981). Sample selection bias as a specification error: A comment. *Econometrica*, 795–798.

Grinblatt, M., Titman, S., & Wermers, R. (1995). Momentum investment strategies, portfolio performance, and herding: A study of mutual fund behavior. *American Economic Review*, *85*(5), 1088–1105.

Gu, S., Kelly, B., & Xiu, D. (2020). Empirical asset pricing via machine learning. *Review of Financial Studies*, *33*(5), 2223–2273.

Hanna, R., Mullainathan, S., & Schwartzstein, J. (2014). Learning through noticing: Theory and evidence from a field experiment. *Quarterly Journal of Economics*, *129*(3), 1311–1353.

Harvey, C. R., Liu, Y., & Zhu, H. (2016). . . . and the cross-section of expected returns. *Review of Financial Studies*, *29*(1), 5–68.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement* (pp. 475–492, Vol. 5). NBER.

Hou, K., Xue, C., & Zhang, L. (2020). Replicating anomalies. *Review of Financial Studies*, *33*(5), 2019–2133.

Huberman, G. (2001). Familiarity breeds investment. *Review of Financial Studies*, *14*(3), 659–680.

Illeditsch, P. K. (2011). Ambiguous information, portfolio inertia, and excess volatility. *Journal of Finance*, *66*(6), 2213–2247.

Jensen, T. I., Kelly, B., & Pedersen, L. H. (2023). Is there a replication crisis in finance? *Journal of Finance*, *78*(5), 2465–2518.

Kacperczyk, M., & Seru, A. (2007). Fund manager use of public information: New evidence on managerial skills. *Journal of Finance*, *62*(2), 485–528.

Kacperczyk, M., Van Nieuwerburgh, S., & Veldkamp, L. (2016). A rational theory of mutual funds' attention allocation. *Econometrica*, *84*(2), 571–626.

Kleinberg, J., Ludwig, J., Mullainathan, S., & Obermeyer, Z. (2015). Prediction policy problems. *American Economic Review*, *105*(5), 491–495.

Lee, J. D., Sun, D. L., Sun, Y., & Taylor, J. E. (2016). Exact post-selection inference, with application to the lasso. *Annals of Statistics*, *44*(3), 907–927.

Lettau, M., Ludvigson, S. C., & Manoel, P. (2018). Characteristics of mutual fund portfolios: Where are the value funds? *NBER Working Paper*, *25381*.

Lewellen, J. (2011). Institutional investors and the limits of arbitrage. *Journal of Financial Economics*, *102*(1), 62–80.

Linnainmaa, J. T. (2011). Why do (some) households trade so much? *Review of Financial Studies*, *24*(5), 1630–1666.

Linnainmaa, J. T., & Roberts, M. R. (2018). The history of the cross-section of stock returns. *The Review of Financial Studies*, *31*(7), 2606–2649.

List, J. A. (2003). Does market experience eliminate market anomalies? *Quarterly Journal of Economics*, *118*(1), 41–71.

Lundberg, S. M., & Lee, S.-I. (2017). A unified approach to interpreting model predictions. *Advances in Neural Information Processing Systems*, *30*.

Markowitz, H. (1952). Portfolio selection. *Journal of Finance*, *7*(1), 77–91.

Martin, I. W., & Nagel, S. (2022). Market efficiency in the age of big data. *Journal of Financial Economics*, *145*(1), 154–177.

Matějka, F., & McKay, A. (2015). Rational inattention to discrete choices: A new foundation for the multinomial logit model. *American Economic Review*, *105*(1), 272–298.

McLean, D. R., & Pontiff, J. (2016). Does academic research destroy stock return predictability? *Journal of Finance*, *71*(1), 5–32.

McLean, D. R., Pontiff, J., & Reilly, C. (2025). Taking sides on return predictability. *Journal of Financial Economics*, *173*, 104158.

Merton, R. C. (1980). On estimating the expected return on the market: An exploratory investigation. *Journal of Financial Economics*, *8*(4), 323–361.

Montiel Olea, J. L., Ortoleva, P., Pai, M. M., & Prat, A. (2022). Competing models. *Quarterly Journal of Economics*, *137*(4), 2419–2457.

Nicolosi, G., Peng, L., & Zhu, N. (2009). Do individual investors learn from their trading experience? *Journal of Financial Markets*, *12*(2), 317–336.

Oprea, R. (2020). What makes a rule complex? *American Economic Review*, *110*(12), 3913–3951.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.

Schwarz, C. G., & Potter, M. E. (2016). Revisiting mutual fund portfolio disclosure. *Review of Financial Studies*, *29*(12), 3519–3544.

Seasholes, M. S., & Zhu, N. (2010). Individual investors and local bias. *Journal of Finance*, *65*(5), 1987–2010.

Seru, A., Shumway, T., & Stoffman, N. (2010). Learning by trading. *Review of Financial Studies*, *23*(2), 705–739.

Sheng, J., Sun, Z., Yang, B., & Zhang, A. L. (forthcoming). Generative ai and asset management. *Review of Financial Studies*.

Tian, X., Loftus, J. R., & Taylor, J. E. (2018). Selective inference with unknown variance via the square-root lasso. *Biometrika*, *105*(4), 755–768.

Tibshirani, R. (1996). Regression shrinkage and selection via the lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, *58*(1), 267–288.

Xu, H., Caramanis, C., & Mannor, S. (2010). Robust regression and Lasso. *IEEE Transactions on Information Theory*, *56*(7), 3561–3574.

Zheng, Y. (2017). Quants in Utopia? Quantopian and its crowd-wisdom hedge-fund model. *HKU Case Study HK1109*.

Appendix

Proofs

To show Proposition 1 requires two lemmas. The first is by Xu, Caramanis, and Mannor (2010), and the second also relies on results by Tian, Loftus, and Taylor (2018).

Lemma 1 (Xu, Caramanis, and Mannor (2010) Theorem 1). *The optimization problem*

$$\min_{\boldsymbol{b}\in\mathbb{R}^m} \max_{\boldsymbol{U}\in\mathcal{U}} ||\boldsymbol{v} - (\boldsymbol{S}+\boldsymbol{U})\boldsymbol{b}||_2 \tag{28}$$

with uncertainty set $\mathcal{U}$ defined by

$$\mathcal{U} := \left\{[\boldsymbol{u}_1\ \boldsymbol{u}_2\ \ldots\ \boldsymbol{u}_m] : ||\boldsymbol{u}_i||_2 \leq \delta_i\ \forall\ i = 1,\ldots,m\right\} \tag{29}$$

is equivalent to the optimization problem

$$\min_{\boldsymbol{b}\in\mathbb{R}^m} ||\boldsymbol{v} - \boldsymbol{S}\boldsymbol{b}||_2 + \sum_{i=1}^{m} \delta_i |b_i|. \tag{30}$$

Proof. See Xu, Caramanis, and Mannor (2010) Theorem 1. □

Lemma 2. *The optimization problem*

$$\min_{\boldsymbol{b}\in\mathbb{R}^m} ||\boldsymbol{v} - \boldsymbol{S}\boldsymbol{b}||_2 + \delta||\boldsymbol{b}||_1 \tag{31}$$

is equivalent to the optimization problem

$$\min_{\boldsymbol{b}\in\mathbb{R}^m} \frac{1}{2}||\boldsymbol{v} - \boldsymbol{S}\boldsymbol{b}||_2^2 + \lambda||\boldsymbol{b}||_1 \tag{32}$$

with the one-to-one change of parameters

$$\lambda = \frac{\delta a_1}{\sqrt{1-\delta^2 a_2}} \quad \Leftrightarrow \quad \delta = \frac{\lambda}{\sqrt{\lambda^2 a_3 + a_4}}, \tag{33}$$

in which each of λ, δ are monotonically increasing in the other, and where $a_1, a_2, a_3, a_4 > 0$ and are constant for any fixed dataset $(\boldsymbol{S}, \boldsymbol{v})$ and common parameter estimates $\boldsymbol{b}^*$.

Proof. Xu, Caramanis, and Mannor (2010, Appendix A) show that the two problems are equivalent up to a change of λ, δ parameters. Tian, Loftus, and Taylor (2018, Lemma 2) derive the

mapping between δ and λ in terms of $\boldsymbol{b}^*, \boldsymbol{S}, \boldsymbol{v}$. It is simple to show that $a_1, a_2, a_3, a_4 > 0$ in their formulation and that $\frac{\partial \lambda}{\partial \delta}, \frac{\partial \delta}{\partial \lambda} > 0$ for $\lambda, \delta > 0$. □

The proof of Proposition 1 now follows.

Proposition 1. *Under Assumption 1, a biased investor solves her original problem* (23) *by solving an equivalent formulation*

$$\min_{\boldsymbol{b} \in \mathbb{R}^m} \frac{1}{2} ||\boldsymbol{v} - \boldsymbol{S}\boldsymbol{b}||_2^2 + \lambda ||\boldsymbol{b}||_1, \tag{25}$$

where $\lambda \geq 0$ *is a scaling of* δ *in* (24).

Proof. Applying Lemma 1 to the optimization problem in Definition 2 with $\delta_i = \delta \ \forall i$ shows that (23) is equivalent to the following optimization problem:

$$\min_{\boldsymbol{b} \in \mathbb{R}^m} ||\boldsymbol{v} - \boldsymbol{S}\boldsymbol{b}||_2 + \delta ||\boldsymbol{b}||_1. \tag{34}$$

By Lemma 2, (34) is equivalent to (25), with the mapping between λ, δ given by the lemma. The optimization problems (23) and (25) are thus equivalent. □

The result in Proposition 2 is due to Tibshirani (1996), who did not provide a proof. For completeness, a brief proof follows.

Proposition 2. *Under Assumption 2, the biased investor will solve her prediction problem* (25) *by predicting*

$$\widehat{\mu} = \boldsymbol{s}\widehat{\boldsymbol{b}}, \tag{26}$$

where $\boldsymbol{s}$ *is an m-dimensional row vector consisting of the m predictive signals to the currently traded asset's expected return* μ*, and* $\widehat{\boldsymbol{b}}$ *is an m-dimensional column vector of weights whose elements are defined by*

$$\widehat{b}_k = \text{sign}(\boldsymbol{s}_k^T \boldsymbol{v}) \max\left\{|\boldsymbol{s}_k^T \boldsymbol{v}| - \lambda, 0\right\}. \tag{27}$$

Proof. I start from the formulation (25). As the ℓ_1 norm is not differentiable, I follow Lee, Sun, Sun, and Taylor (2016) in writing the (necessary and sufficient) Karush–Kuhn–Tucker conditions for a solution in terms of the subdifferential $\partial(||\boldsymbol{b}||_1)$:

$$\boldsymbol{0} \ \in \ \boldsymbol{S}^T(\boldsymbol{S}\widehat{\boldsymbol{b}} - \boldsymbol{v}) + \lambda \partial(||\boldsymbol{b}||_1) \tag{35}$$

The orthonormality assumption $\boldsymbol{S}^T\boldsymbol{S} = \boldsymbol{I}$ then simplifies it to

$$\boldsymbol{0} \ \in \ \widehat{\boldsymbol{b}} - \boldsymbol{S}^T\boldsymbol{v} + \lambda \partial(||\boldsymbol{b}||_1). \tag{36}$$

Note that each element $\widehat{b}_k$ of $\widehat{\boldsymbol{b}}$ does not depend on other elements of $\widehat{\boldsymbol{b}}$:

$$0 \;=\; \widehat{b}_k - \boldsymbol{s}_k^T \boldsymbol{v} + \lambda \widehat{z}_k, \tag{37}$$

where $\widehat{z}_k = \text{sign}(\widehat{b}_k)$ when $\widehat{b}_k \neq 0$ and the set $[-1, +1]$ otherwise. The result (27) then follows.

□

Figures

Figure 1: Stock Return Predictors' Ex-Post Mean Returns. This figure plots the distribution, across the 212 Chen and Zimmermann (2022) predictors, of each predictor's *ex post* mean return based on long–short (top-minus-bottom decile) portfolio sorts. Positive values indicate predictors whose post-publication sorts earn positive average returns. The dashed gray line indicates the sample mean.

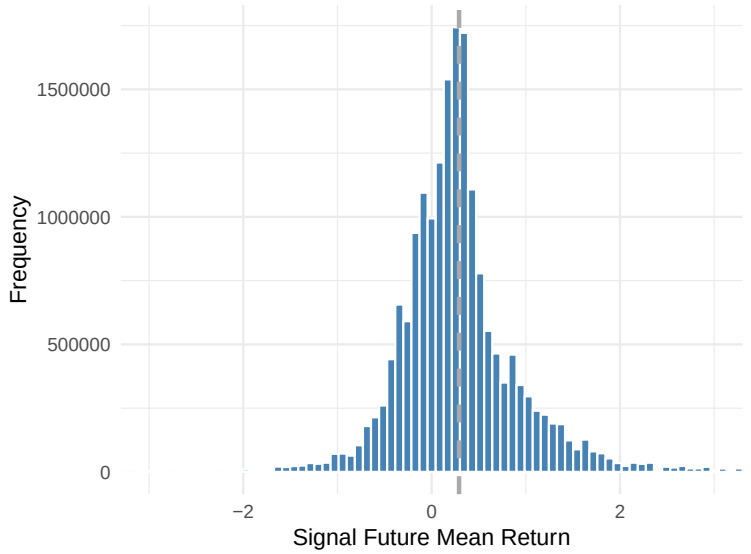


Figure 2: Portfolio–Signal Tilts. This figure plots the distribution of quarterly portfolio–signal tilts across mutual funds and predictors. Tilts are defined by Eqn. (3); positive tilts indicate active loadings toward stocks with higher predicted returns. The dashed gray line indicates the sample mean.

