

# Forecasting Replications with Prediction Markets and Surveys

Michael Gordon, Domenico Viganola, Thomas Pfeiffer

## Introduction

- In 2005 John Ioannidis' essay "Why Most Published Findings Are False" argued that there is a large number of false positives in published research. A number of large-scale replication projects aimed to address this question, including;
  - Replication Project: Psychology (Open Science Collaboration. Science 2015),
  - Experimental Economics Replication Project (Camerer et al. Science 2016),
  - Social Sciences Replication Project (Camerer et al. Nat Hum Behav 2018),
  - Many Labs 2 (Klein et al. Adv Methods Pract Psychol Sci 2018)
- Replication success rates ranged from 39% to 62%.
- Prediction markets and surveys were used to forecast replication outcomes.

## Methods

- Prediction markets:** participants traded securities that pay \$1 for successful replication and \$0 for unsuccessful replication. Final market price is taken as probability of replication success.
- Pre Market Survey:** Participants are asked to give a probability of replication for each study.
- Data:** Pooled data from the four projects allows for higher powered analysis (n=103). R Package is available (check QR Code).

## Prediction Markets vs Surveys

- The prediction market correctly forecasted 75 of 103 replications correctly (73%, Brier Score = 0.192), see figure 1.
- Mean survey responses had an accuracy of 66% (Brier Score = 0.205).
- High correlation was found between variation of participants survey responses and their mean absolute prediction error (see figure 2).
- Using the responses from the top 25% of participants (in terms of response variation) to calculate the mean, produces a more accurate aggregation of the survey than using all users responses. This result reached 76% accuracy (Brier Score = 0.181)

## P-Values

- P-values are highly predictive of replication outcome.
  - P-value smaller than 0.001  $\Rightarrow$  82% replication success
  - P-value between 0.001 and 0.01  $\Rightarrow$  44% replication success
  - P-Value greater than 0.1  $\Rightarrow$  27% replication success
- Market price is more predictive than p-value alone. This indicates p-value is not the only information in the market.
- Market price and p-value combined is more informative than price alone. Thus p-value is not fully priced into the market (Figure 3).

## Market Dynamics

- The market experiences diminishing returns in terms of reductions in absolute errors (figure 4).
- 95% of the reduction in mean absolute error occurs in the first 31 hours of a market.

## Discussion

- Replicability of published studies is not obvious to a broader audience but can be elicited from the scientific community with prediction markets and surveys.
- Identifying the participants who have high response variation, can greatly improve upon the base survey aggregation.
- Market's quickly reach diminishing returns of information gain.
- Results inform upcoming replication projects - Follow QR Code for more information

# Prediction markets reliably forecast replication success in science.



Data, Working Paper and References can be found at <https://github.com/MichaelGordon/PooledMarketR> or scan this QR Code



