What Goes Up Must Come Down?
Experimental Evidence on Intuitive Forecasting

By John Beshears, James J. Choi, Andreas Fuster, David Laibson, and Brigitte C. Madrian

Abstract: Do laboratory subjects correctly perceive the dynamics of a mean-reverting time series? In our experiment, subjects receive historical data and make forecasts at different horizons. The time series process that we use features short-run momentum and long-run partial mean reversion. Half of the subjects see a version of this process in which the momentum and partial mean reversion unfold over 10 periods ('fast'), while the other subjects see a version with dynamics that unfold over 50 periods ('slow'). Typical subjects recognize most of the mean reversion of the fast process and none of the mean reversion of the slow process.

* Beshears: Stanford Graduate School of Business, 655 Knight Way, Stanford, CA 94305 (e-mail: beshears@stanford.edu); Choi: Yale School of Management, 135 Prospect St., New Haven, CT 06520 (e-mail: james.choi@yale.edu); Fuster: Federal Reserve Bank of New York, 33 Liberty St., New York, NY 10045 (e-mail: andreas.fuster@ny.frb.org); Laibson: Dept. of Economics, Harvard University, Cambridge, MA 02138 (e-mail: dlaibson@harvard.edu); Madrian: Harvard Kennedy School, 79 JFK St., Cambridge, MA 02138 (e-mail: brigitte_madrian@harvard.edu). We thank Chris Clayton for excellent research assistance. We acknowledge individual and collective financial support from the National Institute on Aging (grants P01AG005842, R01AG021650 and P30AG034532). The views expressed in this paper are those of the authors and do not necessarily reflect the position of the Federal Reserve Bank of New York or the Federal Reserve System.

Beliefs about the future are central elements of dynamic economic models. While most economic analysis assumes rational expectations, a growing theoretical literature relaxes this restriction, and a growing empirical literature investigates how economic actors actually form their beliefs.[1]

The current paper contributes to this literature by experimentally measuring the degree to which people intuitively recognize mean reversion. Study participants view data generated by an integrated time series process that is characterized by short-run momentum and long-run partial mean reversion. For half of our participants, these dynamics play out completely in 10 periods; we call this the "fast" process. For the other half, the process has the same momentum and mean reversion, but the dynamics play out over 50 rather than 10 periods; we call this the "slow" process.

We give subjects a large sample of past observations of the process and ask them to make a series of forecasts at different horizons. Fitting these forecasts to a set of prespecified candidate models, we infer subjects' beliefs about the underlying data generating process and the extent of mean reversion. Subjects are better at recognizing mean reversion when it unfolds quickly. For the fast process, the median participant makes forecasts that capture 60 percent of the actual mean rever-

[1] See Michael Woodford (2012) for a review.

sion. For the slow process, the median participant makes forecasts that capture *none* of the actual mean reversion. If economic agents in the field also fail to recognize the full extent of mean reversion in economic fundamentals (e.g., corporate earnings), this would explain a wide range of empirical regularities, including cycles in consumption and investment, as well as excess volatility and predictable variation in asset returns (see, e.g., Robert Barsky and Bradford DeLong 1993; Fuster, Laibson, and Brock Mendel 2010; Fuster, Benjamin Hebert, and Laibson 2012).

This paper extends research that has studied expectation formation in the laboratory (e.g., Richard Schmalensee 1976; Gerald Dwyer et al. 1993; John Hey 1994; Cars Hommes 2011; Tobias Rötheli 2011).[2] In the laboratory, researchers can control the data generating process that produces "historical" data. Researchers can also control the information given to subjects and assess subject performance against a known benchmark. Of course, the laboratory setting raises questions of external validity because the forecasting exercise lacks context, subjects face weak financial incentives, and individuals' expectations in the field are influenced by neighbors, co-workers, family, the media, and professional forecasters (Christopher Carroll 2003). Nonetheless, laboratory experiments shed light on individuals' *intuitive forecasts*. Intuitive forecasts may serve as a starting point, or "anchor," that biases people's beliefs (Amos Tversky and Daniel Kahneman 1974).

Our paper also relates to research that studies survey forecasts of future economic outcomes such as stock returns or house price appreciation. This literature finds that people often place too much weight on recent experience and over-extrapolate (see Ulrike Malmendier and Stefan Nagel 2011; Karl Case, Robert Shiller, and Anne Thompson 2012; and Robin Greenwood and Andrei Shleifer 2012 for recent examples). Such over-extrapolation reduces agents' ability to anticipate mean reversion.

I. Experimental Setup

Subjects were recruited for a forecasting experiment in which they were randomly assigned data generated by one of six integrated moving average processes, two of which we analyze in this paper.[3] Figure 1 shows the two processes' impulse response functions. The "fast" process has dynamics that are fully realized in 10 periods: ARIMA(0,1,10). The

[2] There is also a substantial literature, mostly outside of economics, on "judgmental forecasting" (see, e.g., Michael Lawrence et al. 2006).

