Machine Traders, Human Behavior, and Model (Mis)Specification

Ahmed Guecioueur*

March 7, 2023

Abstract

I examine how investors utilize data, exploiting a setting in which investors design machine-driven trading strategies under controlled yet realistic conditions. Investors disagree considerably in how they interpret identical information, leading to widely dispersed trading strategies and performance outcomes. Inexperienced investors underweight variables with predictive power for returns, and instead exhibit a bias towards variables with which they are more familiar. With experience, investors learn to overcome their bias, and benefit substantially from additional data availability. Investors' familiarity bias leads them to mis-specify their models of the world, and is encoded by the machine traders they design.

Keywords: Experience, Familiarity Bias, Predictive Models, Machine Learning, Big Data
JEL Codes: G11, G14, G41

*INSEAD, Boulevard de Constance, 77300 Fontainebleau, France, email: ahmed.guecioueur@insead.edu
I would like to thank participants at the 2022 SFS Cavalcade North America, the 2022 Chicago Booth Machine Learning in Economics Summer Institute, the 2022 Research in Behavioral Finance Conference Poster Session, the 10th Wharton-INSEAD Doctoral Consortium, the 2nd Nova SBE Finance PhD Pitch Perfect, the 2022 AFA PhD Poster Session, the 2021 FMA Annual Meeting, the 2021 EEA-ESEM Congress, the 2nd UT Austin PhD Symposium, the 2021 EFMA Conference and the 4th Dauphine Finance PhD Workshop, as well as Stefano Actis (Quantiacs), Frederico Belo, Olivier Dessaint, Jérôme Dugast (discussant), Bernard Dumas, Gleb Gertsman (discussant), Francisco Gomes, Naveen Gondhi, Georgina Hall, Xing Huang (discussant), Yukun Liu, Joël Peress (my advisor), Rodolfo Prieto, Alexandra Roulet, Amit Seru, Taisiya Sikorskaya, Ke Tang (discussant), Qingyuan Yang (discussant) and Bart Zhou Yueshen for their helpful comments. I gratefully acknowledge the support of the Henry Grunfeld Foundation. A previous version of this paper was circulated under the title "How Do Investors Learn as Data Becomes Bigger? Evidence From a Fintech Platform."

1 Introduction

Machine-driven funds account for a major fraction of both overall institutional trading volume and of hedge fund assets under management.[1] As data becomes more abundant, investors may find themselves relying even more heavily on machines to make trading decisions. Will they benefit? A growing literature finds that machines can detect asset return predictability (Freyberger, Neuhierl, and Weber, 2020; Gu, Kelly, and Xiu, 2020), improve upon human forecasts (Bianchi, Ludvigson, and Ma, 2022; van Binsbergen, Han, and Lopez-Lira, 2022; de Silva and Thesmar, 2021; Cao, Jiang, Wang, and Yang, 2021), and deliver advice that combats some human biases (D'Acunto, Prabhala, and Rossi, 2019). However, every machine must be designed by a human, and we humans can be led astray by our cognitive heuristics and biases (Barberis and Thaler, 2003; Tversky and Kahneman, 1974). It is quite possible that machines might therefore encode our human biases and hold us back from attaining our goals.

This paper examines how investors (mis)specify machine-driven trading strategies, and quantifies the extent to which they benefit from data as a result. I study the performance and behavior of investors who use a trading platform to implement machine traders. Each machine-driven trading strategy forms a portfolio based on a pre-defined set of current and historical market prices and macroeconomic indicators available on the platform. Investors seek to maximize their future performance based on these variables, and therefore face a prediction problem (Martin and Nagel, 2022). The trading platform provides investors with access to machine learning algorithms that they can incorporate into their trading strategies.

My analysis yields five main empirical findings on how investors fare in such a setting. First, investors can indeed benefit substantially from implementing machine-based trading strategies, but there is considerable *performance dispersion* in out-of-sample Sharpe Ratios; this holds true even when comparing machine traders that utilize exactly the same information sets. Second, the dispersion in outcomes stems in large part from *dispersion in predictive models*, as opposed to e.g. differences in execution. Third, *human experience positively impacts outcomes*: experienced investors design machine traders that perform better out-of-sample. Similarly, model disagreement lessens among experienced investors (but remains considerable). Fourth, *inexperienced investors favor familiar signals over genuinely predictive ones*: their trading strategies react strongly to variables that are familiar to them, but only weakly to genuine predictors. Experienced investors design machine traders that do trade upon the release of genuinely predictive information. Fifth, I confirm that this *familiarity bias can prevent investors from benefit-*

[1]With a third of all institutional trading, machine-driven hedge funds trade more than pension and mutual funds combined, and more than human-managed hedge funds. Moreover, out of all assets allocated to hedge funds, approximately 30% are managed by machine-driven hedge funds. (The Economist, *"March of the machines"*, Oct 5th 2019).

ing from data altogether. Inexperienced investors favor some macro indicators without regard to their predictive power, and thus fail to benefit from macro variables at all. Simultaneously, more experienced investors who have learned how to overcome this behavioral bias benefit substantially. I obtain this finding by exploiting a near-doubling in the available number of predictive variables on the trading platform. Altogether, I find behavioral biases influence how we humans design machine traders, and the extent to which we benefit from data as a result.

To come to these findings on the interplay between humans and machines, I exploit trading and performance records from a unique setting — a FinTech platform named Quantiacs that runs futures trading contests for investors who implement machine-based trading strategies using computer code. Unlike a laboratory setting, investors have weeks to formulate a trading strategy, have access to a wide set of actual market prices and macroeconomic indicators, and are strongly motivated by substantial financial incentives.[2] They also enjoy access to sophisticated machine learning and optimization packages that run (for free) on Quantiacs' servers. As well as these realistic features, the contest setting imposes four sets of constraints on participants that are convenient from my empirical perspective. First, participants have a well-defined, common objective, as their rewards are based on maximizing their out-of-sample Sharpe Ratios. Second, participants are prevented from interfering with their machine-based trading strategies once a contest begins; they cannot conduct manual discretionary trades, nor override machine trades. Third, participants in the same contest have identical information sets: they are limited to a common, fixed set of predictive variables, with no ability to upload any private data to the platform. Finally, a number of other potential confounders of portfolio choice are implicitly controlled for, as I describe further in Section 2.1. Thanks to these conditions, investors are driven to maximize their portfolio performance by designing machine traders that use the available data to predict asset returns out-of-sample. And while contestants' code may be confidential, I am able to exploit variation in the availability of predictive variables (across contests) and their release schedules on the Quantiacs platform to investigate the contents of these black box predictive models.

I begin my empirical analysis by documenting substantial heterogeneity in investors' performance outcomes. In each of the dozen contests, the highest decile of Sharpe Ratios (SRs) that investors attain out-of-sample are positive. In ten of these contests, the top-decile performers consistently exceed 1.4 units of SR out-of-sample. I also document considerable dispersion in performance: the interquartile range in a majority of contests is at least 2 units of SR wide, and in many contests the median investor suffers a negative out-of-sample SR.

[2] These high-powered incentives are the offer of profit-sharing contracts on up to \$1 Mln of investment provided by Quantiacs to top-ranked participants. It would be prohibitively costly to provide such significant financial incentives to traders in a laboratory experiment.

Model disagreement (Cookson and Niessner, 2020) may drive the high volume of machine-initiated trading in financial markets. Comparing investors within the same trading contest conditions on investors with a common information set: I find the mean pairwise correlation between daily trading volume series is below 0.4 for each contest (far below the perfect agreement yardstick of +1). Such heterogeneity is notable, given investors share the same goal, horizon, technology and other conditions. I investigate whether this disagreement can be explained by different interpretations of the same information, rather than execution details.

My approach is to compare the reactions of machine traders to precisely the same predictive information. To circumvent the fact that contestants' (confidential) computer code cannot be observed, I focus on trading responses triggered by macroeconomic variable releases. These macro indicators are updated irregularly by the trading platform, which thus varies the information content made available over time. I measure this variation by constructing an index of the actual out-of-sample predictive power of the macro indicators that are updated: I estimate a benchmark machine learning model that successfully detects out-of-sample return predictability for the cross-section of investable assets, and use it to extract aggregate measures of the predictive information content of each macro release day. I next measure the sensitivity of each strategy's (machine-driven) daily turnover to this index of predictive power. Comparing similarities in trading activity to information sensitivity, I find that at least 34% of the heterogeneity in trading patterns overall can be directly attributed to differential interpretations of the same (macro indicator-derived) information. It is remarkable that so much of the machine-driven trading activity can be precisely linked to model disagreement, especially since this analysis excludes the impact of information conveyed through prices. For the same reason, the true extent to which model disagreement drives trading in my setting is likely to be even higher.

Next, I identify an important driver of machine trader outcomes: human experience. Complementing the literature on experience effects,[3] I find that investors design machine traders that attain higher SRs as they gain in experience; I establish the finding using panel regressions with investor fixed effects that control for unobserved heterogeneity such as investor skill. Investor experience positively impacts both in-sample (historical backtest) SRs and out-of-sample SRs. Similarly, investor experience (gradually) resolves model disagreement; nevertheless, model disagreement remains substantial in my sample: even among experienced investors, the mean correlation between trading volume series does not exceed 0.4.

Why might the design of a machine-based trading strategy be influenced by human experience? In this paper's setting, a crucial margin of adjustment for investors is in selecting and

[3]The related literature includes prior findings that experienced investors are less prone to various behavioral biases and, relatedly, that performance improves with experience. This paper contributes evidence on the role of investor experience in designing machine-driven trading strategies and in selecting predictive variables.

combining the predictive variables available on the trading platform. As inexperienced humans are more prone to behavioral biases, I hypothesize that inexperienced investors design trading strategies that sub-optimally use the predictive variables that are available to them. I test this hypothesis by again measuring trading responses to the release of identical predictive information embedded in macro variable releases. A variable that a human investor deems more important when designing her algorithm should trigger larger trades when the input variable's value is updated. Variables are updated at different times, and this variation allows me to tease apart the trades triggered by different variables. I construct two types of index. The first measures how familiar the predictive variables updated on each macro release day are to an inexperienced investor, using data on the frequency of their appearance in news articles or published books;[4] for example, indicators for Exports, Imports and the Unemployment Rate are mentioned very frequently, so investors would be very familiar with these. The second is the index of actual predictive power derived from my benchmark machine learning model.