Figure's 1 & 2 – Prediction Market and Survey Market Forecasts

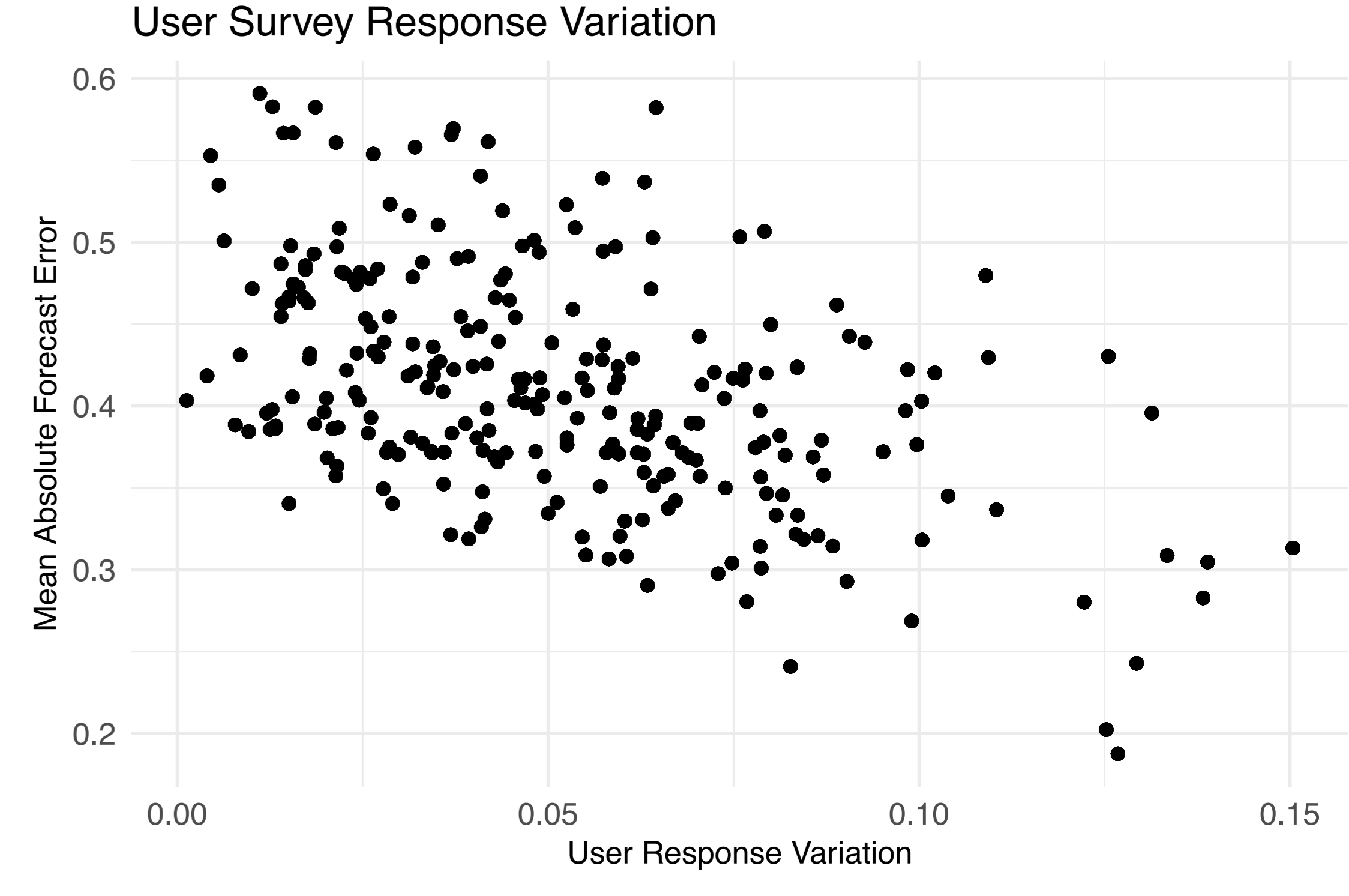
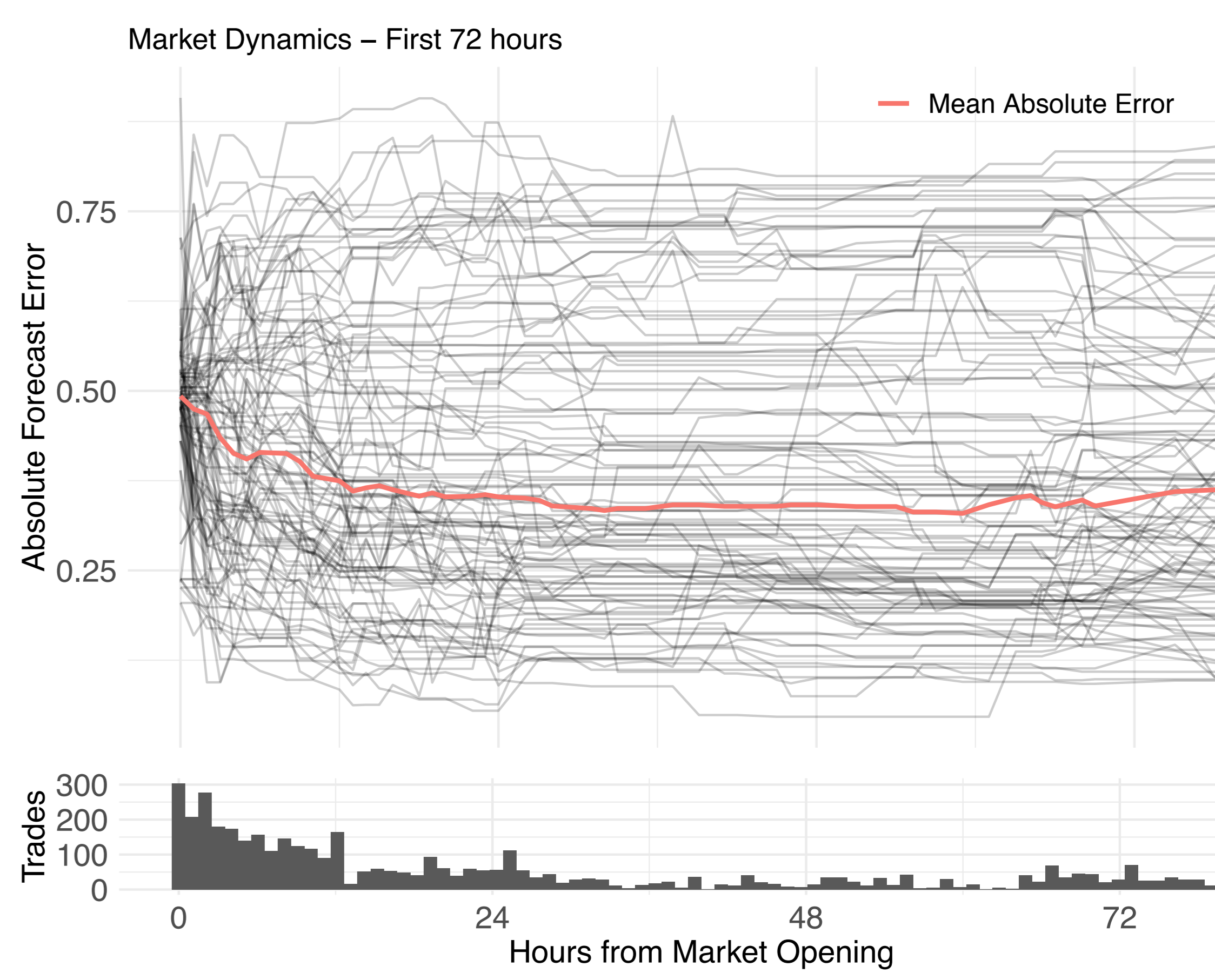


Table 1 – P-Values

	Intercept $\alpha$	Market Price $\beta$	P-value 0.001 cat $\beta$	P-value 0.01 cat $\beta$	AIC
Market Price Model	-0.40*	1.42*			114
P-value Model	0.27*		0.56*	0.17	131
Combination Model	-0.33*	1.16*	0.31*	0.002	105

\* denotes p-value < 0.05

Figure 3: Market Dynamics



Check out our latest project here: <http://www.replicationmarkets.com>