[3] The other processes are described in the Online Appendix and will be analyzed in future work. The Appendix also includes plots of simulated paths of the two processes we analyze, the exact MA coefficients of these processes, the experimental instructions and protocol, and additional details on the analyses in this paper.

"slow" process has dynamics that are fully realized in 50 periods: ARIMA(0,1,50). The slow process is a stretched version of the fast process, with dynamics that take five times as long to play out.[4] Otherwise, the processes are identical.

These ARIMA processes feature short-run momentum and long-run mean reversion. After an impulse is realized, the processes trend in the same direction, peaking at a level 50 percent above the level of the initial impulse before subsequently mean-reverting to a level 50 percent below the level of the initial impulse.

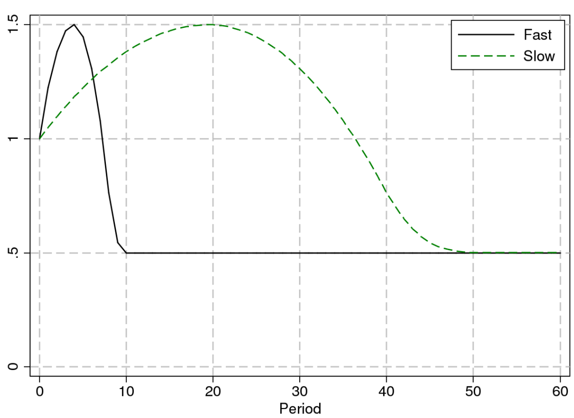

FIGURE 1. IMPULSE RESPONSE FUNCTIONS FOR THE FAST AND SLOW PROCESSES

Short-run momentum and long-run mean reversion characterize the dynamics of macroeconomic variables like GDP, unemployment, and corporate earnings (Fuster, Laibson, and Mendel 2010). Furthermore, many of these time series have relatively slow dynamics, treating their reporting frequency as the time unit.

We conducted the experiment on individual computer stations in the Harvard Decision Science Lab. Participants had access to 100,000 periods of simulated historical data (different for each participant) and a simple interface that displayed past observations in graphical form and in a scrollable list. Participants could change the number of past observations displayed as desired. No other tools (such as calculators) were available. Participants were not shown an impulse response function or given a quantitative description of or any context for the data generating process. They were simply told that the data were generated by statistical rules that would remain unchanged over the course of the experiment and were unaffected by the participants' forecasts.

Experimental sessions comprised 60 periods. In each period, participants made a forecast of the process's *n*-period-ahead realization, where *n* was randomly drawn (for that period) from the set {1, 5, 10, 20, 35, 50}.[5] After a forecast was submitted, the next period's value of the series was revealed, and the

[4] If θ_f is the f^{th} moving average term of the fast process and θ_s is the s^{th} moving average term of the slow process, then $\theta_f = \sum_{s=5f-4}^{5f} \theta_s$.

[5] However, the randomization was set so that the subject would never make the same horizon forecast on consecutive forecasts.

participant was informed of the success or failure of any past forecasts she had made of that next period's value. Successful forecasts, defined as being within 10 units of the realized value, earned a \$0.50 accuracy payment.

Our sample contains 98 subjects, of whom 50 received the fast process and 48 received the slow process. Experimental sessions lasted 30-45 minutes, and subjects earned \$16.68 on average (a \$10 show-up fee plus the accuracy payments, which were earned on slightly less than one quarter of the forecasts).

II. Results

In theory, subject forecasts are a function of all the historical data of the relevant time series (100,000+ observations). It is challenging to infer this mapping, since each subject only made 60 forecasts during the experiment. To surmount this identification problem, we take a structural approach by identifying a set of pre-specified models (with fixed coefficients) and searching for the model that best fits each subject's forecasts.

We assume that subjects make forecasts using an ARIMA(0,1,q) model, the same class of models used to generate the data, but do not know the true order of the ARIMA process, q^*. We calculate the value of q that best fits the forecasts subject i generated in periods 11 to 60.[6] Define $\hat{q}_i$ as:[7]

$$\hat{q}_i \equiv \underset{q\in\{0,1,\ldots,q^*\}}{\arg\min} \sum_{t=11}^{60} \left|\hat{x}_{i,t} - x_{i,t}^{ARIMA(0,1,q)}\right|.$$

We find the model order $\hat{q}_i$ that generates forecasts that minimize the average absolute deviation between the actual forecasts that subject i made at date t for a future period, $\hat{x}_{i,t}$, and the forecast (for the same future period) implied by the ARIMA(0,1,q) model, $x_{i,t}^{ARIMA(0,1,q)}$. To calculate $x_{i,t}^{ARIMA(0,1,q)}$ for a given q, we project the ARIMA(0,1,q) model on a 100,000 period sample generated by the true data generating process (see Appendix). We then apply the coefficients from this estimation (which are the same for each subject) to the historical data available to the subject at period t to calculate the forecast made in period t by the ARIMA(0,1,q) model.