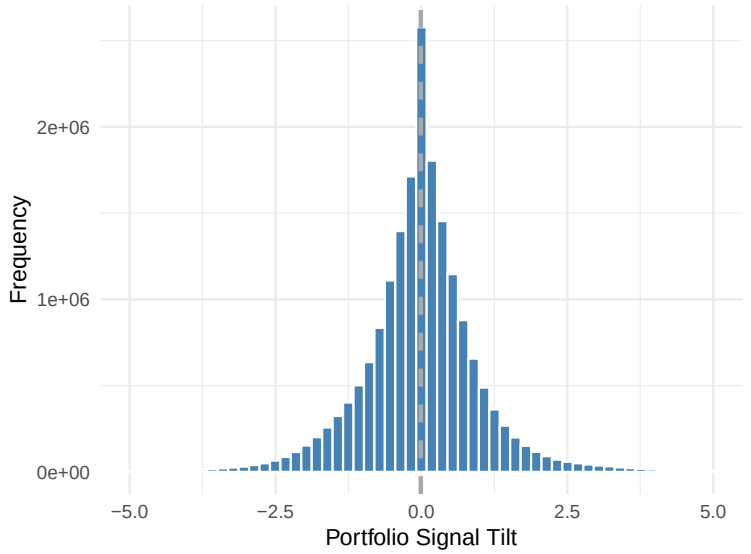


Figure 3: Trading Contest Backtest vs. Live periods. Stylized illustration of how the Backtest period expands with each new contest, whereas the (disjoint) Live period rolls forward from one contest to the next.

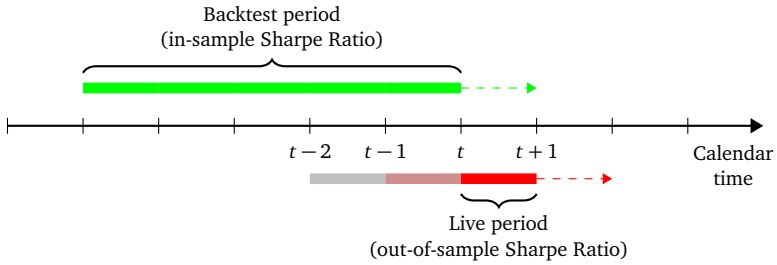


Figure 4: Distinction Between Macro Signal Release Days and Macro Signal Availability. Contestants take part in one or more trading contests t by entering trading strategies that run over an extended calendar period. The historical periods of all contests overlap in calendar time, so each historical calendar day τ can be matched to multiple trading strategies entered into various contests t: this allows entries to be compared for the same days τ. The values of macro indicators are updated by the Quantiacs trading platform on a subset of these calendar days (highlighted in solid colors) but only accessible to trading strategies entered into later contests (highlighted in red).

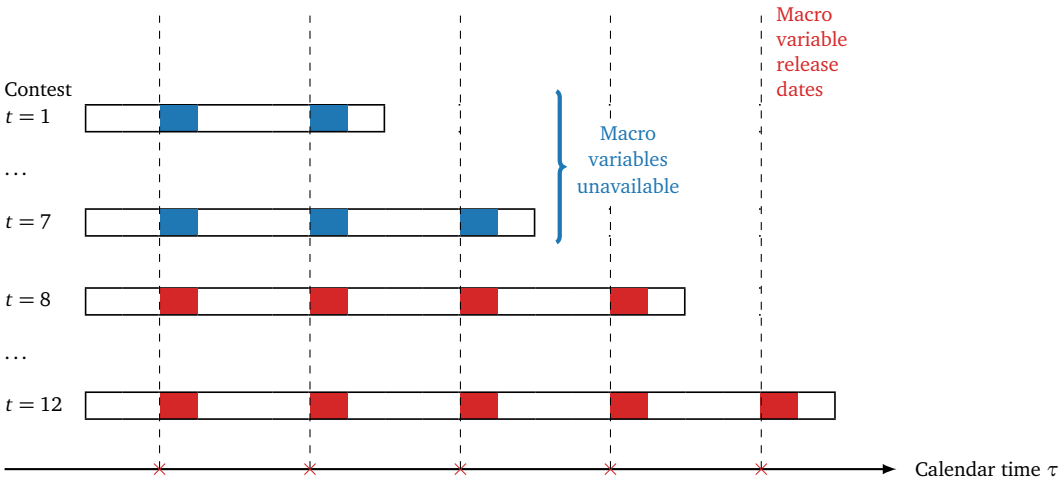


Figure 5: Variation in the Number of Signals Released on Macro Release Dates. This chart counts the number of macro variables that were released to trading strategies on the Quantiacs platform, including both in-sample/Backtest and out-of-sample/Live periods. Variables were released in monthly batches, but not every variable was made available or updated by the platform. The historical calendar dates on which one or more macro variables were updated are termed "macro release days."

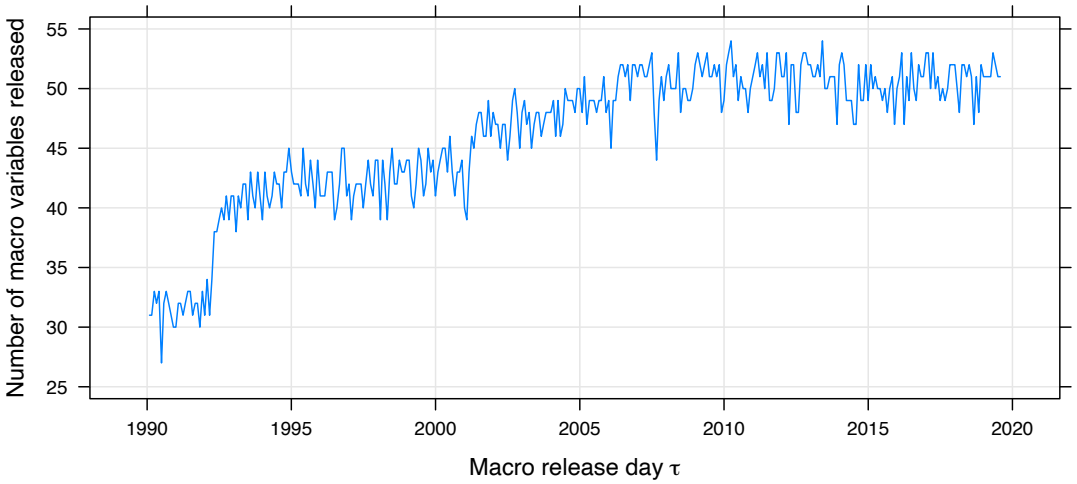


Figure 6: Cross-Sectional Mean Correlation in Trading. For each trading contest index t, this chart displays the cross-sectional mean of the pairwise correlations $\rho_{i,j,t}$ in daily turnovers for all best live entries in the contest. Self-correlations are excluded.

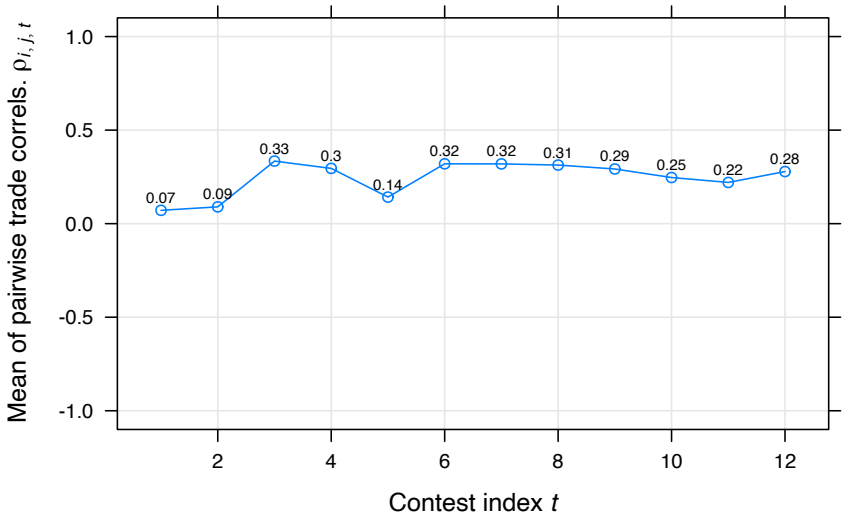


Figure 7: Out-of-Sample Cumulative Predicted Returns for Futures Contracts Using the Macro Signals. This chart plots the cumulative return of a trading strategy based on the benchmark Random Forest cross-sectional return prediction model. The machine learning model uses only the macroeconomic predictive variables available on the Quantiacs platform as its inputs, to predict returns for the full universe of futures contracts.

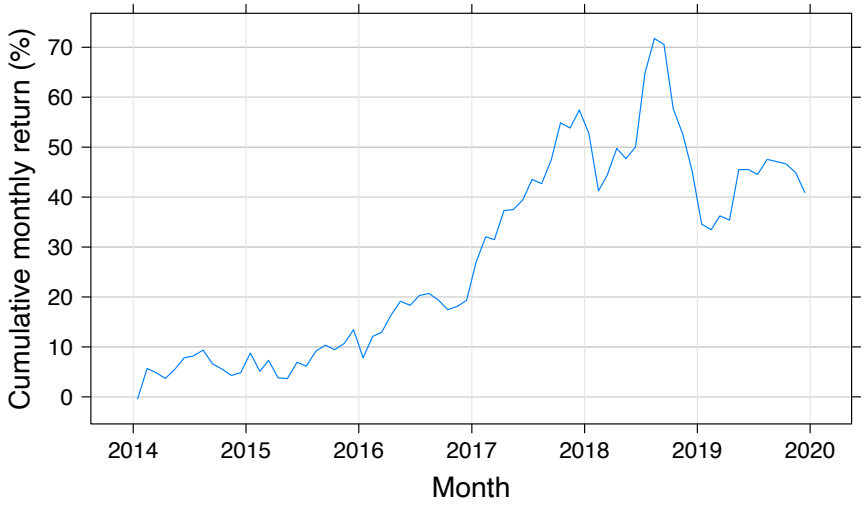


Figure 8: Soft-Thresholding. Illustration of soft-thresholding an input. The dashed line denotes the benchmark $f(x) = x$. The solid blue line $f(x) = \text{sign}(x)\max\{|x| - \lambda, 0\})$ is the result of soft-thresholding the input x at threshold λ.

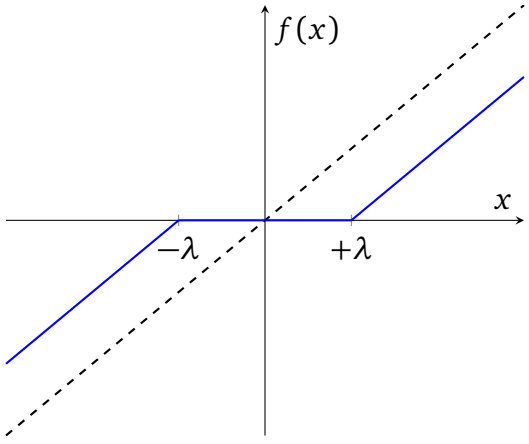
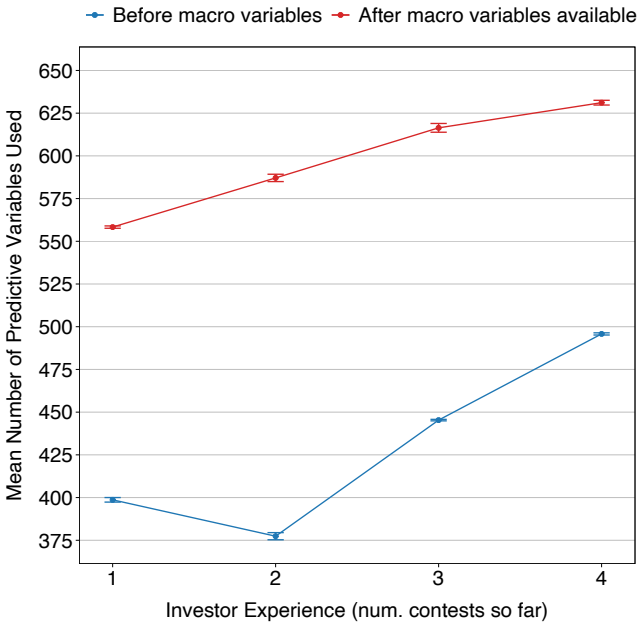


Figure 9: Trading Contest Participants' Usage of Predictive Variables. This chart shows the number of predictive variables used by investors. These are measured according to the model of investor behavior in Section 7.1.2, based on investors' daily portfolio returns and a set of predictive variables. A total of 934 variables are tested, including both return-based and macroeconomic predictive variables. The positive slopes indicate that both groups of investors make use of more predictive variables as they gain in experience. The offset indicates investors with access to more predictive variables use more of them. The split into groups of investors with and without access to macro predictive variables is the same as for the main regression specifications. Bars represent standard errors.