I compare the responses of trading strategies' daily trading volumes to variations in both types of index. All investors' trading strategies exhibit a positive relation between daily turnover and the aggregate familiarity index. However, only experienced investors' machine-driven trades exhibit a highly significant and positive relation with the index of predictive information content (the relation is only weakly significant for inexperienced investors). These results suggest that inexperienced investors first reach for more familiar variables when designing trading strategies, while underweighting genuinely predictive variables, or even ignoring them entirely. With experience, investors make better use of genuinely predictive variables – even though all investors have the same opportunity to detect out-of-sample predictability at any given time. Underweighting genuinely predictive variables constitutes a behavioral bias.

I next show that investors' bias towards familiar macro variables is serious enough that they can fail to gain any benefit from access to macro indicators at all. I compare investors who have access to macroeconomic variables to investors who do not, exploiting the addition of these variables in between trading contests. Inexperienced investors do *not* obtain a net benefit from having access to macroeconomic variables, which suggests that the mis-specification of their machine-based trading strategies negates any potential gains. By contrast, more experienced investors who have access to the macroeconomic indicators outperform similarly-experienced investors who do not. This outperformance occurs out-of-sample, and its magnitude exceeds

[4]Macroeconomic indicators that are widely mentioned in books and news articles are especially likely to be more familiar to the general population, yet it is important to note that my findings are not limited to such variables. I focus on macro variables both for identification purposes (thanks to their irregular release schedule) and because it is simple to proxy for general familiarity. It is likely that other groups of predictive variables will be more or less familiar to investors. Since investors hesitate to adopt even well-established and long-running macroeconomic variables that are gifted to them due to a lack of familiarity, it is possible that this familiarity bias will be even stronger in the case of "alternative data," for example.

2 units of SR. Self-selection and competition effects do not explain this finding.

With data growing more abundant, investors may find it necessary to adopt machine-driven trading strategies. This paper's results show this is not *sufficient* to successfully exploit predictive information. Investors who perceive some variables as unfamiliar may fail to incorporate them into their trading strategies, independently of their information content. As a corollary, this friction may impede the incorporation of unfamiliar information into prices. Furthermore, inexperienced investors (who are more prone to behavioral biases overall) consistently underperform experienced investors, including when macro indicators are unavailable to both; this implies that frictions that slow human investors from benefiting from the data they have to hand are prevalent. Such frictions are likely to be stronger in less controlled settings, where investors would in addition face information processing and acquisition costs.

My empirical results uncover a behavioral bias in how investors select among predictive variables. Recent work by Martin and Nagel (2022) models investors who rationally use historical data to form a predictive model of the world. I show how to incorporate behavioral deviations from the rational baseline of an agent who uses historical data to form beliefs about the expected return of a risky asset. Cao, Han, Hirshleifer, and Zhang (2011) interpret unfamiliarity as a fear of worst-case outcomes. Likewise, I capture the investor's aversion to unfamiliar predictive variables by injecting a fear of worst-case outcomes into her prediction problem. This bias leads her to underweight unfamiliar variables, and even ignore them entirely. Furthermore, I show an equivalence between this concise model of a biased investor's prediction problem and the well-known Lasso prediction problem (Tibshirani, 1996).

Exploiting this equivalence, I conduct Lasso estimations on daily portfolio returns that produce a proxy for the extent to which individual investors ignore predictive variables because of a general familiarity bias.[5] I find that all investors ignore some subset of available variables, and the effect is strongest among inexperienced investors. Investors who gain a contest's worth of experience learn to incorporate around 25-50 additional predictive variables into their models of the world. These semi-structural estimates complement my earlier findings on investor unfamiliarity and variable underweighting, in three ways. First, unfamiliarity can drive investors to underweight variables in general, beyond macroeconomic signals. Second, the underweighting can be severe enough that investors may discard predictive variables altogether. And third, this general form of unfamiliarity bias also lessens with experience. This Lasso-based methodology can be used in other settings to measure the extent of investors' bias.

The remainder of this paper proceeds as follows. I first review the related literature. Section 2 details the institutional setting, Section 3 measures model disagreement, and Section 4 shows that investor experience plays an important role. In Section 5, I find that investors are prone

[5] This empirical procedure is similar to Mullainathan and Obermeyer (2022)'s, with a different interpretation.

to a familiarity bias in variable selection, and Section 6 shows inexperienced investors fail to benefit from data availability as a result. In Section 7, I show how to incorporate biases into an investor's prediction problem. Section 8 concludes. The Appendix contains proofs, figures and tables. An Internet Appendix contains further institutional details and supplementary analyses.

Related Literature

This study provides early empirical evidence on how investors make use of predictive models (Freyberger, Neuhierl, and Weber, 2020; Gu, Kelly, and Xiu, 2020), and highlights a concern that behavioral biases can lead them to mis-specify such predictive models. Biased algorithmic objectives are a source of algorithmic unfairness (Cowgill and Tucker, 2020), and so this study contributes to broad efforts to understand the potential risks of new predictive technologies, including in the field of finance (Fuster, Goldsmith-Pinkham, Ramadorai, and Walther, 2022).

Relatedly, I contribute a new set of insights to the literature on machine-driven trading, which has so far focussed on the commonality in machine traders' actions (Khandani and Lo, 2011; Abis, 2020), the tendency for performance to degenerate out-of-sample (McLean and Pontiff, 2016; Wiecki, Campbell, Lent, and Stauth, 2016; Falck, Rej, and Thesmar, 2022), and the risk exposures of machine traders (Harvey, Rattray, Sinclair, and Van Hemert, 2017). My results highlight the impact of human limitations on the design of machine traders.

This paper adds to recent empirical evidence that model disagreement is a key driver of trading activity (Cookson and Niessner, 2020; Meeuwis, Parker, Schoar, and Simester, 2022). Further, I identify the role of investor experience in (partially) resolving model disagreement, controlling precisely for access to the same data. Recent studies by Cookson, Engelberg, and Mullins (2023) and Faia, Fuster, Pezone, and Zafar (forthcoming) document behavioral biases in how investors acquire information; I find that behavioral biases can in addition lead investors to under-utilize genuinely predictive information that they already have to hand, independently of its salience (Frydman and Wang, 2020). The familiarity bias that I uncover afflicts agents who are equally endowed with access to a sophisticated prediction technology (including machine learning tools) and ample time to implement their trading strategies; it is thus likely that the cost of attending to variables (Aragones, Gilboa, Postlewaite, and Schmeidler, 2005; Hanna, Mullainathan, and Schwartzstein, 2014) is not a binding constraint.

A growing literature analyzes the consequences of various forms of model mis-specification, in both asset pricing (Arthur et al., 1996; Barberis, Shleifer, and Vishny, 1998; Hong, Stein, and Yu, 2007; Branch and Evans, 2010) and from an individual learning perspective (Gagnon-Bartsch, Rabin, and Schwartzstein, 2018; Bohren and Hauser, 2021; Montiel Olea, Ortoleva, Pai, and Prat, 2022). My results suggest that behavioral factors are important drivers of model mis-specification in practice, including when investors have access to a sophisticated prediction technology and a high-dimensional set of predictors. This behavioral view complements studies on how rational agents might specify predictive models (Brock and Hommes, 1997; Al-Najjar, 2009; Balasubramanian and Yang, 2021; Martin and Nagel, 2022).

Similarly, my paper contributes to a literature that studies the consequences of Big Data, and which typically focuses on the equilibrium consequences of data growth (Dugast and Foucault, 2021; Farboodi, Singal, Veldkamp, and Venkateswaran, 2022). I take an empirical approach and find that behavioral biases can prevent investors from fully exploiting an exogenous increase in the number of predictive variables made available to them; i.e. Bigger Data.

My findings that inexperienced investors omit variables from their models of the world are indicative of under-reaction. An extensive empirical literature documents under-reaction in asset prices (e.g. Bernard and Thomas, 1990; Jegadeesh and Titman, 1993) and individual expectations (e.g. Bouchaud, Krueger, Landier, and Thesmar, 2019), and I add to this evidence by showing that machine trader actions also under-react to information. Further, I find that inexperience-driven under-reaction (Mikhail, Walther, and Willis, 2003) is not necessarily due to learning from more signal realizations, but can be due to behavioral heuristics.

This paper is related to three strands of the literature on experience effects. First, prior work finds that investor experience counteracts a number of behavioral biases, including the endowment effect (List, 2003), the disposition effect (Feng and Seasholes, 2005; Dhar and Zhu, 2006; Da Costa Jr et al., 2013), home bias (Abreu, Mendes, and Santos, 2011), a tendency to chase bubbles (Greenwood and Nagel, 2009), and under-diversification (Campbell, Ramadorai, and Ranish, 2014). I find that investors benefit from experience by counteracting a bias in the manner in which they utilize predictive signals. Second, a number of finance studies examine individual trading records to conclude that human investors learn to attain higher performance levels as they gain in experience, including work by Nicolosi, Peng, and Zhu (2009), Seru, Shumway, and Stoffman (2010), Linnainmaa (2011) and Barber et al. (2020). I measure human experience precisely, and find it to be a key driver of performance outcomes even when the day-to-day trading decisions are delegated to machines designed by the human investors.[6] A third stream of literature considers how individuals' lifetime experiences influence their current beliefs (Malmendier and Nagel, 2011; Malmendier, Pouzo, and Vanasco, 2020). Complementing these findings, I show that investor experience impacts their models of the world even when fixing a common dataset, and that they may ignore part of this dataset.

My study joins a recent literature that uses machine learning techniques to understand individuals' actions. In part of my analysis, I set up a benchmark machine learning model for

[6] This is consistent with recent evidence by Chen, Hshieh, Teo, and Zhang (2022) that human capital can have a positive impact on the performance of systematic hedge funds.

comparison, which parallels recent work by Bianchi, Ludvigson, and Ma (2022), van Binsbergen, Han, and Lopez-Lira (2022), de Silva and Thesmar (2021), and Cao, Jiang, Wang, and Yang (2021), all of whom compare human predictions to machine predictions. I use machine predictions to infer the informativeness of individual predictive variables. I then construct an index of the informativeness of variable updates provided by the trading platform, which I compare to machine trading activity to infer how machine trading models are (mis)specified.