Figures 2 and 3 plot the histograms of $\hat{q}_i$ values for the fast and slow data generating processes.[8] For the fast process, subjects' forecasts are largely explained by models whose specification is close to the true data generating process. Thirty-four percent of the

---

[6] We discard the first ten periods in our analysis because responses to a debriefing question, reported in the Appendix, suggest that it took the median subject about ten periods to gain familiarity with the task. We also discard the 1% of predictions that were furthest away from the realization in absolute value, as these were often caused by obvious typos.

[7] Our decision to minimize absolute deviations rather than squared deviations is intended to limit the influence of outliers.

[8] How well the models fit subjects' forecasts is discussed in the Appendix.

participants are best fit by an ARIMA(0,1,10) forecasting model, which corresponds exactly to the true data generating process. Only 12 percent of subjects are best fit by the simplest forecasting model considered, an ARIMA(0,1,0), which is a random walk.[9]

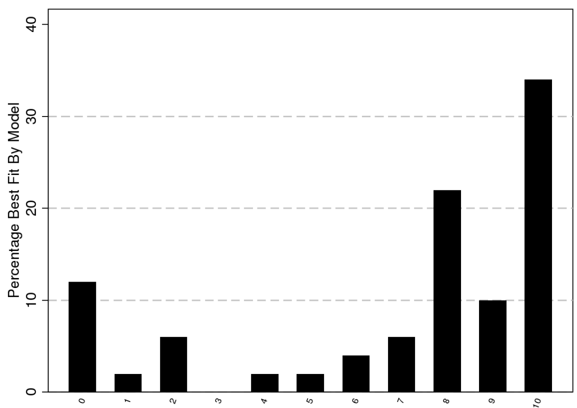

FIGURE 2. MODEL ASSIGNMENTS FOR FAST PROCESS

Note: The fast process is an ARIMA(0,1,10). We study projections of this process onto ARIMA(0,1,q) models, for $0 \leq q \leq 10$. Participants are assigned the ARIMA(0,1,q) model that best fits their forecasts.

For each subject, we also calculate the perceived extent of mean reversion, as implied by the chosen model, relative to the true extent of mean reversion:

$$\frac{1 - IRF(\infty, \hat{q}_i)}{1 - IRF(\infty, q^*)},$$

where $IRF(\infty, q)$ is the asymptotic value of the impulse response function implied by the model of order q. Ranking our subjects by perceived mean reversion, the model assigned to the median subject in the fast condition recognizes 59.5% of the true mean reversion.

In contrast, for the slow process, subjects' forecasts match ARIMA(0,1,q) models that are far from the true data generating process. Only 6 percent of the participants are best fit by the forecasting model that uses the true ARIMA(0,1,50) specification. By contrast, 29 percent of participants are best fit by the simplest forecasting model, the ARIMA(0,1,0). Ranking our subjects by perceived mean reversion, the model assigned to the median subject in the slow condition recognizes 0% of the true mean reversion.[10]

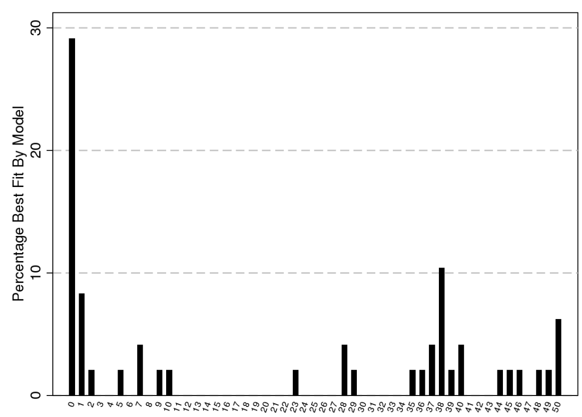

Figure 3. Model Assignments for Slow Process

Note: The slow process is an ARIMA(0,1,50). We study projections of this process onto ARIMA(0,1,q) models, for $0 \leq q \leq 50$. Participants are assigned the ARIMA(0,1,q) model that best fits their forecasts.

We complement our structural analysis with a reduced-form analysis. For each process, we pool data from all subjects and run the median regression

[9] The link between model order and expected performance in our forecasting task is not monotonic. ARIMA(0,1,q) models with "moderate" values of q tend not to predict any mean reversion at all, which leads to forecasts at long horizons that are far from the true data generating process's expectation.