Tables

Table 1: Mutual Fund Portfolio Tilts, Signal Familiarity, and Signal Informativeness. This table shows the results of regressing portfolio–signal tilts (Eqn. (3)) on the time elapsed since the signal was publicy disseminated in the form of an academic publication, as well as on the out-of-sample informativeness of the signal (Figure 1). Additional fund and portfolio controls are included, together with a variety of fixed-effects specifications. Standard errors are double-clustered at the portfolio–quarter level.

Dependent Variable:	Portfolio Signal Tilt$_{p,c,t}$		
	(1)	(2)	(3)
Log Months Since Signal Publication$_{c,t}$	0.0340***	0.0394***	0.0339***
	(0.0053)	(0.0056)	(0.0053)
Signal Predictive Power$_{c,\,t+1\,\text{onwards}}$	-0.1136***	-0.1189***	-0.1133***
	(0.0107)	(0.0114)	(0.0107)
Log Fund Size$_{p,t}$		-0.0156***	
		(0.0025)	
Log Fund Age$_{p,t}$		0.0079	
		(0.0053)	
Fund Expense Ratio$_{p,t}$		-0.0430	
		(1.079)	
Portfolio Turnover$_{p,t}$		0.0074**	
		(0.0029)	
Portfolio FEs	✓	✓	
Quarter FEs	✓	✓	
Portfolio × Quarter FEs			✓
N	16,881,852	12,509,289	16,881,852
R^2	0.03522	0.03532	0.04727

*Note: standard errors (in parentheses) are clustered by portfolio and quarter. *p<0.1; **p<0.05; ***p<0.01*

Table 2: Determinants of Mutual Fund Portfolio Tilts by Portfolio and Fund Type. This table repeats the baseline regression (Table 1 column (3)) separately by portfolio type (long-only vs. long/short) and by whether the fund describes itself as following a quant strategy (Dyer, Guest, and Yu, 2025). Standard errors are double-clustered at the portfolio–quarter level.

Dependent Variable:	Portfolio Signal Tilt$_{p,c,t}$					
	(1)	(2)	(3)	(4)	(5)	(6)
Log Months Since Signal Publication$_{c,t}$	0.0339***	0.0343***	0.0199***	0.0273***	0.0286***	-0.0210*
	(0.0053)	(0.0054)	(0.0065)	(0.0061)	(0.0063)	(0.0116)
Signal Predictive Power$_{c,\,t+1\,\text{onwards}}$	-0.1133***	-0.1171***	0.0123	-0.1006***	-0.1060***	0.1182***
	(0.0107)	(0.0109)	(0.0196)	(0.0182)	(0.0187)	(0.0343)
Sub-sample:						
Strategy	All	All	All	Quant	Quant	Quant
Portfolio	All	Long-Only	Long/Short	All	Long-Only	Long/Short
Portfolio × Quarter FEs	✓	✓	✓	✓	✓	✓
N	16,881,852	16,435,505	384,567	1,165,027	1,135,770	27,511
R^2	0.04727	0.04707	0.07897	0.04408	0.04372	0.18140

*Note: standard errors (in parentheses) are clustered by portfolio and quarter. *p<0.1; **p<0.05; ***p<0.01*

Table 3: Experience Effects Within-Fund Manager. This table shows the results of regressing portfolio–signal tilts on signal predictive power and its interaction with an indicator for whether the fund manager has above-median total experience in the mutual fund industry. The sample consists of single-manager funds only. Control (not shown) are for (log) fund size, (log) fund age, expense ratio, and portfolio turnover. Standard errors are double-clustered at the portfolio–quarter level.

Dependent Variable:	Portfolio Signal Tilt$_{p,c,t}$			
	(1)	(2)	(3)	(4)
Experienced Manager$_{p,t}$	0.0221**	0.0097	-0.0111	-0.0324***
	(0.0085)	(0.0089)	(0.0087)	(0.0093)
Signal Predictive Power$_{c,\,t+1\,\text{onwards}}$	-0.1480***	-0.1733***	-0.1735***	-0.1755***
	(0.0149)	(0.0164)	(0.0164)	(0.0164)
Experienced Manager$_{p,t}$		0.0434**	0.0432**	0.0418**
× Signal Predictive Power$_{c,\,t+1\,\text{onwards}}$		(0.0167)	(0.0167)	(0.0167)
Fund & Portfolio Controls	✓	✓	✓	✓
Manager FEs	✓	✓		
Manager × Portfolio FEs			✓	✓
Quarter FEs				✓
N	3,606,834	3,606,834	3,606,834	3,606,834
R^2	0.03356	0.03372	0.03643	0.03741

*Note: std. errs. (in parentheses) are clustered by portfolio and quarter. *p<0.1; **p<0.05; ***p<0.01*

Table 4: Trading contests. Summaries of each of the 12 contests in the leaderboard sample. "OOS" denotes out-of-sample, "IQR" denotes the inter-quartile range, and "SD" denotes the cross-sectional standard deviation. The Entries columns summarize entries per contestant.

					Backtest/In-Sample Sharpe Ratios					Live/OOS Sharpe Ratios								
			Entries		Percentile			Dispersion		Percentile					Dispersion		Variables	
Index	Live/OOS Period	Contestants	Median	Mean	25	50	75	IQR	SD	10	25	50	75	90	IQR	SD	Price	Macro
1	2014-12-01 – 2015-01-31	13	1.00	1.31	0.72	0.94	1.03	0.31	0.49	-3.85	-2.33	0.00	0.44	1.95	2.77	2.99	✓	
2	2015-07-01 – 2015-09-30	16	2.00	3.62	0.38	0.59	1.38	1.00	0.83	-2.41	-2.07	-1.18	0.74	3.03	2.81	2.12	✓	
3	2015-10-01 – 2015-12-31	23	1.00	1.17	0.09	0.39	0.46	0.37	1.24	-1.97	-0.63	0.26	0.98	1.63	1.61	1.55	✓	
4	2016-01-01 – 2016-03-31	30	1.00	3.53	0.15	0.86	2.38	2.22	2.22	-1.66	-1.20	0.09	1.70	3.11	2.90	2.18	✓	
5	2016-04-01 – 2016-06-30	38	2.50	4.05	0.43	0.76	1.92	1.49	1.41	-2.72	-1.49	-0.21	1.16	2.50	2.65	2.46	✓	
6	2016-08-01 – 2016-10-31	87	1.00	3.02	0.33	0.64	2.24	1.91	1.62	-2.95	-2.13	-0.75	-0.22	0.75	1.91	1.57	✓	
7	2017-01-01 – 2017-03-31	125	1.00	2.54	0.25	0.39	1.53	1.28	1.79	-2.45	-1.38	-1.03	0.67	1.44	2.05	1.69	✓	
8	2017-04-15 – 2017-07-31	92	1.00	3.50	0.38	1.19	2.36	1.98	2.63	-1.46	-0.48	0.68	1.41	1.93	1.89	1.67	✓	✓
9	2017-10-01 – 2018-01-31	163	1.00	2.31	0.30	0.38	1.90	1.60	5.57	-2.07	-0.35	1.00	2.07	2.90	2.42	2.04	✓	✓
10	2018-02-01 – 2018-05-31	63	1.00	3.73	0.40	1.61	4.92	4.52	4.19	-2.67	-1.50	-0.24	0.23	1.56	1.73	2.09	✓	✓
11	2018-07-01 – 2018-10-31	95	1.00	2.88	0.38	0.69	2.50	2.12	5.12	-2.16	-1.65	-0.89	-0.58	0.42	1.06	1.46	✓	✓
12	2019-01-01 – 2019-04-30	129	1.00	3.13	0.36	0.55	2.27	1.91	3.27	-2.90	-2.56	-1.21	1.65	3.43	4.21	2.61	✓	✓
	Overall:	874	1.00	2.92														

Table 5: Relating differences in trading activity to differential interpretations of information. OLS & panel regressions of pairwise dissimilarity in trading activity against pairwise differences in the response to identical predictive information releases. The outcome variable is the negative of the pairwise correlation of the daily trading volume of two machine-driven trading strategies (i, j). Trading strategies i, j are matched by trading contest index t to ensure they have access to identical information sets. The covariates each measure the distance between daily trading volume responses to the predictive information content of macro release days $(\widehat{\beta}_i, \widehat{\beta}_j)$, as measured using individual regressions for i, j.

Dependent Variable:	$-\rho_{i,j}$					
	(1)	(2)	(3)	(4)	(5)	(6)
$\log\big(\lvert\widehat{\beta}_i - \widehat{\beta}_j\rvert\big)$	0.0486***	0.0508***				
	(0.0030)	(0.0026)				
$\log\big(1 + \lvert\widehat{\beta}_i - \widehat{\beta}_j\rvert\big)$			0.2423***	0.2866***		
			(0.0225)	(0.0362)		
$\text{arcsinh}\big(\lvert\widehat{\beta}_i - \widehat{\beta}_j\rvert\big)$					0.1952***	0.2307***
					(0.0170)	(0.0258)
Intercept	✓		✓		✓	
Contest FEs		✓		✓		✓
Observations	19,640	19,640	21,512	21,512	21,512	21,512
R^2	0.343	0.380	0.338	0.420	0.345	0.428
Within R^2		0.371		0.416		0.424

Note: standard errors (in parentheses) are clustered by contest and contestant.
*p<0.1; **p<0.05; ***p<0.01

Table 6: Trading strategy responses to released macro signal informativeness. Standardized daily turnover of best live trading strategies on macro release days, regressed against a measure of release days' predictive information content, interacted with an indicator for investor experience.

Dependent Variable:	$Volume_{i,\tau}$		
	(1)	(2)	(3)
$Experienced_{i,t}$	0.0673**		-0.3229*
	(0.0184)		(0.1208)
$Informativeness_{\tau}$		0.6774**	0.6348*
		(0.2132)	(0.2362)
$Experienced_{i,t} \times Informativeness_{\tau}$			1.025**
			(0.2357)
Intercept	✓		
Model Year FEs		✓	✓
Observations	142,332	142,332	142,332
R^2	0.00014	0.00034	0.00054

Note: standard errors (in parentheses) are clustered by contest and contestant. *p<0.1; **p<0.05; ***p<0.01

Table 7: Trading strategy responses to released macro signal familiarity. Standardized daily turnover of best live trading strategies on macro release days, regressed against two indices of macro release day familiarity, interacted with an indicator for investor experience. Familiarity measures are based on news article mentions in columns (2) and (3), and book mentions in columns (4) and (5).

Dependent Variable:	$Volume_{i,\tau}$				
Familiarity Index:		News Articles		Books	
	(1)	(2)	(3)	(4)	(5)
$Experienced_{i,t}$	0.0673**		-0.3381		-0.1500
	(0.0184)		(0.3085)		(0.2425)
$Familiarity_{\tau}$		0.0027***	0.0026***	0.0026***	0.0026***
		(0.0003)	(0.0003)	(0.0005)	(0.0005)
$Experienced_{i,t} \times Familiarity_{\tau}$			0.0043		0.0023
			(0.0031)		(0.0024)
Intercept	✓				
Benchmark Year FEs		✓	✓	✓	✓
Observations	142,332	142,332	142,332	142,332	142,332
R^2	0.00014	0.00019	0.00034	0.00014	0.00029

Note: standard errors (in parentheses) are clustered by contest and contestant. *p<0.1; **p<0.05; ***p<0.01

Table 8: Performance outcomes and experience, in- and out-of-sample. OLS & panel regressions of in-sample (Backtest) & out-of-sample (Live) performance outcomes against experience.

	Dependent variable:			
	Backtest $SR_{i,t}^{Best}$		Live $SR_{i,t}^{Best}$	
	(1)	(2)	(3)	(4)
Contests experienced$_{i,t}$	1.161***	1.338***	0.445**	1.261***
	(0.055)	(0.505)	(0.178)	(0.456)
Intercept	✓		✓	
Contest FEs		✓		✓
Contestant FEs		✓		✓
Observations	874	874	874	874
R^2	0.156	0.024	0.035	0.040

Note: standard errors (in parentheses) are clustered by contest and contestant. *p<0.1; **p<0.05; ***p<0.01

Table 9: Model agreement and experience. OLS & panel regressions for pairwise correlations of the daily turnover series of best live trading strategies indexed by i, j that have been matched by experience and by trading contest index t.