This paper documents that investors are quick to incorporate familiar variables in their models of the world, and slow to incorporate genuine predictive variables. This tendency towards the familiar constitutes a bias,[7] and may stem from employing two closely-related psychological heuristics. The *availability heuristic* suggests human investors may judge the strength of out-of-sample predictive power based on mental association: since more familiar variables are more likely to come to mind (i.e. be more "available") to human investors when attempting to specify a set of valid predictive variables, this may lead them to infer an "illusory correlation" (Tversky and Kahneman, 1973, 1974) between more familiar macroeconomic variables and asset returns out-of-sample. Alternatively, human investors may be following a simple *recognition heuristic* (Goldstein and Gigerenzer, 1999, 2002) when choosing among variables.[8]

Since investors in my setting use a historical dataset to attain their objectives, a natural approach is to model agents as econometricians, borrowing an analogy from Sargent (1993, pp. 21-23). Martin and Nagel (2022) consider a risk-neutral agent who learns a model to predict asset payoffs based on historical data; I similarly model an agent who learns how to predict expected returns to attain her goal, this time incorporating the investor's familiarity bias as an aversion to worst-case error (Cao, Han, Hirshleifer, and Zhang, 2011). To do this, I lean on results at the intersection of the machine learning & optimization literatures (Xu, Caramanis, and Mannor, 2010; Tibshirani, 1996). Gabaix (2014, 2019) and Molavi, Tahbaz-Salehi, and Vedolin (2021) model bounded rationality using notions of sparsity; my framework also allows investors to ignore a subset of their environment, but stems from a different microfoundation. My approach also relates to, but is distinct from, models of ambiguity aversion (Epstein and Schneider, 2008; Illeditsch, 2011; Garlappi, Uppal, and Wang, 2007).

Finally, I join other studies that tackle economic questions using data sourced from FinTech settings, such as crowd-sourcing (Da, Huang, and Jin, 2021) or crowdfunding (Tang, 2019) platforms. My institutional setting emulates realistic trading conditions.

[7]A literature on geographic home bias debates whether familiarity in portfolio choice is irrational or due to rational information acquisition (Huberman, 2001; Grinblatt and Keloharju, 2001; Massa and Simonov, 2006; Goetzmann and Kumar, 2008; Van Nieuwerburgh and Veldkamp, 2009). Investors in my setting fail to benefit from predictive variables until gaining experience; this negative effect on performance indicates a behavioral bias.

[8]One psychological mechanism relates to making mental judgments of association (i.e. predictability, in this setting), and the other relates to making choices. Typically (and in my setting), an investor's choice among predictive variables and her judgment about the relative strength of predictability are observationally equivalent.

2 Institutional Setting and Data

This section describes the Quantiacs institutional setting. Among other features, this unique setting allows me to measure two key aspects of the datasets available to trading strategies: data availability (the presence or absence of predictive variables) and data releases (the addition of a new realization for a variable). I can thus infer how machine-driven trading strategies make use of data. Furthermore, I can precisely measure performance and investors' experience levels to study the interplay between human and machine.

2.1 The Quantiacs Trading Platform and Trading Contests

Systematic trading[9] is the act of implementing an algorithm that takes positions in various financial assets based upon a trading strategy that its (human) designers have specified at the outset. It has traditionally been associated with statistical arbitrage hedge funds (who buy and short portfolios of stocks) and so-called "Commodity Trading Advisors" (who trade futures contracts and other derivatives).

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.[10] 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.[11] 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

[9] I use the terms "systematic trading" and "systematic investing" interchangeably. Another equivalent term is "quantitative" trading or investing. I avoid the terms "algorithmic" or "automated" trading because they are ambiguous, and may refer instead to the algorithmic execution of human-initiated trades. In this study, by contrast, all trades are initiated by machines, and the role of humans is to design these machines.

[10] 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.

[11] 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.

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 1 illustrates the distinction between the in-sample/backtest and out-of-sample live periods of a contest.

[Insert Figure 1 around 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.

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 ensure the relevance of this study's findings.

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.[12]

[12]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.

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 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 2.2.4 and 2.3 provide further details, and additionally describe the (irregular) release schedules of different macroeconomic indicators.

2.2 Data

2.2.1 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.

2.2.2 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.

2.2.3 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 and weights are not available, I can observe two daily positioning indices per trading strategy, one for its aggregate long positioning and the other for its short positioning: for trading strategy j and day τ, I observe the sequences $\{Long_{j,\tau}\}$ and $\{Short_{j,\tau}\}$, respectively. The daily sum of these indices is bounded between zero and one, $0 \leq \max_{\tau}(Long_{j,\tau} + Short_{j,\tau}) \leq 1,\ \forall\ i$, and these may be interpreted as portfolio shares of each strategy's assets under management allocated to aggregate long and short positions. I use these measures to proxy for daily trading volume (or turnover) as

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

2.2.4 Macroeconomic Variables 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 variables. 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 2 around here]

Figure 2 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.

2.3 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 3 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 3 around 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 Model and Performance Heterogeneity

3.1 Differences in Performance

Table 1 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 1 around 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 4 around here]

Figure 4 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.

3.2 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.

3.2.1 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 a suitable benchmark to compare human behavior to (van Binsbergen, Han, and Lopez-Lira, 2022). 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:

$$r_{i,\tau+1} = \mathbb{E}_t[r_{i,\tau+1}] + \epsilon_{i,\tau+1}, \tag{2}$$

with the conditional prediction for each next-month return

$$\mathbb{E}_\tau[r_{i,\tau+1}] = f(\boldsymbol{x}_\tau) \tag{3}$$

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;[13] 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,[14] 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.

[13] 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.

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

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 5 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).[15]

[Insert Figure 5 around 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).[16] 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

[15] 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 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.

[16] "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.

by summing up the SHAP values of the variables released that day:

$$SHAP(\tau, m) = \sum_{k \in \mathcal{R}(t)} SHAP(k, m). \tag{4}$$

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

$$\text{Informativeness}_\tau = \frac{1}{M} \sum_m \left| SHAP(\tau, m) \right|. \tag{5}$$

Equation (5) 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.

3.2.2 Explaining Trading Heterogeneity with Model Disagreement

Armed with my index (5) of the predictive informational content of macro release days, I measure the sensitivity of different trading strategies to identical predictive signals. Using Equation (1), 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:

$$\text{Volume}_{i,\tau} = \alpha_i + \beta_i \times \text{Information}_\tau + \epsilon_{i,\tau} \tag{6}$$

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

$$\underbrace{-\rho_{i,j}}_{\substack{\text{Trading} \\ \text{dissimilarity}}} = \gamma \times \underbrace{\delta(\widehat{\beta}_i, \widehat{\beta}_j)}_{\substack{\text{Disagreement} \\ \text{about macro} \\ \text{predictive info.}}} + \phi_t + \epsilon_{i,j}, \tag{7}$$

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 (6).

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 2 around here]

Table 2 presents regression results for each of the variants of Eqn. (7). 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.

3.2.3 In Search of a Behavioral Explanation

From the standpoint of a rational Bayesian updating benchmark, the extent to which investors disagree in their interpretations of the same signals is puzzling. Assume each investor places some weight on each predictive variable when forming her beliefs and, since she is a Bayesian, assigns a prior to each weight, which may differ across investors. The group of investors who enter the 8th trading contest (the first in which macro indicators were available) did so in 2017, and have thus observed monthly macro indicator realizations over the previous 27-year backtest period. If investors update their posterior weightings after observing each of the (over 300) realizations, one might expect their models of the world to have converged more closely to one another by that point. Instead, the differences of opinion that I document remain substantial.

Furthermore, the need to produce out-of-sample forecasts is not a satisfactory explanation for the extent of model disagreement. When specifying a predictive model of the world, investors possess sufficient historical data to retain a hold-out sample to assess their model's out-of-sample performance. This procedure is common in the empirical literature, including in recent work (Freyberger, Neuhierl, and Weber, 2020; Gu, Kelly, and Xiu, 2020) and indeed, in the present study (Section 3.2.1). Investors are capable of following a similar procedure in the Quantiacs setting, and thus have an (equal) opportunity to detect out-of-sample predictability. Nevertheless, they disagree about how to specify predictive models.

The remainder of the paper studies whether and how investor behavior influences the models of the world that they specify.

4 Role of Experience

Motivated by the well-accepted link between behavioral biases and investor (in)experience (e.g. List, 2003), I investigate whether experience plays a role in the present setting. If inexperienced investors are indeed prone to behavioral biases when specifying their models of the world, their performance is likely to suffer as a result. Furthermore, more experienced investors who succeed in specifying a more accurate model of the world should agree more, and thus trade more similarly to one another. This section tests these hypotheses.

4.1 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

$$\mathrm{SR}_{i,t}^{Best} = \beta \times \text{Contests experienced}_{i,t} + \pi_i + \phi_t + \epsilon_{i,t}, \tag{8}$$

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 3 around here]

Table 3 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.[17] 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 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).

4.2 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 4). 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:

$$\rho_{i,j,t} = \beta \times \text{Contests experienced}_{i,j,t} + \pi_i + \phi_t + \epsilon_{i,j,t}. \tag{9}$$

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 4 around here]

Table 4 displays the regression results. The simplest specification in column (1) produces an easily-interpretable measure of the mean pairwise correlation between the (machine-initiated)

[17]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.

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.

5 Data Usage

In this study's institutional setting, the role that human investors play is limited to designing machine traders.[18] 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.[19] 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.

[18]As described in Section 2.1 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).

[19]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.

5.1 Usage of Genuine Predictive Variables

Given my measure of the information content of macro release days (5), I now analyze the responses of machine trading strategies to this predictive information content. Following Eqn. (1), 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:

$$\begin{aligned} \text{Volume}_{i,\tau} = \beta_1 \times \text{Experienced}_{i,t} + \beta_2 \times \text{Informativeness}_{\tau} \\ + \beta_3 \times \text{Experienced}_{i,t} \times \text{Informativeness}_{\tau} + \phi_t + \epsilon_{i,t,\tau} \end{aligned} \tag{10}$$

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. (5). 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 5 around here]

Table 5 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).

5.2 Usage of Familiar Variables

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.

5.2.1 Measuring Variable 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 frequency 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 total familiarity of each macroeconomic indicator k among the set of macroeconomic indicators $\mathcal{R}(\tau)$ released together on date τ by the Quantiacs platform:

$$\text{Familiarity}_{\tau} = \frac{1}{\text{Total mentions}} \sum_{k \in \mathcal{R}(t)} \text{Mentions}(k). \tag{11}$$

Equation (11) measures the total share of media mentions of the macroeconomic variables 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 (11) 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.

5.2.2 All Investors Rely on Familiar Macro Variables

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 5.1. 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:

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

When calculating familiarity indices using Equation (11), 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 6 around here]

Table 6 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 5) and familiarity regressions (Table 6). Neither index drives out the other, and results are similar across both familiarity measures.