[10] This is an exact zero, since the subjects who are assigned the random walk model as the best-fit approximation for their forecasts have the median level of perceived mean reversion.

$$\hat{x}_{i,t} - c_{i,t} = \alpha + \beta\left(x_{i,t}^{RE} - c_{i,t}\right) + \eta_{i,t},$$

where $x_{i,t}^{RE}$ is the forecast that would be issued at period t by an agent with rational expectations, $\hat{x}_{i,t}$ is the forecast that was actually issued at period t, and $c_{i,t}$ is the current value of the process at period t.[11] The null hypothesis of rational expectations implies $\alpha = 0$ and $\beta = 1$. The parameter β provides an index of congruence with rational expectations. When $\beta = 1$, actual forecasts move one for one with rational expectations. When $\beta = 0$, actual forecasts are orthogonal to rational expectations forecasts. For the fast process, the estimated $\hat{\beta}$ equals 0.60 (s.e.=0.03). For the slow process, the estimated $\hat{\beta}$ is 0.09 (s.e.=0.04), which implies that subjects' forecasts are nearly orthogonal to rational forecasts. The fast process is far more transparent to the subjects than the slow process.

III. Conclusion

Most participants failed to correctly perceive the degree of mean reversion in the processes that they analyzed. This bias was particularly acute for the statistical process with relatively slow dynamics. Worse performance on the slow process might be expected, since the individual moving average coefficients for the slow process are smaller in absolute value than the individual moving average coefficients for the fast process. However, even when we use our experimental methodology to study special cases in which the coefficient magnitudes are the *same* across two processes, we still find that slower processes tend to be far harder for subjects to parse correctly.[12]

Picking an as-if model of each subject's beliefs from a small pre-specified set of ARIMA models, as we have done here, provides only a first pass for studying forecasting behavior. Economics would greatly benefit from a general theory that explains how people recognize patterns in data and use those patterns to make forecasts.

REFERENCES

Barsky, Robert B. and J. Bradford DeLong. 1993. "Why Does the Stock Market Fluctuate?" *Quarterly Journal of Economics,* 108(2): 291-312.

Carroll, Christopher. 2003. "Macroeconomic Expectations of Households and Professional Forecasters." *Quarterly Journal of Economics,* 118(1): 269–298.

[11] Running separate median regressions for each subject produces qualitatively similar findings.

[12] Here we refer to two of the processes from our experiment that we are not able to discuss in this paper because of space constraints. These results will be discussed in future work.

Case, Karl E., Robert J. Shiller, and Anne Thompson. 2012. “What Have They Been Thinking? Home Buyer Behavior in Hot and Cold Markets,” *NBER Working Paper* 18400, National Bureau of Economic Research.

Dwyer, Gerald P., Arlington W. Williams, Raymond C. Battalio, and Timothy I. Mason. 1993. “Tests of Rational Expectations in a Stark Setting.” *Economic Journal*, 103(418): 586–601.

Fuster, Andreas, Benjamin Hebert, and David Laibson. 2012. “Natural Expectations, Macroeconomic Dynamics, and Asset Pricing.” *NBER Macroeconomics Annual 2011*. Edited by Daron Acemoglu and Michael Woodford. University of Chicago Press. 26: 1-48.

Fuster, Andreas, David Laibson, and Brock Mendel. 2010. “Natural Expectations and Macroeconomic Fluctuations.” *Journal of Economic Perspectives,* 24(4): 67-84.

Greenwood, Robin and Andrei Shleifer. 2012. “Expectations of Returns and Expected Returns.” Working Paper, Harvard University.

Hey, John D. 1994. “Expectations Formation: Rational or Adaptive or ...?” *Journal of Economic Behavior and Organization*, 25(3): 329-349.

Hommes, Cars. 2011. “The Heterogeneous Expectations Hypothesis: Some Evidence From the Lab.” *Journal of Economic Dynamics and Control*, 35(1): 1-24.

Lawrence, Michael, Paul Goodwin, Marcus O’Connor, and Dilek Önkal. 2006. “Judgmental Forecasting: A Review of Progress Over the Last 25 Years.” *International Journal of Forecasting,* 22(3): 493-518.

Malmendier, Ulrike and Stefan Nagel. 2011. “Depression Babies: Do Macroeconomic Experiences Affect Risk-Taking?” *Quarterly Journal of Economics*. 126(1): 373-416.

Rötheli, Tobias F. 2011. “Pattern-Based Expectations: International Experimental Evidence and Applications in Financial Economics.” *Review of Economics and Statistics,* 93(4): 1319-1330.

Schmalensee, Richard. 1976. “An Experimental Study of Expectation Formation.” *Econometrica,* 44(1): 17-41.

Tversky, Amos, and Daniel Kahneman. 1974. "Judgment Under Uncertainty: Heuristics and Biases." *Science* 185(4157): 1124-1131.

Woodford, Michael. 2012. "Macroeconomic Analysis without the Rational Expectations Hypothesis." *Annual Review of Economics*, forthcoming.