Dependent Variable:	$\rho_{i,j,t}$		
	(1)	(2)	(3)
Contests experienced$_{i,j,t}$	0.0625***	0.0640***	0.0929*
	(0.0176)	(0.0182)	(0.0436)
(Intercept)	0.3087***		
	(0.0207)		
Contest FEs		✓	✓
Contestant i FEs			✓
Observations	25,829	25,829	25,829
R^2	0.00112	0.00939	0.30492

Note: standard errors (in parentheses) are clustered by contest and contestant. *p<0.1; **p<0.05; ***p<0.01

Table 10: Investor balance across information environments. Balance of investor populations across the Control & Treatment groups, as defined in Section 6.1: investors in the Control group have no access to macroeconomic predictive variables, and investors in the Treatment group do. Panel A shows participation attributes of individual contestants i at contests t. Panel B shows participation attributes aggregated at the contest level (t). The comparisons are of observable attributes before vs. after the additional macroeconomic indicators were made available to contestants by the trading platform.

Panel A: Individual-level balance

	Control (N=289)		**Treatment** (N=502)			
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Diff. in Means**	**p value**
Contests experienced$_{i,t}$	1.0796	0.3183	1.1058	0.3985	0.0262	0.3108
Percentile(Score$^{Best}_{i,t-1}$)	0.6445	0.2570	0.7102	0.2588	0.0657	0.3592

Panel B: Contest-level balance

	Control (N=6)		**Treatment** (N=4)			
	Mean	**Std. Dev.**	**Mean**	**Std. Dev.**	**Diff. in Means**	**p value**
Mean$_t$(Contests experienced$_{i,t}$)	1.2092	0.0544	1.2191	0.0810	0.0099	0.8389
Fraction of first-time contestants at t	0.8415	0.0534	0.8580	0.0396	0.0165	0.5914
Fraction of last-time contestants at t	0.7813	0.0778	0.8542	0.0567	0.0729	0.1263

Note: the reported p-values are from two-sided t-tests. For mechanical reasons, in Panel B the first contest is excluded (because its fraction of first-time contestants would be 1.0) and the last contest is excluded (because its fraction of last-time contestants would be 1.0).

Table 11: Benefiting from macro data availability, and the interaction with experience. Regressions of investors' out-of-sample performance against macro data availability interacted with experience. The dependent variable is the excess best live Sharpe Ratio over the Sharpe Ratio of the benchmark index during that out-of-sample period, which effectively controls for time periods/contests. In all specifications, the dummy "Macro variables available$_t$" indicates whether the contestant is a member of the Treatment group, defined in Section 6.1, or the Control group.

Panel A: Linear relationship

	Dependent variable:		
	Excess Live $SR_{i,t}^{Best}$		
	(1)	(2)	(3)
(Intercept)	-2.181^{***}	-1.959^{***}	-1.470^{**}
	(0.566)	(0.525)	(0.626)
Contests experienced$_{i,t}$	1.030^{***}	1.003^{***}	0.589^{***}
	(0.263)	(0.240)	(0.186)
Macro variables available$_t$		-0.318	-1.352
		(0.609)	(1.069)
Contests experienced$_{i,t}$ × Macro variables available$_t$			0.908^{**}
			(0.439)
Observations	830	830	830
R^2	0.048	0.053	0.063

Panel B: Indicator variables

	Dependent variable:		
	Excess Live $SR_{i,t}^{Best}$		
	(1)	(2)	(3)
(Intercept)	-1.162^{***}	-0.970^{**}	-0.899^{**}
	(0.341)	(0.425)	(0.451)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 2\}$	1.332^{***}	1.293^{***}	1.044^{***}
	(0.266)	(0.257)	(0.271)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 3\}$	1.673^{*}	1.654^{*}	-0.004
	(0.914)	(0.896)	(0.459)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 4\}$	3.021^{***}	2.918^{***}	2.245^{***}
	(0.433)	(0.421)	(0.507)
Macro variables available$_t$		-0.311	-0.428
		(0.614)	(0.659)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 2\}$ × Macro variables available$_t$			0.478
			(0.504)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 3\}$ × Macro variables available$_t$			2.973^{**}
			(1.491)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 4\}$ × Macro variables available$_t$			2.221^{***}
			(0.527)
Observations	830	830	830
R^2	0.050	0.055	0.067

Note: standard errors (in parentheses) are clustered by contest and contestant. *p<0.1; **p<0.05; ***p<0.01

Table 12: Ruling out selection effects. Heckman (1976) (i.e. Type II Tobit) two-stage selection models for implementing selection bias corrections to the regression of out-of-sample performance against experience. The first stage models the probability of participation, while the second stage models the outcome of interest.

		All contests	*Contests 1-7*	*Contests 8-12*
Stage		Live $SR_{i,t}^{Best}$	Live $SR_{i,t}^{Best}$	Live $SR_{i,t}^{Best}$
1. *Selection*	(Intercept)	1.95^{***}	2.38^{***}	1.50^{***}
		(0.18)	(0.29)	(0.34)
	Contests experienced$_{i,t}$	-0.88^{***}	-1.22^{***}	-0.77^{***}
		(0.05)	(0.10)	(0.05)
	Quantopian search index$_t$	-0.01^{**}	-0.00	-0.00
		(0.00)	(0.00)	(0.00)
	Ratio of entries to contest mean$_{i,t-1}$	0.23^{***}	0.24^{***}	0.22^{***}
		(0.04)	(0.08)	(0.04)
2. *Outcome*	(Intercept)	-0.42^{***}	-0.79^{**}	-0.04
		(0.13)	(0.32)	(0.17)
	Contests experienced$_{i,t}$	0.84^{***}	0.55	0.86^{***}
		(0.17)	(0.39)	(0.20)
	Inverse Mills Ratio	-0.83^{***}	0.33	-1.27^{***}
		(0.32)	(0.54)	(0.41)
	σ	2.13	1.89	2.31
	ρ	-0.39	0.17	-0.55
	R^2	0.04	0.07	0.04
	Num. obs.	1482	482	1000
	Censored	621	163	458
	Observed	861	319	542

Standard errors (in parentheses) computed according to Greene (1981). $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

Internet Appendix

A Additional Figures . . . 1
B Additional Tables . . . 5
C Entering a Trading Contest . . . 7
D Futures Contracts and the Benchmark Portfolio . . . 11
D.1 Market Data for Futures Contracts . . . 11
D.2 Benchmark Portfolio to Adjust for Market Conditions . . . 14
E Macroeconomic Variables Added in Between Contests . . . 15
F Response to Macroeconomic Variable Releases . . . 17
G Realized Moments of Daily Portfolio Returns . . . 19
H Data Abundance and Competition Effects . . . 23
I Robustness to Using the Official Score . . . 26
References in Internet Appendix . . . 29

A Additional Figures

Cross-sectional dispersion in OOS performance outcomes. Figure IA.1 illustrates the cross-sectional dispersion in the Live/out-of-sample Sharpe Ratios of machine-driven trading strategies in all 12 contests in the contest leaderboard sample. These are conditioned by the contest index, and thus depict the time-varying conditional distribution of investor performance outcomes. Immediately, it is evident that this distribution is widely dispersed. Some investors are able to benefit substantially by implementing machine-based trading strategies, with high empirical SR, while others experience negative SRs. The median empirical out-of-sample SR is negative for many contests, indicating that a large fraction of machine trading strategies do not benefit their human investor designers, when measured out-of-sample. Since contestants within the same contest have exactly the same information set, this figure also illustrates the extent of model disagreement.

Participation and attrition. Figure IA.2 shows that a high proportion of contestants in each contest are first-time participants and that, furthermore, many participate only once. The latter fact is consistent with the prior literature and intuition discussed by Linnainmaa (2011). One important conclusion from this chart is that investors do not appear to be noticeably more or less willing to participate in contests just before or after the new predictive variables were introduced in between contests 7 & 8; i.e. the introduction of the additional predictive variables does not appear to be a structural break in terms of participation.

Informativeness vs. familiarity at the individual macro variable level. Figure IA.3 illustrates to what extent the aggregate individual macro variable informativeness and familiarity indices are concentrated among the top 3 and top 10 ranked individual variables each year. Familiarity is highly concentrated, while informativeness is more evenly-spread. Figure IA.4 displays the correlations of individual macroeconomic variables' rankings of informativeness and familiarity, recomputed each year. Both familiarity measures are highly correlated at the individual variable level. Variables' informativeness and familiarity are only weakly correlated with one another; still the correlation is typically positive.

Figure IA.1: Cross-section of out-of-sample performance outcomes. Box-and-whisker plots of the distributions of Live (out-of-sample) trading strategy entry Sharpe Ratios, conditioning by contest index. The LHS plot shows all available entries; the RHS plot shows the best live entries per contest. Boxes cover the 25th and 75th percentiles of the empirical distribution; black dots denote the median value.

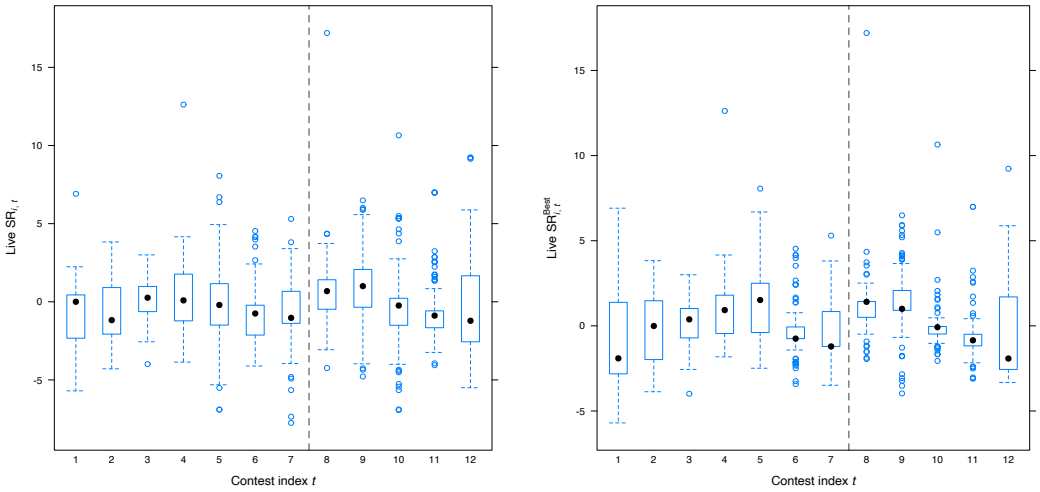


Figure IA.2: Participation and attrition. Proportions of contestants for whom this is the first contest, and those for whom this is the last contest (with breakdowns of the latter).

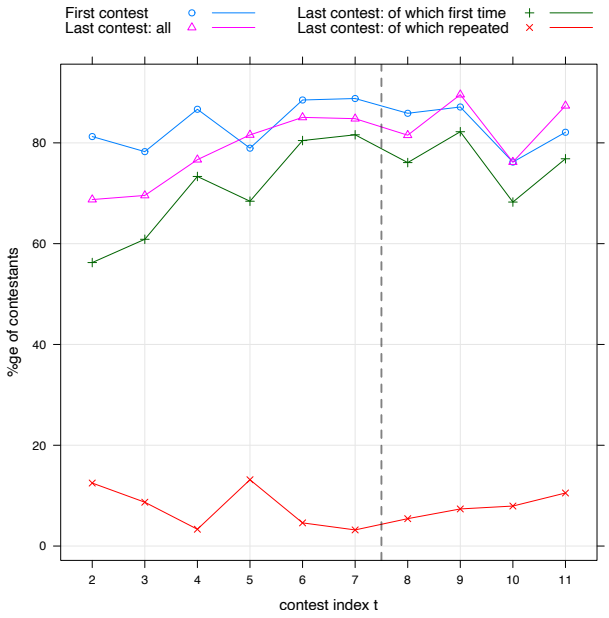


Figure IA.3: Concentration measures of individual macro variable informativeness and familiarity. Percentages of the total annual macro variable informativeness, and of the total annual macro variable mentions in news articles and books (i.e. familiarity), that are attributed to the top 3-ranked (LHS) and top 10-ranked (RHS) individual macro variables, for each year.