6 Gains from Access to (Bigger) Data

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 predictive 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 Predictive Variables

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 7 around 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 7 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 8 around here]

The regressions in Table 8 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 8 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.[20]

The regression results in Panel A of Table 8 are variants of the following linear specification:

$$\begin{aligned}
&\text{Excess Live SR}_{i,t}^{Best} \\
&\quad = \beta_1 + \beta_2 \times \text{Contests experienced}_{i,t} + \beta_3 \times \text{New variables available}_t \\
&\qquad + \beta_4 \times \text{Contests experienced}_{i,t} \times \text{New variables available}_t + \epsilon_{i,t}.
\end{aligned} \tag{13}$$

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 8 are variants of the following specification:

$$\begin{aligned}
&\text{Excess Live SR}_{i,t}^{Best} \\
&\quad = \beta_1 + \sum_{k=2}^{4} \left[\beta_k \times \mathbb{1}\{\text{Contests experienced}_{i,t} = k\} \right] + \gamma_1 \times \text{New variables available}_t \\
&\qquad + \sum_{k=2}^{4} \left[\gamma_k \times \mathbb{1}\{\text{Contests experienced}_{i,t} = k\} \times \text{New variables available}_t \right] + \epsilon_{i,t}.
\end{aligned} \tag{14}$$

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 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

[20]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.

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 9 around 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"[21] 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 9, and the estimates indicate that contestants are less likely to participate in the next contest as they gain in experience and as 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.

[21] 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 second-stage coefficients represent the outcome equation of interest, and these are shown in the middle section of Table 9. 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 8 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.[22] 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.

7 Investor Behavior When Solving a Prediction Problem

This section introduces a model of investor behavior that allows me to measure investors' usage of individual predictive variables. 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 variables that are available to them. I also estimate this behavioral model.

[22]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.

7.1 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

$$\max_{w} \frac{\boldsymbol{\mu}^T \boldsymbol{w}}{\sqrt{\boldsymbol{w}^T \boldsymbol{\Sigma} \boldsymbol{w}}}, \tag{15}$$

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).[23] 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.[24] Investors in the Quantiacs setting behave consistently with this assumption: the results in Internet Appendix 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,

$$\mu = \sum_{i=1}^{m} b_i s_i = \boldsymbol{s}\,\boldsymbol{b}, \tag{16}$$

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

[23]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.

[24]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.

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 $\text{Err} : \mathbb{R} \times \mathbb{R} \to [0, +\infty)$ 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 $\boldsymbol{S}$:

$$\min_{\boldsymbol{b} \in \mathbb{R}^m} \text{Err}(\boldsymbol{v}, \boldsymbol{S}\boldsymbol{b}) \tag{17}$$

Minimizing the errors in the predictions of these historical realized returns (17) 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 (17) 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*

$$\min_{\boldsymbol{b} \in \mathbb{R}^m} ||\boldsymbol{v} - \boldsymbol{S}\boldsymbol{b}||_2. \tag{18}$$

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. (16).

7.2 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 (16), 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,*

$$\min_{\boldsymbol{b}\in\mathbb{R}^m} \max_{\boldsymbol{U}\in\mathcal{U}} ||\boldsymbol{v}-(\boldsymbol{S}+\boldsymbol{U})\boldsymbol{b}||_2, \tag{19}$$

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

$$\mathcal{U} := \left\{[\boldsymbol{u}_1\ \boldsymbol{u}_2\ \dots\ \boldsymbol{u}_m] : ||\boldsymbol{u}_i||_2 \leq \delta_i\ \forall\ i=1,\dots,m\right\} \tag{20}$$

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. 16). 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 (19) 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* (19) *by solving an equivalent formulation*

$$\min_{\boldsymbol{b} \in \mathbb{R}^m} \frac{1}{2} ||\boldsymbol{v} - \boldsymbol{S}\boldsymbol{b}||_2^2 + \lambda ||\boldsymbol{b}||_1, \tag{21}$$

where $\lambda \geq 0$ *is a scaling of* δ *in* (20).

The Lasso prediction problem (21) 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* (21) *by predicting*

$$\widehat{\mu} = \boldsymbol{s}\widehat{\boldsymbol{b}}, \tag{22}$$

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*

$$\widehat{b}_k = \text{sign}(\boldsymbol{s}_k^T \boldsymbol{v}) \max\left\{|\boldsymbol{s}_k^T \boldsymbol{v}| - \lambda, 0\right\}. \tag{23}$$

Figure 6 illustrates the functional form given by Equation (23). Over the region $|\boldsymbol{s}_k^T \boldsymbol{v}| \leq \lambda$ 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 $\boldsymbol{s}_k^T \boldsymbol{v}$ and so will still be treated differently by the investor.

[Insert Figure 6 around here]

7.3 Consequences of Investor Bias

Under-reaction & model mis-specification One interpretation of the biased investor's model of the world (23) is to note that $\boldsymbol{s}_k^T \boldsymbol{v}$ 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 6. 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.[25] According to my empirical findings in Section 6 and in my next analysis, Bigger Data does not 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.

7.4 Empirically Measuring the Extent of Investor Bias

I now take the predictive model (21) 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 5. 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 with experience, as hypothesized earlier.

7.4.1 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 (21) 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. (21).

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.

[25] 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.

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 $\boldsymbol{v}$,[26] 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,[27] I report the estimated number of predictive variables that each investor entry uses; that is, the number of non-zero elements in the corresponding vector of parameter solutions $\widehat{\boldsymbol{b}}^*$.

7.4.2 Results and Discussion

Figure 7 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 7 around 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 5).

[26] 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.

[27] 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.

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 7 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

Using a unique panel of systematic investor outcomes and daily turnover from an institutional setting that controls investors' preferences, horizons and (crucially) the information they can use to make trading decisions, I study how investors specify machine-driven trading strategies that make use of predictive variables. The controlled conditions allow me to interpret these algorithms as models of the world defined by investors.

I find that investors perform better with experience, and (gradually) alleviate their model disagreement. Delving into investors' usage of macroeconomic predictive variables, I find that inexperienced investors appear to underweight genuine predictive variables, in favor of variables that they perceive to be more familiar. This bias holds inexperienced investors back from benefiting from Bigger Data.

I also contribute an empirical framework that can be used to measure investors' usage of predictive variables. I find that investors of all experience levels ignore a subset of their information environments, but learn to incorporate additional signals into their models of the world with experience.

My findings shed light on the role of human limitations in the design of machine traders. The familiarity bias that investors exhibit in their use of available information is likely to be prevalent in other settings, and is a driver of model disagreement.

References

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

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.

Al-Najjar, N. I. (2009). Decision makers as statisticians: Diversity, ambiguity, and learning. *Econometrica*, *77*(5), 1371–1401.

Aragones, E., Gilboa, I., Postlewaite, A., & Schmeidler, D. (2005). Fact-free learning. *American Economic Review*, *95*(5), 1355–1368.

Arthur, W. B., Holland, J. H., LeBaron, B., Palmer, R., & Tayler, P. (1996). Asset pricing under endogenous expectations in an artificial stock market. *The economy as an evolving complex system II*, *27*.

Balasubramanian, A., & Yang, D. Y. (2021). Trading with high-dimensional data. *Available at SSRN*, *3583217*.

Barber, B. M., Lee, Y.-T., Liu, Y.-J., Odean, T., & Zhang, K. (2020). Learning, fast or slow. *Review of Asset Pricing Studies*, *10*(1), 61–93.

Barberis, N., Shleifer, A., & Vishny, R. (1998). A model of investor sentiment. *Journal of Financial Economics*, *49*(3), 307–343.

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

Bernard, V. L., & Thomas, J. K. (1990). Evidence that stock prices do not fully reflect the implications of current earnings for future earnings. *Journal of Accounting and Economics*, *13*(4), 305–340.

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.

Bianchi, F., Ludvigson, S. C., & Ma, S. (2022). Belief distortions and macroeconomic fluctuations. *American Economic Review*, *112*(7), 2269–2315.

Bohren, J. A., & Hauser, D. N. (2021). Learning with heterogeneous misspecified models: Characterization and robustness. *Econometrica*, *89*(6), 3025–3077.

Bouchaud, J.-P., Krueger, P., Landier, A., & Thesmar, D. (2019). Sticky expectations and the profitability anomaly. *Journal of Finance*, *74*(2), 639–674.

Branch, W. A., & Evans, G. W. (2010). Asset return dynamics and learning. *Review of Financial Studies*, *23*(4), 1651–1680.

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

Brock, W. A., & Hommes, C. H. (1997). A rational route to randomness. *Econometrica*, *65*(5), 1059–1095.

Campbell, J. Y., Ramadorai, T., & Ranish, B. (2014). Getting better or feeling better? How equity investors respond to investment experience. *NBER Working Paper*, *20000*.

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.

Cao, S., Jiang, W., Wang, J. L., & Yang, B. (2021). From man vs. machine to man + machine: The art and AI of stock analyses. *NBER Working Paper*, *28800*.

Chen, J., Hshieh, S., Teo, M., & Zhang, F. (2022). Foreign talent and hedge funds. *Available at SSRN*, *4025639*.

Cookson, J. A., Engelberg, J. E., & Mullins, W. (2023). Echo chambers. *Review of Financial Studies*, *36*(2), 450–500.

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.

Cowgill, B., & Tucker, C. E. (2020). Algorithmic fairness and economics. *Columbia Business School Research Paper*.

Da, Z., Huang, X., & Jin, L. J. (2021). Extrapolative beliefs in the cross-section: What can we learn from the crowds? *Journal of Financial Economics*, *140*(1), 175–196.

Da Costa Jr, N., Goulart, M., Cupertino, C., Macedo Jr, J., & Da Silva, S. (2013). The disposition effect and investor experience. *Journal of Banking & Finance*, *37*(5), 1669–1675.

D'Acunto, F., Prabhala, N., & Rossi, A. G. (2019). The promises and pitfalls of robo-advising. *Review of Financial Studies*, *32*(5), 1983–2020.

de Silva, T., & Thesmar, D. (2021). Noise in expectations: Evidence from analyst forecasts. *NBER Working Paper*, *28963*.

Dhar, R., & Zhu, N. (2006). Up close and personal: Investor sophistication and the disposition effect. *Management Science*, *52*(5), 726–740.

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*.

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

Faia, E., Fuster, A., Pezone, V., & Zafar, B. (forthcoming). Biases in information selection and processing: Survey evidence from the pandemic. *Review of Economics and Statistics*.

Falck, A., Rej, A., & Thesmar, D. (2022). When do systematic strategies decay? *Quantitative Finance*, *22*(11), 1955–1969.

Farboodi, M., Singal, D., Veldkamp, L., & Venkateswaran, V. (2022). Valuing financial data. *NBER Working Paper*, *29894*.

Feng, L., & Seasholes, M. S. (2005). Do investor sophistication and trading experience eliminate behavioral biases in financial markets? *Review of Finance*, *9*(3), 305–351.

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.

Frydman, C., & Wang, B. (2020). The impact of salience on investor behavior: Evidence from a natural experiment. *Journal of Finance*, *75*(1), 229–276.