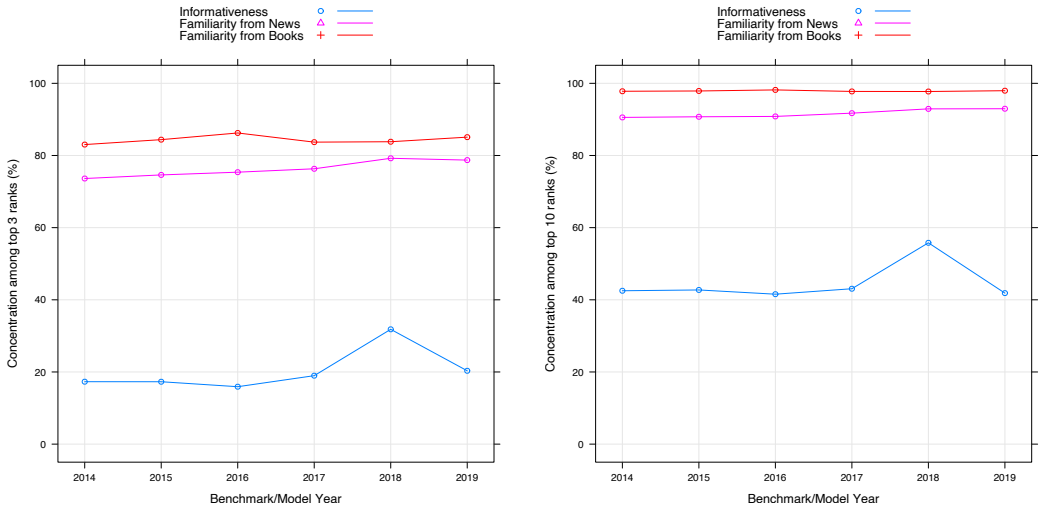
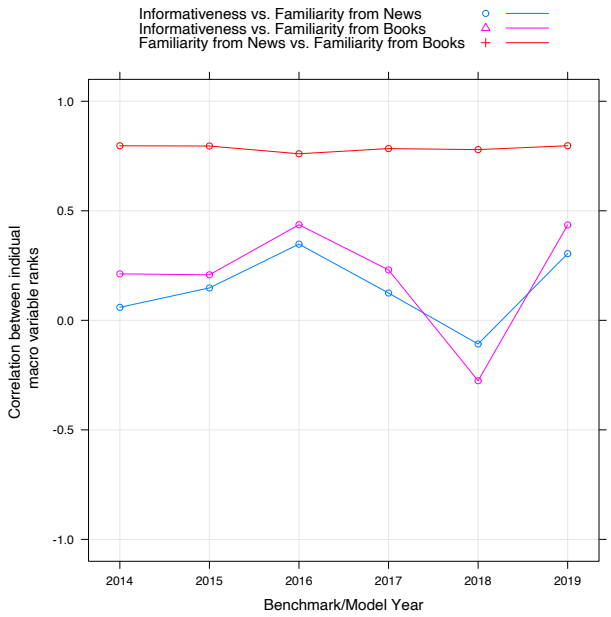


Figure IA.4: Rank-correlations of individual macro variable informativeness and familiarity. Correlations between macro variable informativeness and familiarity rankings. For both familiarity measures, rankings are computed each year using article and book mentions for that year. The informativeness measures are computed according to a benchmark machine learning model that uses historical data up to the end of the prior year.



B Additional Tables

Informativeness and familiarity at the individual macro variable level. Table IA.1 lists the top 20-ranked macro variables by informativeness and by each familiarity measure, aggregated over the benchmark period.

Turnover responses to both aggregate informativeness and familiarity. Table IA.2 combines the regression specifications of Tables 6 and 7 to jointly analyze the responses of turnover to macro release day informativeness and macro release day familiarity. The estimates in this joint specification are consistent with earlier findings from individual specifications.

Table IA.1: Mean macro variable rankings. Top-20 ranked macroeconomic predictive variables, by informativeness and familiarity, utilizing the means of annual 2014-2019 rankings.

Rank	By Informativeness	By Familiarity (News)	By Familiarity (Books)
1	Export Prices	Exports	Exports
2	Imports	Imports	Imports
3	Producer Prices Change	Unemployment Rate	Unemployment Rate
4	Inflation Rate	Industrial Production	Industrial Production
5	Import Prices	Business Confidence	Inflation Rate
6	Capacity Utilization	Inflation Rate	Capital Flows
7	Exports	Consumer Credit	Consumer Credit
8	Balance of Trade	Capacity Utilization	Balance of Trade
9	Core Inflation Rate	Export Prices	Capacity Utilization
10	ADP Employment Change	Producer Prices	Job Offers
11	Government Payrolls	Import Prices	Export Prices
12	Consumer Price Index CPI	Capital Flows	Import Prices
13	Industrial Production	Consumer Price Index CPI	Labor Force Participation Rate
14	Unemployment Rate	Job Offers	Business Confidence
15	Non Farm Payrolls	Manufacturing Production	Producer Prices
16	Philadelphia Fed Manufacturing Index	Non Farm Payrolls	Manufacturing Production
17	Manufacturing Payrolls	Pending Home Sales	Average Hourly Earnings
18	Housing Index	Average Hourly Earnings	Average Weekly Hours
19	Average Weekly Hours	Durable Goods Orders	Business Inventories
20	Manufacturing Production	Business Inventories	Core Inflation Rate

Table IA.2: Trading strategy responses to both released macro signal familiarity and informativeness. Standardized daily turnover of best live trading strategies on macro release days, regressed against the index of macro release day informativeness, as well as one of two indices of macro release day familiarity. The familiarity and informativeness indices are interacted with an indicator for investor experience. Familiarity measures are based on news article mentions in columns (1) and (2), and book mentions in columns (3) and (4), respectively.

Dependent Variable:	$\text{Volume}_{i,\tau}$			
Familiarity Index:	News Articles		Books	
	(1)	(2)	(3)	(4)
$\text{Familiarity}_{\tau}$	0.0016**	0.0016**	0.0017**	0.0017**
	(0.0006)	(0.0006)	(0.0005)	(0.0005)
$\text{Informativeness}_{\tau}$	0.5714*	0.5282*	0.6089**	0.5646*
	(0.2219)	(0.2462)	(0.2137)	(0.2378)
$\text{Experienced}_{i,t}$	-0.3047	-0.4709	-0.1192	-0.3483
	(0.3103)	(0.3715)	(0.2448)	(0.3374)
$\text{Experienced}_{i,t} \times \text{Familiarity}_{\tau}$	0.0039	0.0016	0.0019	0.0002
	(0.0031)	(0.0027)	(0.0024)	(0.0020)
$\text{Experienced}_{i,t} \times \text{Informativeness}_{\tau}$		1.006***		1.037***
		(0.1402)		(0.1944)
Benchmark/Model Year FEs	✓	✓	✓	✓
Observations	142,332	142,332	142,332	142,332
R^2	0.00055	0.00060	0.00054	0.00059

Note: standard errors (in parentheses) are clustered by contest and contestant.
*p<0.1; **p<0.05; ***p<0.01

C Entering a Trading Contest

To make the institutional setting of Section 3.1 more concrete, I review the steps necessary to participate in a contest from the point of view of an individual investor, and include screenshots to illustrate the process.

Figure IA.5 shows an example of how to write (in Python) a systematic trading strategy that simply takes equally-weighted long positions in all available futures contracts. This code is presented as a default template to the contestant by the Quantiacs web platform. More sophisticated strategies are, of course, encouraged.

After coding up such a systematic trading strategy, the systematic investor can then use the Quantiacs platform to run a backtest of the strategy using historical market data. Figure IA.6 shows the output of such a backtest, for the long-only strategy of my above example. Various performance metrics are provided, but all are calculated on historical data, and I refer to this pre-contest period as the in-sample/backtest period.

The systematic investor may then decide to officially submit this trading strategy to a contest: if so, I term this the contestant's "entry." Once a contest has begun (i.e. during the live/out-of-sample period), entries can no longer be modified, and the contestant has effectively committed to following that systematic trading strategy for the duration of the live/out-of-sample period. The distinction between the in-sample/backtest and out-of-sample/live periods for a single contest is illustrated by Figure 3 in the manuscript.

After the contest ends, the out-of-sample Sharpe Ratios are made available, the contestants are ranked, and results are displayed in a leaderboard, as in Figure IA.7.

Figure IA.5: Coding up a contest entry. Example of writing code (in Python) while logged into the trading platform in order to define a systematic trading strategy, ahead of possible entry into a trading contest. This code is presented as a default template to the contestant by the Quantiacs web platform. Lines 17-18 define a simple long-only equally-weighted portfolio, with no use of market data or macroeconomic variables. Note that uncommenting line 6 loads the `scikit-learn` (Pedregosa et al., 2011) machine learning software package.

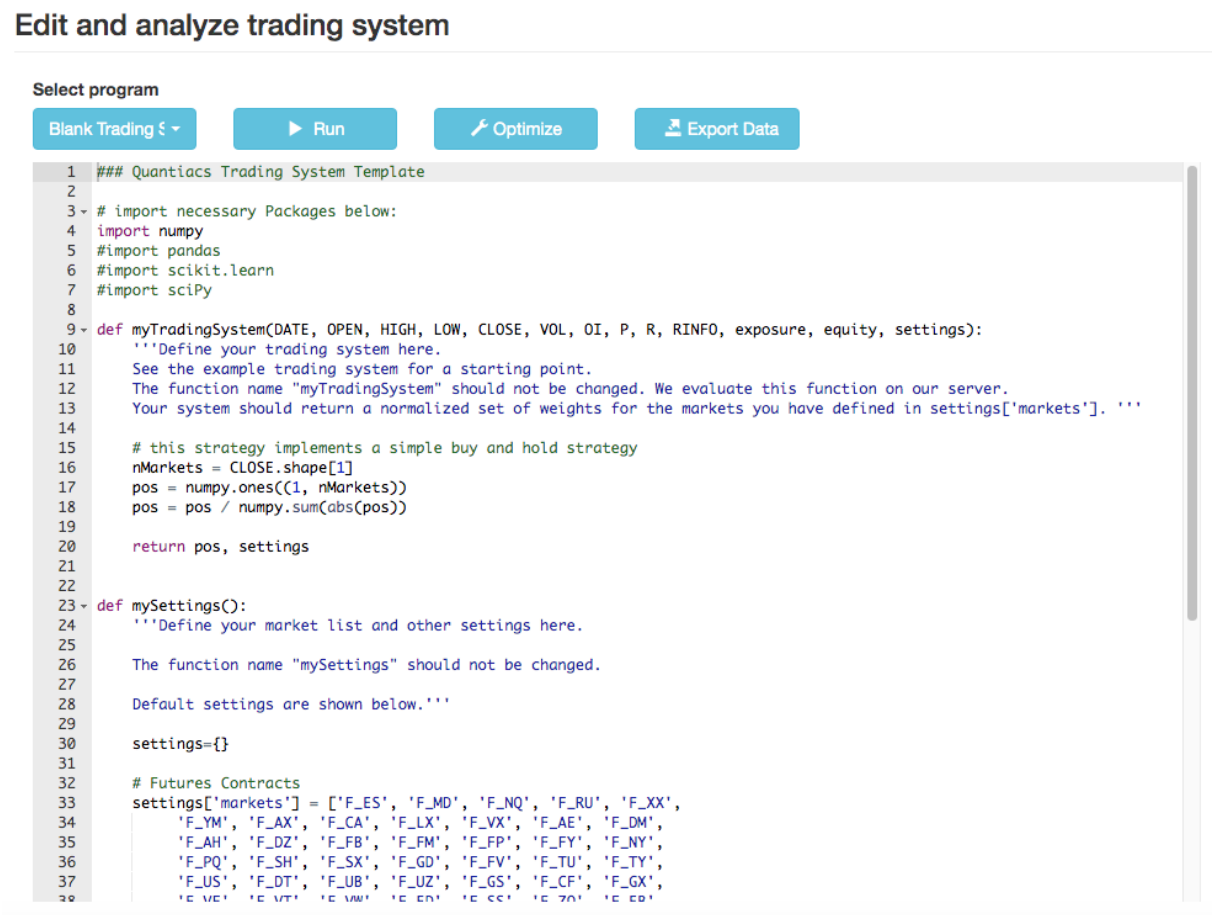


Figure IA.6: Backtesting prior to contest entry. Example of in-sample backtest results for a systematic trading strategy, before an entry is submitted. At this point, a trading contest has not (yet) begun.

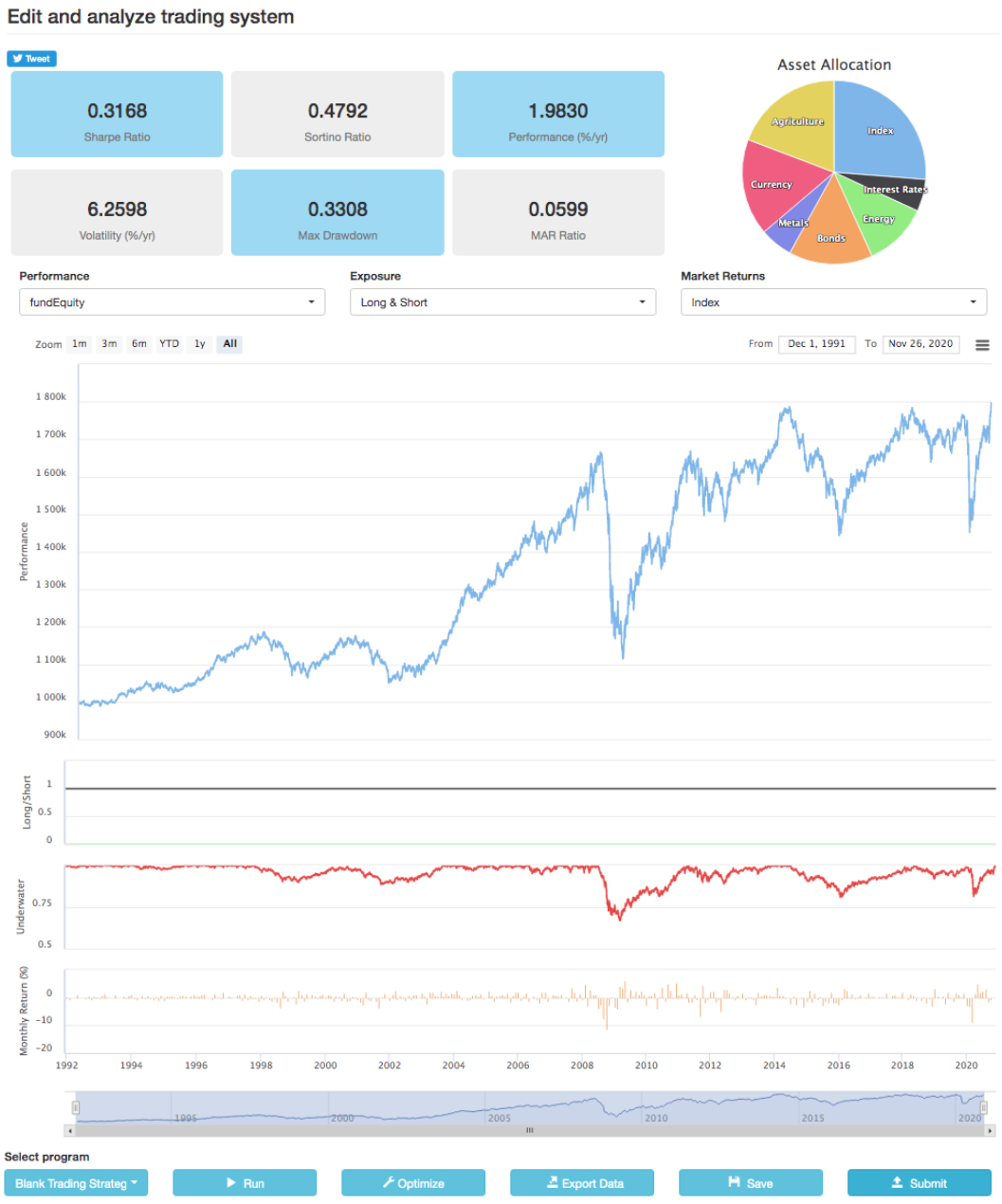
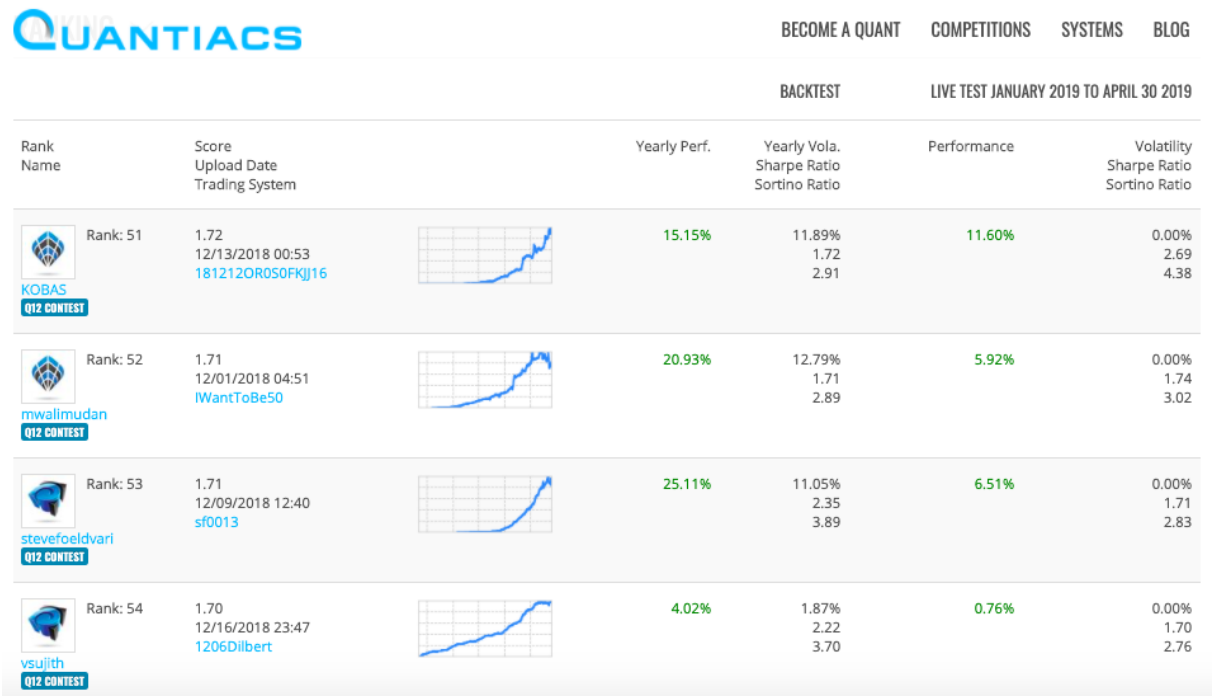


Figure IA.7: Contest outcomes once Live period ends. Excerpt from a contest leaderboard, showing both in-sample and out-of-sample Sharpe Ratios, after a trading contest has ended. The "live test" performance metrics were calculated once the contest period (1 January to 30 April 2019, in this case) had ended.



D Futures Contracts and the Benchmark Portfolio

D.1 Market Data for Futures Contracts

As discussed in Section 3 of the paper, the Quantiacs trading platform allowed investors to take positions in 88 futures contracts, provided investors with historical market data for this investment universe, and used live market data during the out-of-sample evaluation period for each contest. For each future, the trading platform selected a single contract maturity that was available for trading at each point in time.

Like the platform, I use actual market data for the same universe of futures contracts, with the exception of the Russell 2000 index future, for which I simply use the daily index level.[25] I source the data from Bloomberg, stitching together multiple actively-traded futures contracts in a contiguous ratio-adjusted daily price series; this standard methodology ensures that return calculations are accurate. The futures (and their Bloomberg identifiers) are listed in Table IA.3.

Table IA.3: Universe of futures. Futures instruments available on Quantiacs for contestants' trading strategies to take daily positions in. These are also the instruments I used to construct the benchmark index.

	Quantiacs ticker	Name	Type	Matching Bloomberg base code
1	F_AD	Australian Dollar	Currency	AD Curncy
2	F_AE	AEX Index	Index	EO Index
3	F_AH	Bloomberg Commodity Index	Index	DN Index
4	F_AX	DAX	Index	GX Index
5	F_BC	Brent Crude Oil	Energy	CO Comdty
6	F_BG	Gas Oil	Energy	QS Comdty
7	F_BO	Soybean Oil	Agriculture	BO Comdty
8	F_BP	British Pound	Currency	BP Curncy
9	F_C	Corn	Agriculture	C Comdty
10	F_CA	CAC40	Index	CF Index
11	F_CC	Cocoa	Agriculture	CC Comdty

[25] I use the Russell 2000 index itself to proxy for the series of Russell 2000 index futures contracts because I am unable to source and combine price data for these futures contracts: the Russell 2000 index future has changed its listing multiple times between the ICE and CME exchanges and I do not have access to older historical data. I judge it more useful for my benchmark to possess a long timeseries of the underlying index over the full backtest and live periods. The downside is that the absence of the "basis" between the derivative price and underlying price means that contango/backwardation effects will be omitted, but I expect these to be negligible on the most-active contract of a non-commodity future — such as this one.

	Quantiacs ticker	Name	Type	Matching Bloomberg base code
12	F_CD	Canadian Dollar	Currency	CD Curncy
13	F_CF	10y Swiss Note	Bond	SWC Comdty
14	F_CL	WTI Crude Oil	Energy	CL Comdty
15	F_CT	Cotton	Agriculture	CT Comdty
16	F_DL	Milk Class III	Agriculture	DA Comdty
17	F_DM	MDAX Index	Index	MF Index
18	F_DT	EURO Bond	Bond	RX Comdty
19	F_DX	US Dollar Index	Currency	DX Curncy
20	F_DZ	TechDAX	Index	DP Index
21	F_EB	3-Month EuriBor	Interest Rate	ER Comdty
22	F_EC	Euro FX	Currency	EC Curncy
23	F_ED	Eurodollars	Interest Rate	ED Comdty
24	F_ES	E-mini S&P 500 Index	Index	ES Index
25	F_F	3-Month EuroSwiss	Interest Rate	ES Comdty
26	F_FB	DJ Stoxx Bank 600	Index	BJ Index
27	F_FC	Feeder Cattle	Agriculture	FC Comdty
28	F_FL	Chicago Ethanol	Energy	CUA Comdty
29	F_FM	Stoxx Europe Mid 200	Index	SXR Index
30	F_FP	OMX Helsinki 25	Index	OT Index
31	F_FV	5-year Treasury Note	Bond	FV Comdty
32	F_FY	Stoxx Europe 600	Index	SXO Index
33	F_GC	Gold	Metal	GC Comdty
34	F_GD	Goldman Sachs Commodity Index	Index	GI Index
35	F_GS	10-Year Long Gilt	Bond	G Comdty
36	F_GX	Euro BUXL	Bond	UB Comdty
37	F_HG	Copper	Metal	HG Comdty
38	F_HO	Heating Oil	Energy	HO Comdty
39	F_HP	Natural Gas Penultimate	Energy	ZA Comdty
40	F_JY	Japanese Yen	Currency	JY Curncy
41	F_KC	Coffee	Agriculture	KC Comdty
42	F_LB	Lumber	Agriculture	LB Comdty
43	F_LC	Live Cattle	Agriculture	LC Comdty
44	F_LN	Lean Hogs	Agriculture	LH Comdty
45	F_LQ	Newcastle Coal	Energy	XW Comdty
46	F_LR	Brazilian Real	Currency	BR Curncy
47	F_LU	Rotterdam Coal	Energy	XA Comdty
48	F_LX	FTSE 100 Index	Index	Z Index
49	F_MD	E-mini S&P 400	Index	FA Index
50	F_MP	Mexican Peso	Currency	PE Curncy
51	F_ND	New Zealand Dollar	Currency	NV Curncy
52	F_NG	Natural Gas	Energy	NG Comdty

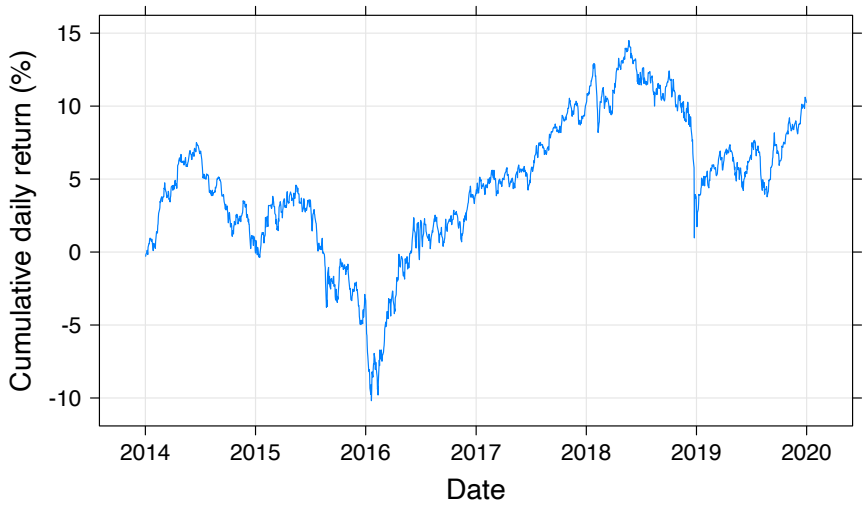
	Quantiacs ticker	Name	Type	Matching Bloomberg base code
53	F_NQ	E-mini Nasdaq 100 Index	Index	NQ Index
54	F_NR	Rough Rice	Agriculture	RR Comdty
55	F_NY	Nikkei 225	Index	NI Index
56	F_O	Oats	Agriculture	O Comdty
57	F_OJ	Orange Juice	Agriculture	JO Comdty
58	F_PA	Palladium	Metal	PA Comdty
59	F_PL	Platinum	Metal	PL Comdty
60	F_PQ	PSI20	Index	PP Index
61	F_RB	Gasoline	Energy	XB Comdty
62	F_RF	EURO FX/Swiss Franc	Currency	RF Curncy
63	F_RP	EURO FX/British Pound	Currency	RP Curncy
64	F_RR	Russian Ruble	Currency	RU Curncy
65	F_RU	Russell 2000	Index	RTY Index
66	F_RY	EURO FX/Japanese Yen	Currency	RY Curncy
67	F_S	Soybeans	Agriculture	S Comdty
68	F_SB	Sugar	Agriculture	SB Comdty
69	F_SF	Swiss Franc	Currency	SF Curncy
70	F_SH	Swiss Mid Cap	Index	S1 Index
71	F_SI	Silver	Metal	SI Comdty
72	F_SM	Soybean Meal	Agriculture	SM Comdty
73	F_SS	3-Month Short Sterling	Interest Rate	L Comdty
74	F_SX	Swiss Market	Index	SM Index
75	F_TR	South African Rand	Currency	RA Curncy
76	F_TU	2-year Treasury Note	Bond	TU Comdty
77	F_TY	10-year Treasury Note	Bond	TY Comdty
78	F_UB	EURO Bobl	Bond	OE Comdty
79	F_US	30-year Treasury Bond	Bond	US Comdty
80	F_UZ	EURO Schatz	Bond	DU Comdty
81	F_VF	5-Year Euro Swapnote	Bond	T Comdty
82	F_VT	10-Year Euro Swapnote	Bond	P Comdty
83	F_VW	2-Year Euro Swapnote	Bond	RW Comdty
84	F_VX	Volatilty Index	Index	UX Index
85	F_W	Wheat	Agriculture	W Comdty
86	F_XX	Dow Jones STOXX 50	Index	VG Index
87	F_YM	E-mini Dow Jones Industrial Average	Index	DM Index
88	F_ZQ	30-Day Fed Funds	Interest Rate	FF Comdty

D.2 Benchmark Portfolio to Adjust for Market Conditions

The trading platform does not formally judge contestants against a benchmark, though it does provide each contestant with an example systematic trading strategy that takes equally-weighted long positions in each of the 88 futures contracts available on the platform (Figure IA.5). Benchmark indices exist for certain sectors – such as the 24-contract Goldman Sachs Commodity Index (GSCI) – but I am not aware of any benchmark for futures as an overall asset class. I therefore construct the benchmark similarly to the simple default trading strategy suggested by Quantiacs to new users: based on daily-rebalanced equally-weighted returns of long positions in the most-active contracts of the underlying 88 futures (or, in the special case of the Russell 2000 index, the future's underlying index itself). A timeseries of returns for the benchmark portfolio is displayed in Figure IA.8.