Fuster, A., Goldsmith-Pinkham, P., Ramadorai, T., & Walther, A. (2022). Predictably unequal? The effects of machine learning on credit markets. *Journal of Finance*, *77*(1), 5–47.

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

Gabaix, X. (2019). Behavioral inattention. In *Handbook of Behavioral Economics: Applications and Foundations* (pp. 261–343). Elsevier.

Gagnon-Bartsch, T., Rabin, M., & Schwartzstein, J. (2018). Channeled attention and stable errors. *Harvard Business School Working Paper*, *18-108*.

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.

Goetzmann, W. N., & Kumar, A. (2008). Equity portfolio diversification. *Review of Finance*, *12*(3), 433–463.

Goldstein, D. G., & Gigerenzer, G. (1999). The recognition heuristic: How ignorance makes us smart. In G. Gigerenzer & P. M. Todd (Eds.), *Simple heuristics that make us smart* (pp. 37–58). Oxford University Press.

Goldstein, D. G., & Gigerenzer, G. (2002). Models of ecological rationality: The recognition heuristic. *Psychological Review*, *109*(1), 75.

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

Greenwood, R., & Nagel, S. (2009). Inexperienced investors and bubbles. *Journal of Financial Economics*, *93*(2), 239–258.

Grinblatt, M., & Keloharju, M. (2001). How distance, language, and culture influence stockholdings and trades. *Journal of Finance*, *56*(3), 1053–1073.

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., Rattray, S., Sinclair, A., & Van Hemert, O. (2017). Man vs. machine: Comparing discretionary and systematic hedge fund performance. *Journal of Portfolio Management*, *43*(4), 55–69.

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). NBER.

Hong, H., Stein, J. C., & Yu, J. (2007). Simple forecasts and paradigm shifts. *Journal of Finance*, *62*(3), 1207–1242.

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.

Jegadeesh, N., & Titman, S. (1993). Returns to buying winners and selling losers: Implications for stock market efficiency. *Journal of Finance*, *48*(1), 65–91.

Khandani, A. E., & Lo, A. W. (2011). What happened to the quants in August 2007? Evidence from factors and transactions data. *Journal of Financial Markets*, *14*(1), 1–46.

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.

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

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*.

Malmendier, U., & Nagel, S. (2011). Depression babies: Do macroeconomic experiences affect risk taking? *Quarterly Journal of Economics*, *126*(1), 373–416.

Malmendier, U., Pouzo, D., & Vanasco, V. (2020). Investor experiences and financial market dynamics. *Journal of Financial Economics*, *136*(3), 597–622.

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.

Massa, M., & Simonov, A. (2006). Hedging, familiarity and portfolio choice. *Review of Financial Studies*, *19*(2), 633–685.

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

Meeuwis, M., Parker, J. A., Schoar, A., & Simester, D. (2022). Belief disagreement and portfolio choice. *Journal of Finance*, *77*(6), 3191–3247.

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

Mikhail, M. B., Walther, B. R., & Willis, R. H. (2003). The effect of experience on security analyst underreaction. *Journal of Accounting and Economics*, *35*(1), 101–116.

Molavi, P., Tahbaz-Salehi, A., & Vedolin, A. (2021). Model complexity, expectations, and asset prices. *NBER Working Paper*, *28408*.

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

Mullainathan, S., & Obermeyer, Z. (2022). Diagnosing physician error: A machine learning approach to low-value health care. *Quarterly Journal of Economics*, *137*(2), 679–727.

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

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.

Sargent, T. J. (1993). *Bounded Rationality in Macroeconomics*. Oxford University Press.

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

Tang, H. (2019). Peer-to-peer lenders versus banks: Substitutes or complements? *Review of Financial Studies*, *32*(5), 1900–1938.

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.

Tversky, A., & Kahneman, D. (1973). Availability: A heuristic for judging frequency and probability. *Cognitive Psychology*, *5*(2), 207–232.

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases: Biases in judgments reveal some heuristics of thinking under uncertainty. *Science*, *185*(4157), 1124–1131.

Van Nieuwerburgh, S., & Veldkamp, L. (2009). Information immobility and the home bias puzzle. *Journal of Finance*, *64*(3), 1187–1215.

van Binsbergen, J. H., Han, X., & Lopez-Lira, A. (2022). Man versus machine learning: The term structure of earnings expectations and conditional biases. *Review of Financial Studies*.

Wiecki, T., Campbell, A., Lent, J., & Stauth, J. (2016). All that glitters is not gold: Comparing backtest and out-of-sample performance on a large cohort of trading algorithms. *Journal of Investing*, *25*(3), 69–80.

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*

$$\min_{\boldsymbol{b}\in\mathbb{R}^m} \max_{\boldsymbol{U}\in\mathcal{U}} ||\boldsymbol{v} - (\boldsymbol{S}+\boldsymbol{U})\boldsymbol{b}||_2 \tag{24}$$

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

$$\mathcal{U} := \left\{[\boldsymbol{u}_1\ \boldsymbol{u}_2\ \dots\ \boldsymbol{u}_m] : ||\boldsymbol{u}_i||_2 \le \delta_i\ \forall\ i = 1,\dots,m\right\} \tag{25}$$

is equivalent to the optimization problem

$$\min_{\boldsymbol{b}\in\mathbb{R}^m} ||\boldsymbol{v} - \boldsymbol{S}\boldsymbol{b}||_2 + \sum_{i=1}^{m} \delta_i |b_i|. \tag{26}$$

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

Lemma 2. *The optimization problem*

$$\min_{\boldsymbol{b}\in\mathbb{R}^m} ||\boldsymbol{v} - \boldsymbol{S}\boldsymbol{b}||_2 + \delta||\boldsymbol{b}||_1 \tag{27}$$

is equivalent to the optimization problem

$$\min_{\boldsymbol{b}\in\mathbb{R}^m} \frac{1}{2}||\boldsymbol{v} - \boldsymbol{S}\boldsymbol{b}||_2^2 + \lambda||\boldsymbol{b}||_1 \tag{28}$$

with the one-to-one change of parameters

$$\lambda = \frac{\delta a_1}{\sqrt{1-\delta^2 a_2}} \quad \Leftrightarrow \quad \delta = \frac{\lambda}{\sqrt{\lambda^2 a_3 + a_4}}, \tag{29}$$

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* (19) *by solving an equivalent formulation*

$$\min_{\boldsymbol{b}\in\mathbb{R}^m} \frac{1}{2}||\boldsymbol{v}-\boldsymbol{S}\boldsymbol{b}||_2^2+\lambda||\boldsymbol{b}||_1, \tag{21}$$

where $\lambda \geq 0$ *is a scaling of* δ *in* (20).

Proof. Applying Lemma 1 to the optimization problem in Definition 2 with $\delta_i = \delta\ \forall i$ shows that (19) is equivalent to the following optimization problem:

$$\min_{\boldsymbol{b}\in\mathbb{R}^m} ||\boldsymbol{v}-\boldsymbol{S}\boldsymbol{b}||_2+\delta||\boldsymbol{b}||_1. \tag{30}$$

By Lemma 2, (30) is equivalent to (21), with the mapping between λ, δ given by the lemma. The optimization problems (19) and (21) 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* (21) *by predicting*

$$\widehat{\mu} = \boldsymbol{s}\widehat{\boldsymbol{b}}, \tag{22}$$

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*

$$\widehat{b}_k = \operatorname{sign}(\boldsymbol{s}_k^T\boldsymbol{v})\max\left\{|\boldsymbol{s}_k^T\boldsymbol{v}|-\lambda, 0\right\}. \tag{23}$$

Proof. I start from the formulation (21). 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 $\partial(||\boldsymbol{b}||_1)$:

$$\boldsymbol{0} \ \in\ \boldsymbol{S}^T(\boldsymbol{S}\widehat{\boldsymbol{b}}-\boldsymbol{v})+\lambda\partial(||\boldsymbol{b}||_1) \tag{31}$$

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

$$\boldsymbol{0} \ \in\ \widehat{\boldsymbol{b}}-\boldsymbol{S}^T\boldsymbol{v}+\lambda\partial(||\boldsymbol{b}||_1). \tag{32}$$

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

$$0 \ =\ \widehat{b}_k-\boldsymbol{s}_k^T\boldsymbol{v}+\lambda\widehat{z}_k, \tag{33}$$

where $\widehat{z}_k = \text{sign}(\widehat{b}_k)$ when $\widehat{b}_k \neq 0$ and the set $[-1, +1]$ otherwise. The result (23) then follows. $\square$

Figures

Figure 1: 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.

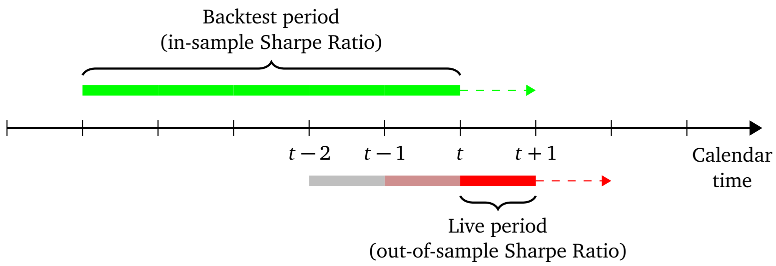


Figure 2: Stylized distinction between macro variable release days and macro variable availability. Contestants take part in one or more trading contests t by entering trading strategies that run over an extended calendar period. The historical periods of all contests overlap in calendar time, so each 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).

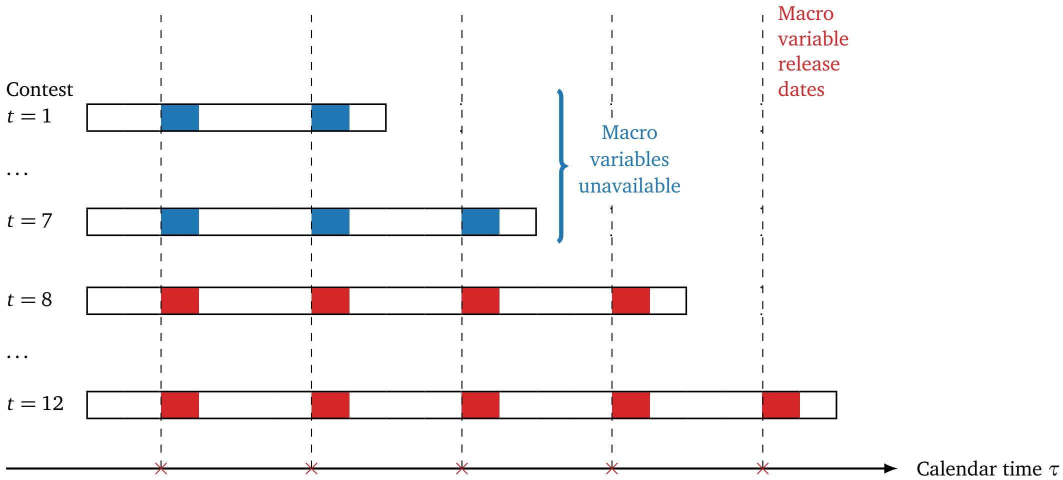


Figure 3: Variation in the number of variables released on macro release dates. This chart counts the number of macro variables that were released to trading strategies on the Quantiacs platform, including both in-sample/Backtest and out-of-sample/Live periods. Variables were released in monthly batches, 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."

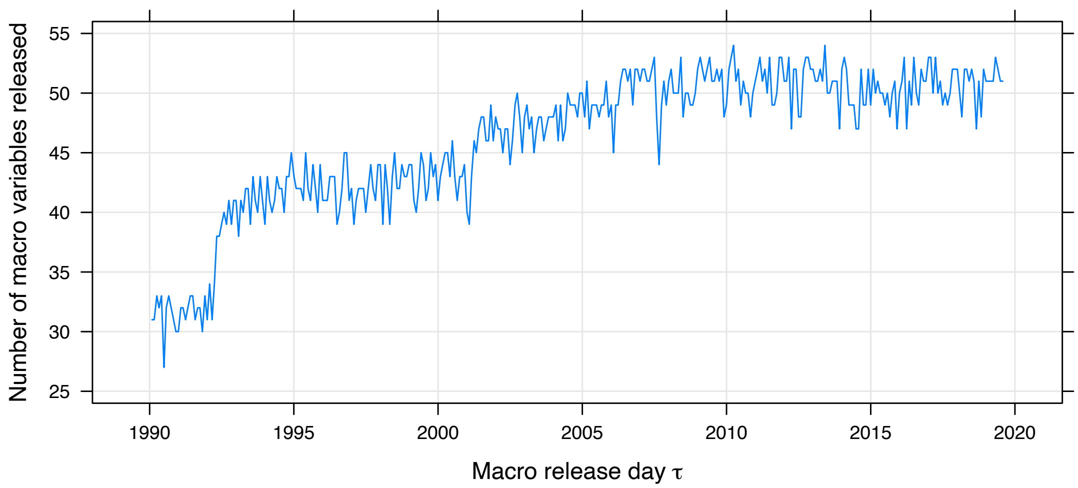


Figure 4: Cross-sectional mean correlation in trades. For each contest index t, this chart displays the cross-sectional mean of the pairwise correlations $\rho_{i,j,t}$ in daily turnovers for all best live entries in the contest. Self-correlations are excluded.

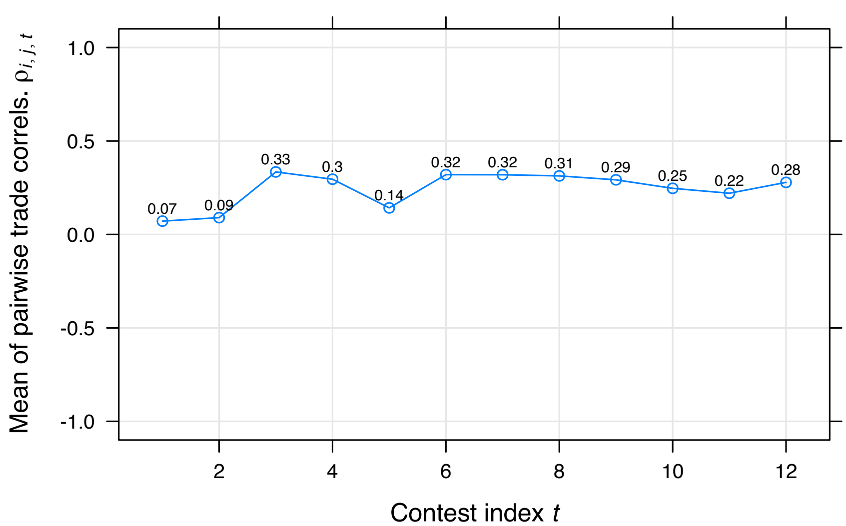


Figure 5: Out-of-sample cumulative predicted returns using the macro predictive variables. This chart plots the cumulative return of an example trading strategy based on the benchmark Random Forest cross-sectional return prediction model. The machine learning model uses only the macroeconomic predictive variables available on the Quantiacs platform as its inputs, to predict returns for the full universe of futures contracts.

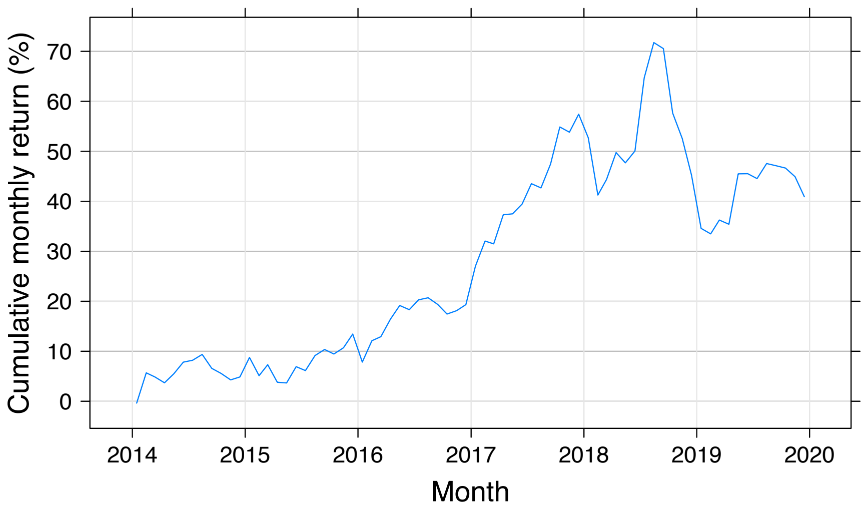


Figure 6: Soft-thresholding. Illustration of soft-thresholding an input. The dashed line denotes the benchmark $f(x) = x$. The solid blue line $f(x) = \text{sign}(x)\max\{|x| - \lambda, 0\})$ is the result of soft-thresholding the input x at threshold λ.

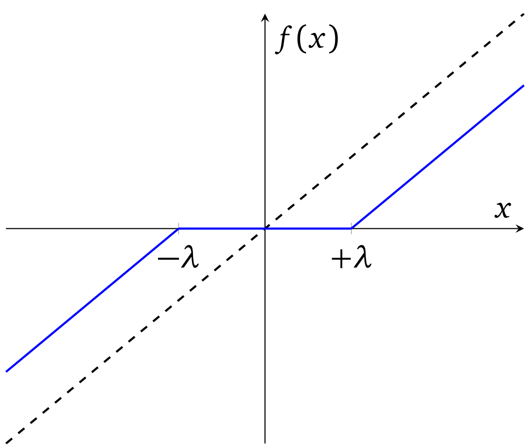
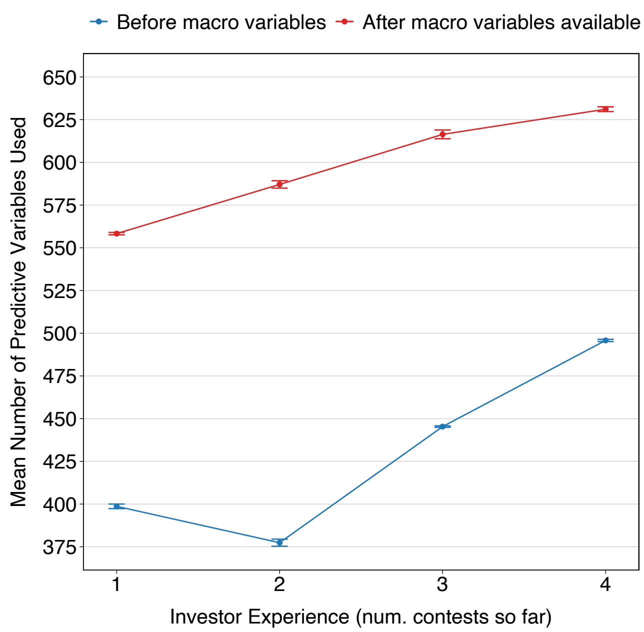


Figure 7: Investors' estimated usage of predictive variables. This chart shows the number of predictive variables used by investors. These are measured according to the model of investor behavior in Section 7.2, based on investors' daily portfolio returns and a set of predictive variables. A total of 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: 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 2: 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)
$\log(\lvert\widehat{\beta}_i - \widehat{\beta}_j\rvert)$	0.0486***	0.0508***				
	(0.0030)	(0.0026)				
$\log(1 + \lvert\widehat{\beta}_i - \widehat{\beta}_j\rvert)$			0.2423***	0.2866***		
			(0.0225)	(0.0362)		
$\text{arcsinh}(\lvert\widehat{\beta}_i - \widehat{\beta}_j\rvert)$					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 3: 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 $\text{SR}_{i,t}^{Best}$		Live $\text{SR}_{i,t}^{Best}$	
	(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 4: 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 5: 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$_{i,t}$ × 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 6: 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:	$\text{Volume}_{i,\tau}$				
Familiarity Index:		News Articles		Books	
	(1)	(2)	(3)	(4)	(5)
$\text{Experienced}_{i,t}$	0.0673**		-0.3381		-0.1500
	(0.0184)		(0.3085)		(0.2425)
$\text{Familiarity}_{\tau}$		0.0027***	0.0026***	0.0026***	0.0026***
		(0.0003)	(0.0003)	(0.0005)	(0.0005)
$\text{Experienced}_{i,t} \times \text{Familiarity}_{\tau}$			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 7: 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 $\text{experienced}_{i,t}$	1.0796	0.3183	1.1058	0.3985	0.0262	0.3108
$\text{Percentile}(\text{Score}^{Best}_{i,t-1})$	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**
$\text{Mean}_t(\text{Contests experienced}_{i,t})$	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 8: 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 $SR^{Best}_{i,t}$		
	(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 $SR^{Best}_{i,t}$		
	(1)	(2)	(3)
(Intercept)	−1.162***	−0.970**	−0.899**
	(0.341)	(0.425)	(0.451)
$\mathbb{1}$\{Contests experienced$_{i,t}$ = 2\}	1.332***	1.293***	1.044***
	(0.266)	(0.257)	(0.271)
$\mathbb{1}$\{Contests experienced$_{i,t}$ = 3\}	1.673*	1.654*	−0.004
	(0.914)	(0.896)	(0.459)
$\mathbb{1}$\{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)
$\mathbb{1}$\{Contests experienced$_{i,t}$ = 2\} × Macro variables available$_t$			0.478
			(0.504)
$\mathbb{1}$\{Contests experienced$_{i,t}$ = 3\} × Macro variables available$_t$			2.973**
			(1.491)
$\mathbb{1}$\{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 9: 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 $SR^{Best}_{i,t}$	Live $SR^{Best}_{i,t}$	Live $SR^{Best}_{i,t}$
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

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.