Besides the fact that this benchmark portfolio is the default presented to new users, the methodology has additional merits. The use of long-only positions (similar to the GSCI commodity index methodology) is justified because the buyers of futures contracts receive delivery of the underlying physical/cash asset at maturity. The use of equally-weighted positions is justified by the absence of any alternative for weighting derivatives in such diverse underlying assets: unlike stocks or bonds, there is no fixed supply that limits the open interest that is possible in futures contracts.

Figure IA.8: Benchmark portfolio. Cumulative return of the benchmark portfolio over the years in which contest Live periods took place.



E Macroeconomic Variables Added in Between Contests

On 22 March 2017, Quantiacs announced in a blog post that 54 macroeconomic data series would be made available to all contestants from the 8th contest onwards, for use in both backtesting and live trading. The variables are listed in Table IA.4.

Table IA.4: Macroeconomic indicators. Macroeconomic variables added to the Quantiacs platform from the 8th trading contest onwards, which could be used as input predictive variables for all contestants' trading strategies from that contest onwards.

	Macroeconomic variable	**Quantiacs identifier**
1	ADP Employment Change	USA_ADP
2	Average Hourly Earnings	USA_EARN
3	Average Weekly Hours	USA_HRS
4	Balance of Trade	USA_BOT
5	Business Confidence	USA_BC
6	Business Inventories	USA_BI
7	Capacity Utilization	USA_CU
8	Capital Flows	USA_CF
9	Challenger Job Cuts	USA_CHJC
10	Chicago Fed National Activity Index	USA_CFNAI
11	Chicago Pmi	USA_CP
12	Consumer Credit	USA_CCR
13	Consumer Price Index CPI	USA_CPI
14	Core Consumer Prices	USA_CCPI
15	Core Inflation Rate	USA_CINF
16	Dallas Fed Manufacturing Index	USA_DFMI
17	Durable Goods Orders	USA_DUR
18	Durable Goods Orders Ex Transportation	USA_DURET
19	Export Prices	USA_EXPX
20	Exports	USA_EXVOL
21	Factory Orders Ex Transportation	USA_FRET
22	Foreign Bond Investment	USA_FBI
23	Government Budget Value	USA_GBVL
24	Government Payrolls	USA_GPAY
25	Housing Index	USA_HI
26	Import Prices	USA_IMPX
27	Imports	USA_IMVOL
28	Industrial Production	USA_IP
29	Industrial Production Mom	USA_IPMOM
30	Inflation Rate	USA_CPIC
31	Inflation Rate Mom	USA_CPICM

	Macroeconomic variable	Quantiacs identifier
32	Job Offers	USA_JBO
33	Labor Force Participation Rate	USA_LFPR
34	Leading Economic Index	USA_LEI
35	Manufacturing Payrolls	USA_MPAY
36	Manufacturing Production	USA_MP
37	Nahb Housing Market Index	USA_NAHB
38	Net Long Term Tic Flows	USA_NLTTF
39	NFIB Business Optimism Index	USA_NFIB
40	Non Farm Payrolls	USA_NFP
41	Non Manufacturing PMI	USA_NMPMI
42	Nonfarm Payrolls Private	USA_NPP
43	NY Empire State Manufacturing Index	USA_EMPST
44	Pending Home Sales	USA_PHS
45	Philadelphia Fed Manufacturing Index	USA_PFED
46	Producer Prices	USA_PP
47	Producer Prices Change	USA_PPIC
48	Retail Sales MoM	USA_RSM
49	Retail Sales YoY	USA_RSY
50	Retail Sales Ex Autos	USA_RSEA
51	Richmond Fed Manufacturing Index	USA_RFMI
52	Total Vehicle Sales	USA_TVS
53	Unemployment Rate	USA_UNR
54	Wholesale Inventories	USA_WINV

F Response to Macroeconomic Variable Releases

This appendix analyzes the responses of investors' trading strategy turnovers to the release/update of macroeconomic variables on the platform. As a placebo test, I repeat this analysis for trading strategies that did not actually have access to these macroeconomic indicators (see Figure 4 in the paper).

I regress the standardized daily trading volumes of investors' best Live-period entries against various indicators for experience, the occurrence of macro release days, the availability of macro data in the current contest, and their interactions, and present the results in Table IA.5. The first two columns focus on investors who could access the macroeconomic predictive variables and use their realizations to make systematic trading decisions. The coefficient on the interaction in column (2) captures the interaction between experience and the timing of macro variable releases. Its value is positive and significant, which suggests that investors with access to the macro predictive variables trade significantly more than their own daily average on the days when new releases of these macro variables were updated on the Quantiacs trading platform.

To complete the placebo test, the second two columns in Table IA.5 focus on investors who did not have access to the macro predictors. The coefficient on the interaction in column (4) is not significantly different to zero: there was no significant above- or below-average change to the placebo group investors' daily trading volumes on the days the macro variables would have been released to the platform (in the counterfactual situation of them entering that trading strategy to a later contest).

This analysis suggests that investors' trading strategies respond to the release of macroeconomic predictive variables. There is also a differential response based on the human investors' experience. For the placebo in which macroeconomic indicators were not available, there was no discernible response by the trading strategies implemented by either experienced or inexperienced human investors on macro release days.

Table IA.5: Daily trading volume and macro release days. Regressions of the daily volume (standardized) of investors' best live trading strategies against an indicator for the daily release of macro variables to contestants interacted with an indicator for experienced investors (who are taking part in their 3rd or 4th contest). Columns (1) and (2) are for the sub-sample of investors who could access the macro variables. Columns (3) and (4) are for the sub-sample of investors who could not access the macro variables via their trading strategies.

	$\text{Volume}_{i,\tau}$			
	With data availability		*Placebo group*	
	(1)	(2)	(3)	(4)
(Intercept)	0.002^{***}	0.003^{***}	0.002^{***}	0.002^{***}
	(0.0004)	(0.0001)	(0.001)	(0.001)
$\text{Macro release}_{\tau}$	-0.009^{***}	-0.011^{***}	0.0002	0.0002
	(0.003)	(0.002)	(0.012)	(0.012)
$\text{Experienced}_{i,t}$		−0.009		0.0001
		(0.007)		(0.001)
$\text{Macro release}_{\tau} \times \text{Experienced}_{i,t}$		0.076^{**}		0.001
		(0.036)		(0.017)
Observations	3,234,115	3,234,115	2,059,615	2,059,615
R^2	0.00000	0.00001	0.000	0.000

Note: standard errors (in parentheses) are clustered by contest and contestant.
*p<0.1; **p<0.05; ***p<0.01

G Realized Moments of Daily Portfolio Returns

This appendix examines the realized moments of investors' daily portfolio returns, and relates them to experience and data availability.

An investor who seeks to maximize her out-of-sample portfolio Sharpe Ratio, as Quantiacs contestants do, is faced with the joint problem of maximizing the out-of-sample returns of the portfolio while minimizing their variance; this is due to the construction of the Sharpe Ratio itself. Using timeseries of trading strategy daily returns, I decompose the realized Sharpe Ratios by estimating the mean and standard deviation of these contestants' out-of-sample portfolio daily returns. I then relate these to experience levels and the availability of additional predictive variables.

I modify the specification (18) in the paper by regressing the *ex post* estimated means of the out-of-sample daily returns of trading strategies against interacted dummies for the availability of additional predictive variables and investor experience levels:

$$\begin{aligned}
&\text{Mean}_{i,t}(\text{Best Live entry daily returns}) \\
&= \beta_1 + \sum_{k=2}^{4}\Big[\beta_k \times \mathbb{1}\{\text{Contests experienced}_{i,t} = k\}\Big] + \gamma_1 \times \text{New variables available}_t \\
&\quad + \sum_{k=2}^{4}\Big[\gamma_k \times \mathbb{1}\{\text{Contests experienced}_{i,t} = k\} \times \text{New variables available}_t\Big] + \epsilon_{i,t}. \qquad (38)
\end{aligned}$$

Regression results are displayed in Table IA.6, with the dependent variable in columns (1)-(3) constructed using raw returns, and the dependent variable in columns (4)-(6) adjusting for the benchmark portfolio's daily returns as an implicit control. Whether or not the benchmark adjustment is applied, the availability of additional predictive variables is associated with higher out-of-sample mean returns for more experienced investors, and the estimates for these coefficients γ_k are significantly different to zero.

A similar modification of the main specification (18) and previous specification (38) is to regress the *ex post* estimated standard deviations of the out-of-sample daily returns of trading strategies against interacted dummies for the availability of additional predictive variables and

investor experience levels:

$$\begin{aligned}
&\text{SD}_{i,t}(\text{Best Live entry daily returns}) \\
&= \beta_1 + \sum_{k=2}^{4}\Big[\beta_k \times \mathbb{1}\{\text{Contests experienced}_{i,t} = k\}\Big] + \gamma_1 \times \text{New variables available}_t \\
&\quad + \sum_{k=2}^{4}\Big[\gamma_k \times \mathbb{1}\{\text{Contests experienced}_{i,t} = k\} \times \text{New variables available}_t\Big] + \epsilon_{i,t}. \qquad (39)
\end{aligned}$$

Regression results are displayed in Table IA.7. As above, columns (1)-(3) use raw returns and columns (4)-(6) adjust for the benchmark portfolio's daily returns before estimating their standard deviation, in order to implicitly control for differing contest time periods. Once again, the conclusions do not depend on whether the benchmark adjustment is applied. While the estimates for coefficients γ_k in columns (3) & (6) again agree with intuition, this time they are mostly not significantly different to zero.

These results indicate that the improvements in out-of-sample Sharpe Ratios that are associated with the availability of additional predictive variables (in Table 11, for example) can be attributed mainly to higher out-of-sample mean returns (Table IA.6). While I also detect lower out-of-sample standard deviations, these additional effects are mostly statistically insignificant (Table IA.7).

Table IA.6: Means of daily trading strategy returns. Regressions of investors' out-of-sample daily return means against additional data availability & experience. The dependent variable in columns (1)-(3) does not incorporate a benchmark, so there is no control for time periods; the dependent variable in columns (4)-(6) adjusts the daily returns using the benchmark portfolio. In all specifications, the dummy "Macro variables available$_t$" indicates whether the contestant is a member of the Treatment group, defined in Section 6.1, or the Control group.

	Dependent variable:					
	Mean of Live daily returns (%) for contestant i's best entry in contest t					
	Raw returns			*Excess returns over the benchmark*		
	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	−0.143**	−0.006	0.008	−0.159**	−0.012	0.002
	(0.066)	(0.009)	(0.006)	(0.066)	(0.009)	(0.006)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 2\}$	0.165**	0.137**	0.022	0.172**	0.142**	0.025
	(0.067)	(0.057)	(0.016)	(0.067)	(0.057)	(0.016)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 3\}$	0.148**	0.130*	−0.016	0.161**	0.141**	−0.018
	(0.067)	(0.067)	(0.011)	(0.067)	(0.068)	(0.014)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 4\}$	0.212***	0.149***	0.067***	0.208***	0.141**	0.048***
	(0.068)	(0.057)	(0.023)	(0.068)	(0.061)	(0.018)
Macro variables available$_t$		−0.223**	−0.246**		−0.238**	−0.262**
		(0.098)	(0.108)		(0.098)	(0.108)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 2\} \times$ Macro variables available$_t$			0.230**			0.233**
			(0.110)			(0.110)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 3\} \times$ Macro variables available$_t$			0.270**			0.295***
			(0.109)			(0.109)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 4\} \times$ Macro variables available$_t$			0.225**			0.258**
			(0.111)			(0.110)
Observations	713	713	713	713	713	713
R^2	0.001	0.005	0.006	0.001	0.006	0.007

Note: standard errors (in parentheses) are clustered by contest. *p<0.1; **p<0.05; ***p<0.01

Table IA.7: Standard deviations of daily trading strategy returns. Regressions of investors' out-of-sample daily return standard deviations against additional data availability & experience. The dependent variable in columns (1)-(3) does not incorporate a benchmark, so there is no control for time periods; the dependent variable in columns (4)-(6) adjusts the daily returns using the benchmark portfolio. In all specifications, the dummy "Macro variables available$_t$" indicates whether the contestant is a member of the Treatment group, defined in Section 6.1, or the Control group.

	Dependent variable:					
	SD of Live daily returns (%) for contestant i's best entry in contest t					
	Raw returns			*Excess returns over the benchmark*		
	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	0.816***	0.705***	0.686***	0.893***	0.769***	0.747***
	(0.140)	(0.060)	(0.062)	(0.139)	(0.058)	(0.060)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 2\}$	−0.389***	−0.367***	−0.226**	−0.380***	−0.355***	−0.175**
	(0.146)	(0.124)	(0.089)	(0.144)	(0.124)	(0.085)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 3\}$	−0.416***	−0.401***	−0.189	−0.367**	−0.351**	−0.101
	(0.155)	(0.147)	(0.129)	(0.159)	(0.153)	(0.149)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 4\}$	−0.413***	−0.362***	−0.225**	−0.478***	−0.421***	−0.316***
	(0.152)	(0.124)	(0.096)	(0.144)	(0.108)	(0.080)
Macro variables available$_t$		0.181	0.211		0.202	0.238
		(0.210)	(0.233)		(0.209)	(0.232)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 2\} \times$ Macro variables available$_t$			−0.279			−0.358
			(0.246)			(0.243)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 3\} \times$ Macro variables available$_t$			−0.393			−0.462*
			(0.266)			(0.275)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 4\} \times$ Macro variables available$_t$			−0.386			−0.285
			(0.245)			(0.240)
Observations	713	713	713	713	713	713
R^2	0.001	0.002	0.002	0.001	0.002	0.002

Note: standard errors (in parentheses) are clustered by contest. *p<0.1; **p<0.05; ***p<0.01

H Data Abundance and Competition Effects

This appendix checks whether and how competitive pressures can explain the benefits that investors derive (or fail to gain) from access to additional predictive variables. The analysis indicates that an interaction between competitive intensity and data abundance (Dugast and Foucault, 2021) does not drive the results in the Quantiacs setting. The regression results that I present next do not apply any clustering to standard errors, in order to set a lower hurdle for coefficient significance to overcome; nonetheless, the key coefficients of interest remains non-significantly different to zero, as I explain shortly.

I proxy for the competitive intensity experienced by an investor by the (log of the) total number of contestants taking part in the same competition. Table IA.8 displays the results of OLS regressions of this competitive intensity against investors' out-of-sample performance outcomes (with the usual benchmark adjustment). I find significant and negative coefficients on competitive intensity, indicating that individual investors do seem to experience poorer absolute performance outcomes in highly-competitive contests. However, the coefficients on the interaction of competitive intensity with the availability of additional predictors are not significantly different to zero, suggesting that competition does not impact on any data abundance-related gains or losses experienced by an investor. Furthermore, the main results (that are repeated in columns (1) & (3)) are not qualitatively affected by the inclusion of the additional covariates in columns (4) & (6), respectively.

I next conduct an additional and related analysis of the effort a contestant exerts (proxied by the number of entries she makes into a competition) and compare it to the competitive pressure that she experiences. Table IA.9 shows Poisson GLM regressions that find no association between an investor's observed effort and the competitive pressures she experiences, controlling for her experience level. Furthermore, there is also no interaction with the availability of additional predictive variables. I conclude there is no direct relationship between data abundance and the effort exerted by a competitor in my particular setting.

Table IA.8: Competitive intensity interacted with the availability of macro signals. Regressions of investors' out-of-sample performance (over the benchmark SR) against measures of experience, availability of additional predictive variables, and competitive intensity. In all specifications, the dummy "Macro variables available$_t$" indicates whether the contestant is a member of the Treatment group, defined in Section 6.1, or the Control group.

	Dependent variable:					
	Excess Live $SR_{i,t}^{Best}$					
	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	-1.162^{***}	-0.970^{***}	-0.899^{***}	4.777^{***}	5.073^{***}	5.214^{***}
	(0.075)	(0.127)	(0.130)	(0.667)	(0.791)	(0.795)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 2\}$	1.332^{***}	1.293^{***}	1.044^{**}	1.033^{***}	1.054^{***}	0.684
	(0.355)	(0.359)	(0.444)	(0.347)	(0.342)	(0.426)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 3\}$	1.673^{***}	1.654^{***}	-0.004	1.428^{**}	1.390^{**}	-0.168
	(0.523)	(0.545)	(0.395)	(0.607)	(0.590)	(0.496)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 4\}$	3.021^{***}	2.918^{***}	2.245^{***}	3.009^{***}	3.176^{***}	2.636^{***}
	(0.490)	(0.529)	(0.504)	(0.545)	(0.496)	(0.562)
Macro variables available$_t$		-0.311^{**}	-0.428^{***}		2.569^{*}	2.100
		(0.153)	(0.159)		(1.428)	(1.418)
Macro variables available$_t$ × $\mathbb{1}\{\text{Contests experienced}_{i,t} = 2\}$			0.478			0.722
			(0.717)			(0.682)
Macro variables available$_t$ × $\mathbb{1}\{\text{Contests experienced}_{i,t} = 3\}$			2.973^{***}			2.801^{***}
			(0.809)			(0.943)
Macro variables available$_t$ × $\mathbb{1}\{\text{Contests experienced}_{i,t} = 4\}$			2.221^{***}			1.756^{***}
			(0.714)			(0.643)
Log Contestants$_t$				-1.311^{***}	-1.448^{***}	-1.463^{***}
				(0.143)	(0.175)	(0.176)
Macro variables available$_t$ × Log Contestants$_t$					-0.431	-0.357
					(0.307)	(0.305)
Observations	830	830	830	830	830	830
R^2	0.050	0.055	0.067	0.180	0.196	0.207

Note: standard errors (in parentheses) are robust to heteroskedasticity. *p<0.1; **p<0.05; ***p<0.01

Table IA.9: Competitive intensity. Poisson GLM regressions of contestants' number of entries in a contest against measures of experience, availability of additional predictive variables, and competitive intensity. In all specifications, the dummy "Macro variables available$_t$" indicates whether the contestant is a member of the Treatment group, defined in Section 6.1, or the Control group.

	Dependent variable:			
	Entries$_{i,t}$			
	(1)	(2)	(3)	(4)
(Intercept)	0.737***	1.085**	1.080**	0.885*
	(0.074)	(0.435)	(0.460)	(0.502)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 2\}$	0.812***	0.794***	0.794***	0.792***
	(0.168)	(0.172)	(0.172)	(0.173)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 3\}$	1.537***	1.523***	1.523***	1.505***
	(0.350)	(0.352)	(0.353)	(0.347)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 4\}$	2.070***	2.070***	2.069***	2.055***
	(0.389)	(0.393)	(0.398)	(0.393)
Log Contestants$_t$		−0.077	−0.075	−0.027
		(0.094)	(0.107)	(0.115)
Macro variables available$_t$			−0.004	1.138
			(0.157)	(1.267)
Macro variables available$_t$ × Log Contestants$_t$				−0.249
				(0.270)
Observations	830	830	830	830

Note: standard errors (in parentheses) are robust to heteroskedasticity.
*p<0.1; **p<0.05; ***p<0.01

I Robustness to Using the Official Score

This appendix repeats key results that use contestants' out-of-sample Sharpe Ratios as dependent variables and instead uses the official scores assigned by the Quantiacs platform:

- Table IA.10 modifies the specification of Table 11.
- Table IA.11 modifies the specification of Table 12.

There is little change to the estimates and no change to the conclusions drawn.

Table IA.10: Benefiting from macro data availability, and its interaction with experience. Regressions of investors' official scores against additional data availability interacted with experience. The dependent variable in columns (1)-(3) does not incorporate a benchmark, so there is no control for time periods; the dependent variable in columns (4)-(6) is the excess best official score over the Sharpe Ratio of the benchmark index during the matched time period, which effectively controls for time periods/contests. In all specifications, the dummy "Macro variables available$_t$" indicates whether the contestant is a member of the Treatment group, defined in Section 6.1, or the Control group.

	Dependent variable:					
	$\text{Score}_{i,t}^{Best}$			$\text{Score}_{i,t}^{Best} -$ Benchmark SR_t		
	(1)	(2)	(3)	(4)	(5)	(6)
(Intercept)	-0.592^{**}	-0.716^{***}	-0.698^{***}	-2.002^{***}	-1.754^{***}	-1.704^{***}
	(0.232)	(0.127)	(0.133)	(0.376)	(0.350)	(0.367)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 2\}$	0.678^{***}	0.703^{***}	0.671^{***}	0.945^{***}	0.895^{***}	0.840^{***}
	(0.221)	(0.192)	(0.255)	(0.192)	(0.201)	(0.244)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 3\}$	0.920	0.932	0.644^{***}	1.746^{***}	1.722^{***}	0.571^{*}
	(0.585)	(0.573)	(0.209)	(0.435)	(0.420)	(0.323)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 4\}$	2.493^{***}	2.560^{***}	2.020^{***}	2.770^{***}	2.637^{***}	1.530^{***}
	(0.675)	(0.577)	(0.426)	(0.731)	(0.704)	(0.444)
Macro variables available$_t$		0.201	0.172		-0.403	-0.483
		(0.342)	(0.384)		(0.660)	(0.667)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 2\} \times$ Macro variables available$_t$			0.057			0.092
			(0.382)			(0.407)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 3\} \times$ Macro variables available$_t$			0.515			2.063^{***}
			(1.034)			(0.626)
$\mathbb{1}\{\text{Contests experienced}_{i,t} = 4\} \times$ Macro variables available$_t$			1.856^{**}			3.783^{***}
			(0.855)			(0.511)
Observations	830	830	830	830	830	830
R^2	0.061	0.068	0.072	0.065	0.080	0.097

Note: standard errors (in parentheses) are clustered by contest and contestant. *p<0.1; **p<0.05; ***p<0.01

Table IA.11: Ruling out selection effects. Heckman (1976) (i.e. Type II Tobit) two-stage selection models for implementing selection bias corrections to the regression of official score against experience. The first stage models the probability of participation. The second stage models the outcome of interest.

		All contests	*Contests 1-7*	*Contests 8-12*
Stage		$\text{Score}_{i,t}^{Best}$	$\text{Score}_{i,t}^{Best}$	$\text{Score}_{i,t}^{Best}$
1. *Selection*	(Intercept)	1.95^{***}	2.38^{***}	1.50^{***}
		(0.18)	(0.29)	(0.34)
	$\text{Contests experienced}_{i,t}$	-0.88^{***}	-1.22^{***}	-0.77^{***}
		(0.05)	(0.10)	(0.05)
	$\text{Quantopian search index}_{t}$	-0.01^{**}	-0.00	-0.00
		(0.00)	(0.00)	(0.00)
	$\text{Ratio of entries to contest mean}_{i,t-1}$	0.23^{***}	0.24^{***}	0.22^{***}
		(0.04)	(0.08)	(0.04)
2. *Outcome*	(Intercept)	-0.99^{***}	-0.85^{***}	-0.14
		(0.08)	(0.30)	(0.14)
	$\text{Contests experienced}_{i,t}$	0.75^{***}	0.51	0.70^{***}
		(0.11)	(0.36)	(0.17)
	Inverse Mills Ratio	-0.76^{***}	0.39	-0.94^{***}
		(0.19)	(0.49)	(0.36)
	σ	1.33	1.74	2.00
	ρ	-0.57	0.22	-0.47
	R^2	0.09	0.09	0.04
	Num. obs.	1482	482	1000
	Censored	621	163	458
	Observed	861	319	542

Standard errors (in parentheses) computed according to Greene (1981). $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$

References in Internet Appendix

Dugast, J., & Foucault, T. (2021). Equilibrium data mining and data abundance. *HEC Paris Research Paper No. FIN-2020-1393, Université Paris-Dauphine Research Paper*, *3710495*.

Greene, W. H. (1981). Sample selection bias as a specification error: A comment. *Econometrica*, 795–798.

Heckman, J. J. (1976). The common structure of statistical models of truncation, sample selection and limited dependent variables and a simple estimator for such models. In *Annals of economic and social measurement* (pp. 475–492, Vol. 5). NBER.

Linnainmaa, J. T. (2011). Why do (some) households trade so much? *Review of Financial Studies*, *24*(5), 1630–1666.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., Blondel, M., Prettenhofer, P., Weiss, R., Dubourg, V., Vanderplas, J., Passos, A., Cournapeau, D., Brucher, M., Perrot, M., & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. *Journal of Machine Learning Research*, *12*, 2825–2830.