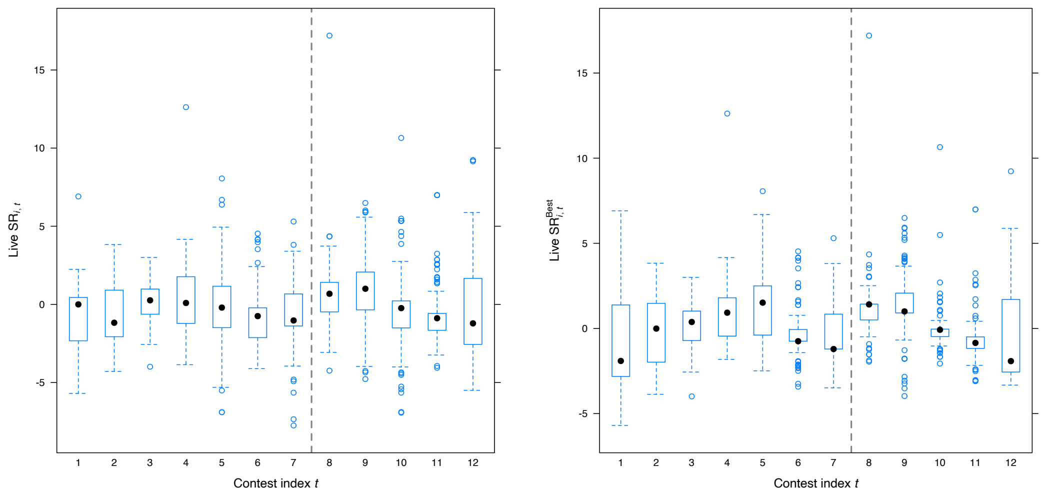


Figure IA.2: Participation and attrition. Proportions of contestants for whom this is the first contest, and those for whom this is the last contest (with breakdowns of the latter).

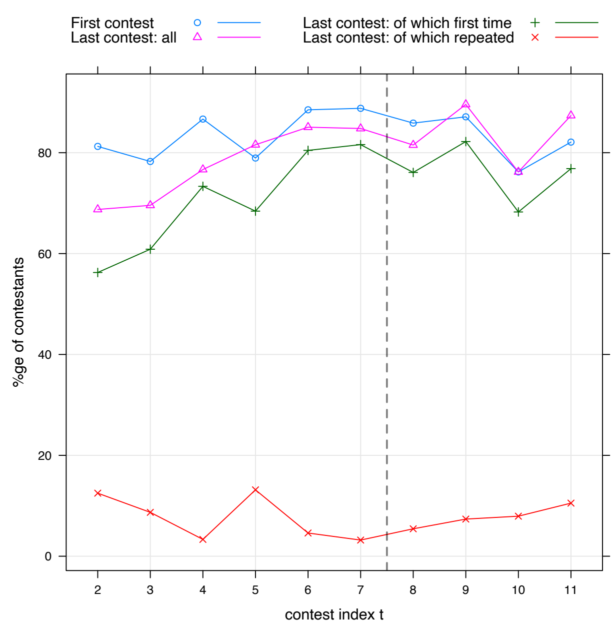


Figure IA.3: Concentration measures of individual macro variable informativeness and familiarity. Percentages of the total annual macro variable informativeness, and of the total annual macro variable mentions in news articles and books (i.e. familiarity), that are attributed to the top 3-ranked (LHS)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

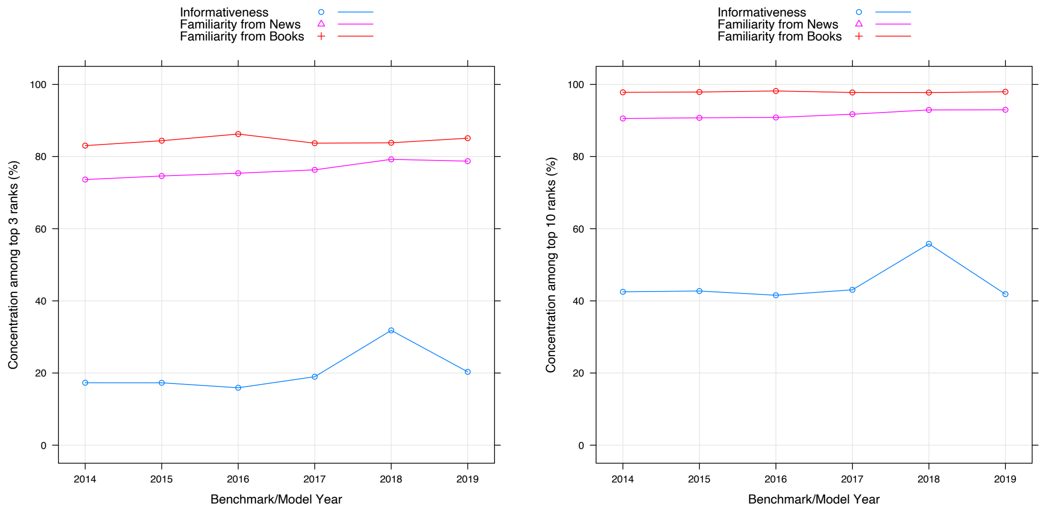
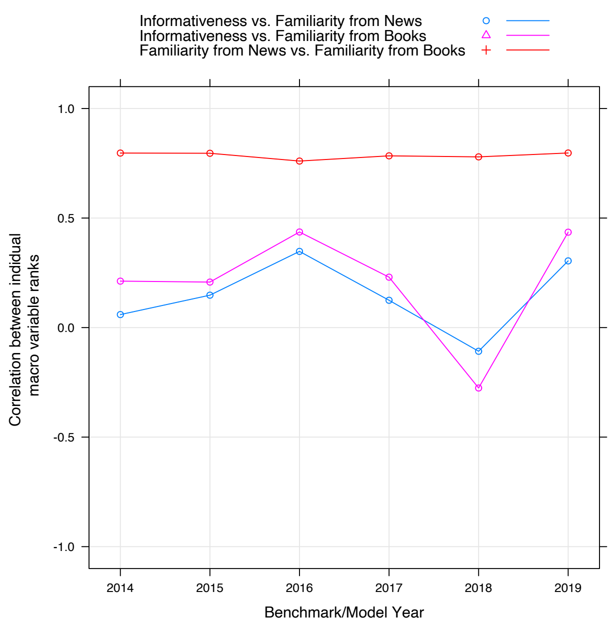


Figure IA.4: Rank-correlations of individual macro variable informativeness and familiarity. Correlations between macro variable informativeness and familiarity rankings. For both familiarity measures, rankings are computed each year using article and book mentions for that year. The informativeness 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 5 and 6 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)
Familiarity_τ	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 2.1 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 1 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.

Edit and analyze trading system

Select program

Blank Trading S ▾ | ▶ Run | Optimize | Export Data

```
1  ### Quantiacs Trading System Template
2
3  # import necessary Packages below:
4  import numpy
5  #import pandas
6  #import scikit.learn
7  #import sciPy
8
9  def myTradingSystem(DATE, OPEN, HIGH, LOW, CLOSE, VOL, OI, P, R, RINFO, exposure, equity, settings):
10     '''Define your trading system here.
11     See the example trading system for a starting point.
12     The function name "myTradingSystem" should not be changed. We evaluate this function on our server.
13     Your system should return a normalized set of weights for the markets you have defined in settings['markets']. '''
14
15     # this strategy implements a simple buy and hold strategy
16     nMarkets = CLOSE.shape[1]
17     pos = numpy.ones((1, nMarkets))
18     pos = pos / numpy.sum(abs(pos))
19
20     return pos, settings
21
22
23 def mySettings():
24     '''Define your market list and other settings here.
25
26     The function name "mySettings" should not be changed.
27
28     Default settings are shown below.'''
29
30     settings={}
31
32     # Futures Contracts
33     settings['markets'] = ['F_ES', 'F_MD', 'F_NQ', 'F_RU', 'F_XX',
34         'F_YM', 'F_AX', 'F_CA', 'F_LX', 'F_VX', 'F_AE', 'F_DM',
35         'F_AH', 'F_DZ', 'F_FB', 'F_FM', 'F_FP', 'F_FY', 'F_NY',
36         'F_PQ', 'F_SH', 'F_SX', 'F_GD', 'F_FV', 'F_TU', 'F_TY',
37         'F_US', 'F_DT', 'F_UB', 'F_UZ', 'F_GS', 'F_CF', 'F_GX',
```

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.

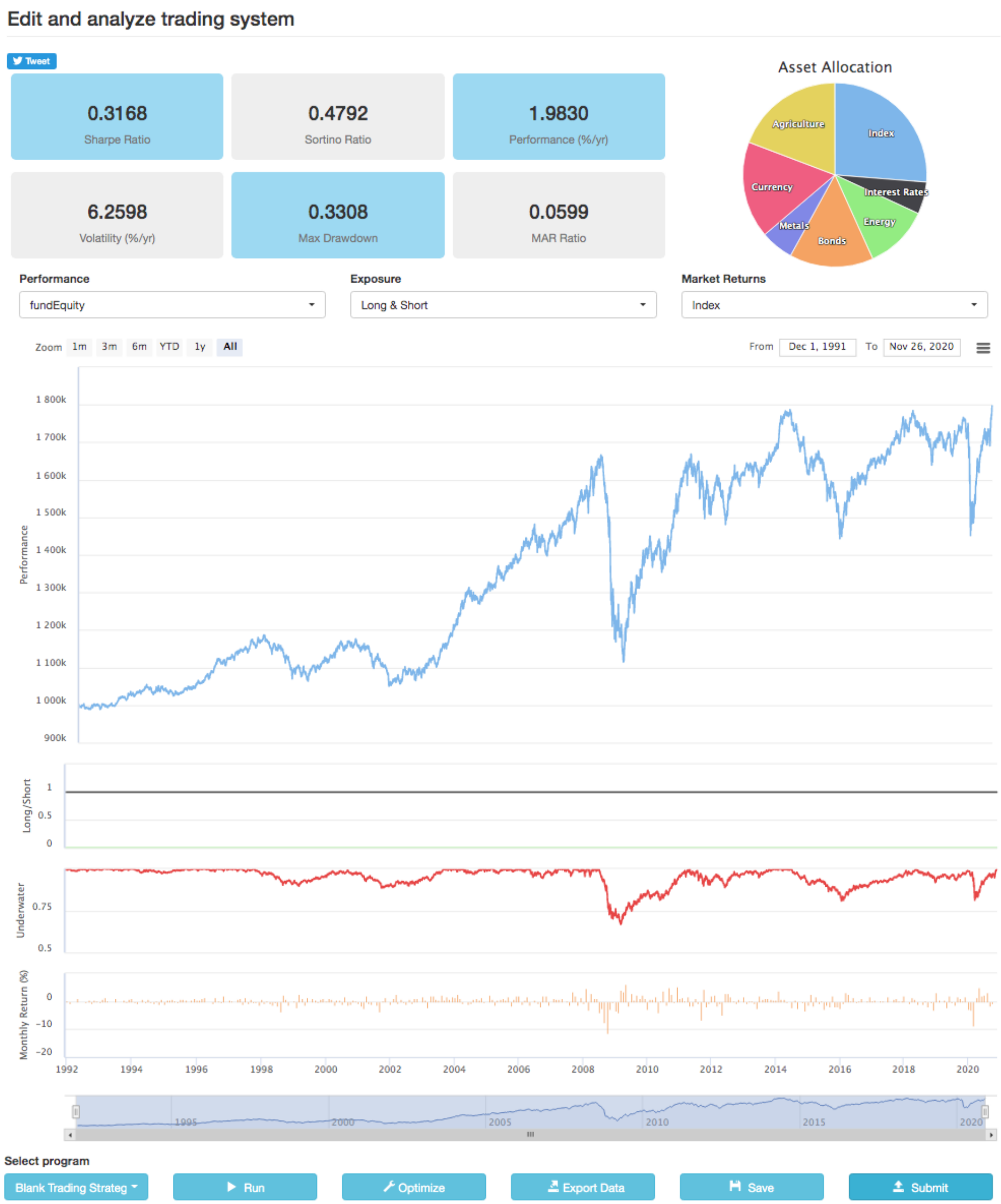
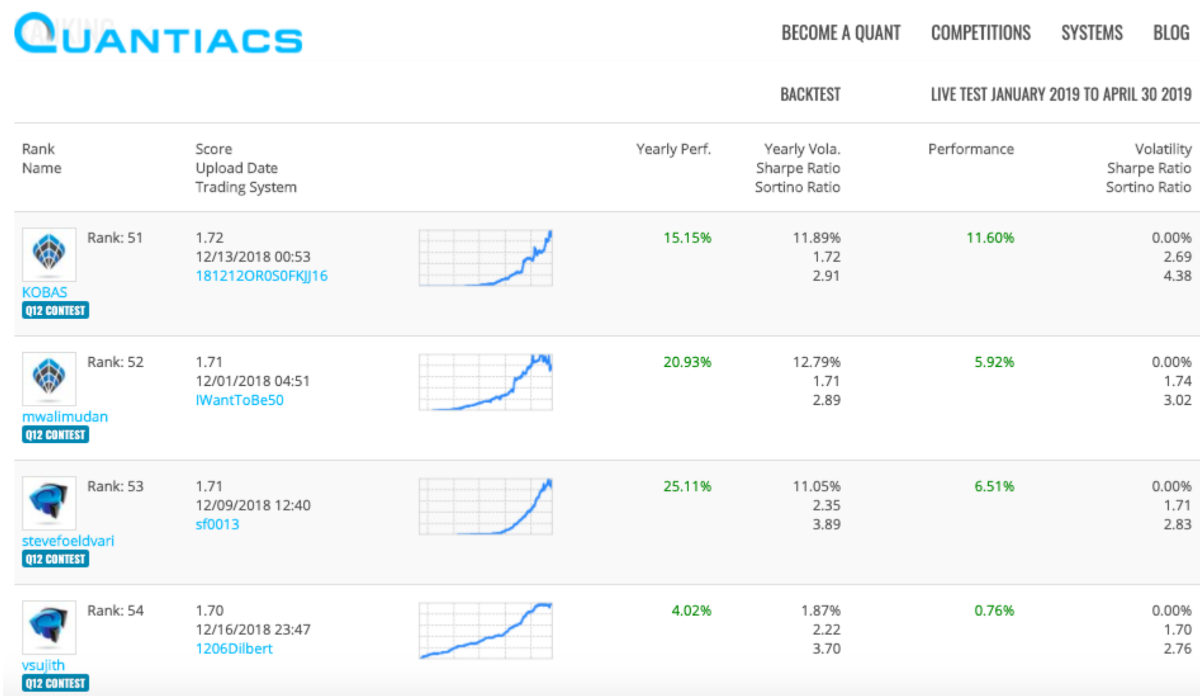


Figure IA.7: Contest outcomes once Live period ends. Excerpt from a contest leaderboard, showing both in-sample and out-of-sample Sharpe Ratios, after a trading contest has ended. The "live test" performance metrics were calculated once the contest period (1 January to 30 April 2019, in this case) had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



		BACKTEST		LIVE TEST JANUARY 2019 TO APRIL 30 2019	
Rank Name	Score Upload Date Trading System	Yearly Perf.	Yearly Vola. Sharpe Ratio Sortino Ratio	Performance	Volatility Sharpe Ratio Sortino Ratio
Rank: 51 KOBAS Q12 CONTEST	1.72 12/13/2018 00:53 181212OR0S0FKJJ16	15.15%	11.89% 1.72 2.91	11.60%	0.00% 2.69 4.38
Rank: 52 mwalimudan Q12 CONTEST	1.71 12/01/2018 04:51 IWantToBe50	20.93%	12.79% 1.71 2.89	5.92%	0.00% 1.74 3.02
Rank: 53 stevefoeldvari Q12 CONTEST	1.71 12/09/2018 12:40 sf0013	25.11%	11.05% 2.35 3.89	6.51%	0.00% 1.71 2.83
Rank: 54 vsujith Q12 CONTEST	1.70 12/16/2018 23:47 1206Dilbert	4.02%	1.87% 2.22 3.70	0.76%	0.00% 1.70 2.76

D Futures Contracts and the Benchmark Portfolio

D.1 Market Data for Futures Contracts

As discussed in Section 2 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.[28] 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
12	F_CD	Canadian Dollar	Currency	CD Curncy

[28] 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
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
53	F_NQ	E-mini Nasdaq 100 Index	Index	NQ Index

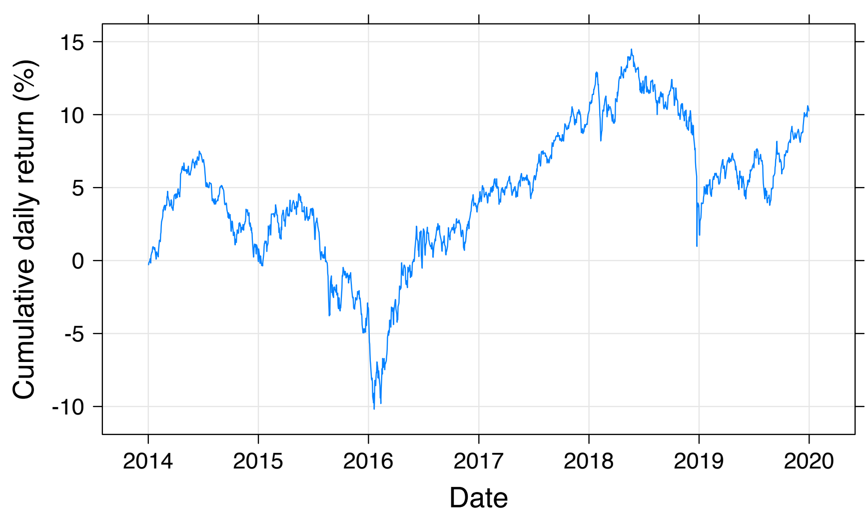
	Quantiacs ticker	Name	Type	Matching Bloomberg base code
54	F_NR	Rough Rice	Agriculture	RR Comdty
55	F_NY	Nikkei 225	Index	NI Index
56	F_O	Oats	Agriculture	O Comdty
57	F_OJ	Orange Juice	Agriculture	JO Comdty
58	F_PA	Palladium	Metal	PA Comdty
59	F_PL	Platinum	Metal	PL Comdty
60	F_PQ	PSI20	Index	PP Index
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 2 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 in the Treatment group, 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 in the Control group, 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 Control 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 Treated investors, for whom the macro variables were accessible. Columns (3) and (4) are for the sub-sample of Control investors, who could not access the macro variables via their trading strategies.

	$\text{Volume}_{i,\tau}$			
	Treated investors		*Control investors*	
	(1)	(2)	(3)	(4)
(Intercept)	0.002***	0.003***	0.002***	0.002***
	(0.0004)	(0.0001)	(0.001)	(0.001)
New variables available_t	−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)
Macro release_τ × $\text{Experienced}_{i,t}$		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 (14) 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:

$$\begin{aligned}
&\text{Mean}_{i,t}(\text{Best Live entry daily returns}) \\
&= \beta_1 + \sum_{k=2}^{4}\left[\beta_k \times \mathbb{1}\{\text{Contests experienced}_{i,t} = k\}\right] + \gamma_1 \times \text{New variables available}_t \\
&\quad + \sum_{k=2}^{4}\left[\gamma_k \times \mathbb{1}\{\text{Contests experienced}_{i,t} = k\} \times \text{New variables available}_t\right] + \epsilon_{i,t}. \qquad (34)
\end{aligned}$$

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 (14) and previous specification (34) 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:

$$\begin{aligned}
&\text{SD}_{i,t}(\text{Best Live entry daily returns}) \\
&= \beta_1 + \sum_{k=2}^{4}\left[\beta_k \times \mathbb{1}\{\text{Contests experienced}_{i,t} = k\}\right] + \gamma_1 \times \text{New variables available}_t \\
&\quad + \sum_{k=2}^{4}\left[\gamma_k \times \mathbb{1}\{\text{Contests experienced}_{i,t} = k\} \times \text{New variables available}_t\right] + \epsilon_{i,t}. \qquad (35)
\end{aligned}$$

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 8, 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 $\text{SR}_{i,t}^{Best}$					
	(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 8.
- Table IA.11 modifies the specification of Table 9.

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:					
	$\text{Score}^{Best}_{i,t}$			$\text{Score}^{Best}_{i,t} - \text{Benchmark SR}_t$		
	(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		$\text{Score}^{Best}_{i,t}$	$\text{Score}^{Best}_{i,t}$	$\text{Score}^{Best}_{i,t}$
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)	-0.99^{***}	-0.85^{***}	-0.14
		(0.08)	(0.30)	(0.14)
	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). $^{***}p < 0.01$; $^{**}p < 0.05$; $^{*}p < 0.1$