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.

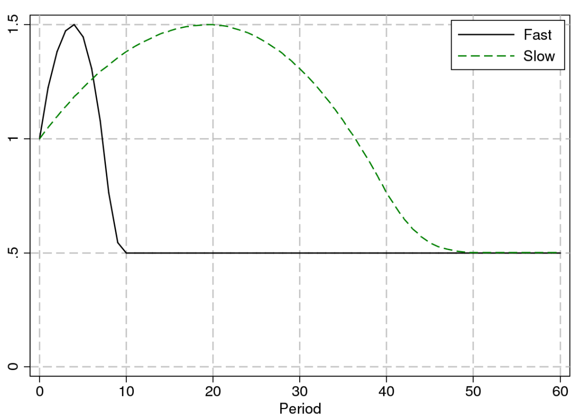


FIGURE 1. IMPULSE RESPONSE FUNCTIONS FOR THE FAST AND SLOW PROCESSES

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]

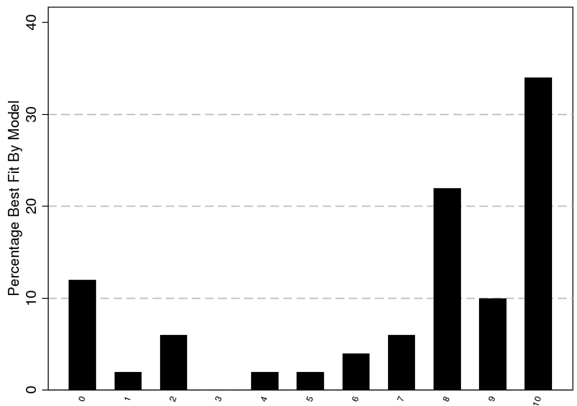


FIGURE 2. MODEL ASSIGNMENTS FOR FAST PROCESS

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]

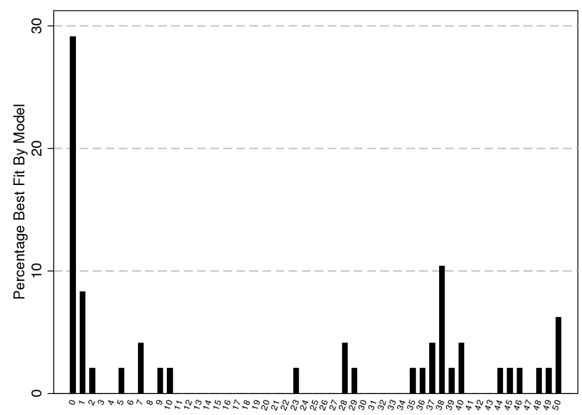


Figure 3. Model Assignments for Slow Process

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.

Appendix to “What Goes Up Must Come Down? Experimental Evidence on Intuitive Forecasting” (Beshears, Choi, Fuster, Laibson, and Madrian), 2013.

Appendix A: Data Generating Processes

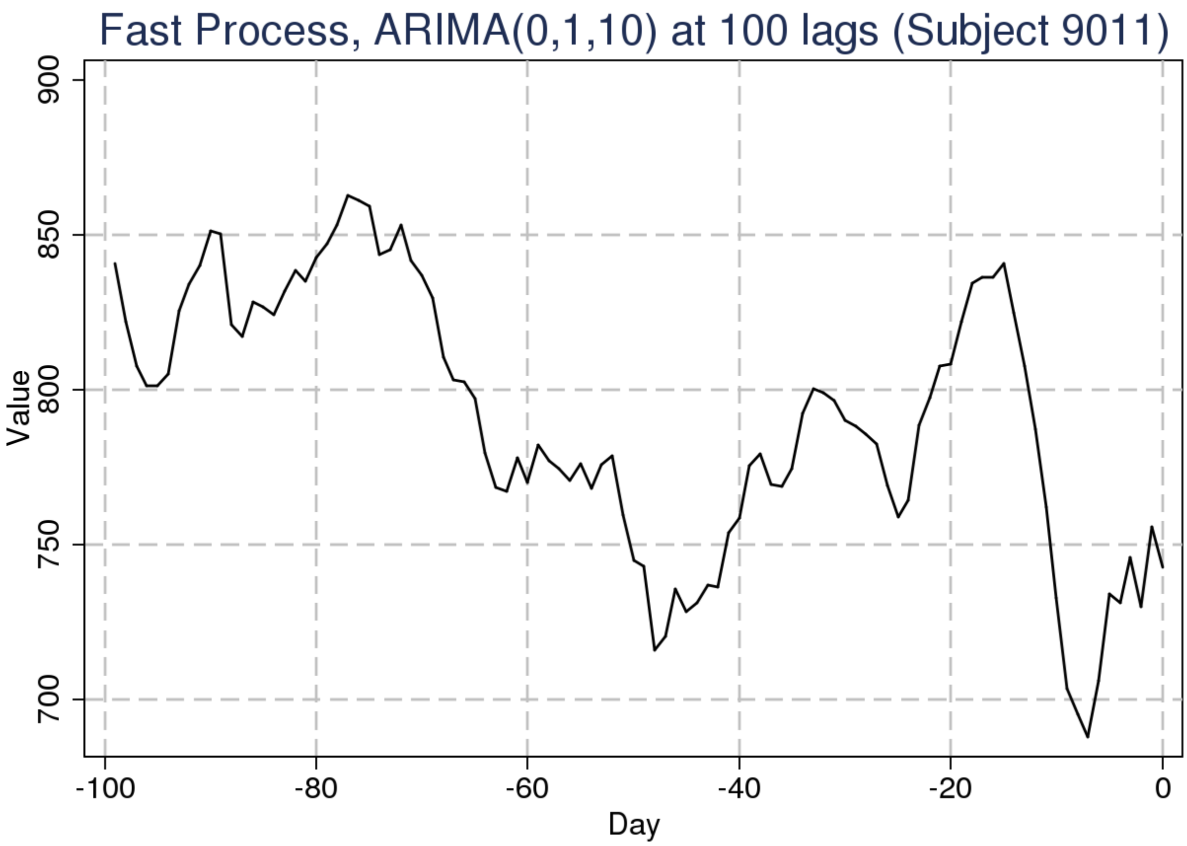
Appendix Table A1 lists the moving average coefficients[1] for each data generating process. Shock terms were drawn independently from a N(0,100) distribution. All process realizations were modified by an additive constant that set the “Day 0” value to 742.8. Appendix Figures A2-A4 show plots of 100, 500, and 10,000 periods of the realization of the fast (ARIMA(0,1,10)) process drawn for a randomly chosen subject. Figures A5-A7 show plots of 100, 500, and 10,000 periods of the realization of the slow (ARIMA(0,1,50)) process drawn for a randomly chosen subject.

Appendix Table A1: MA Coefficients

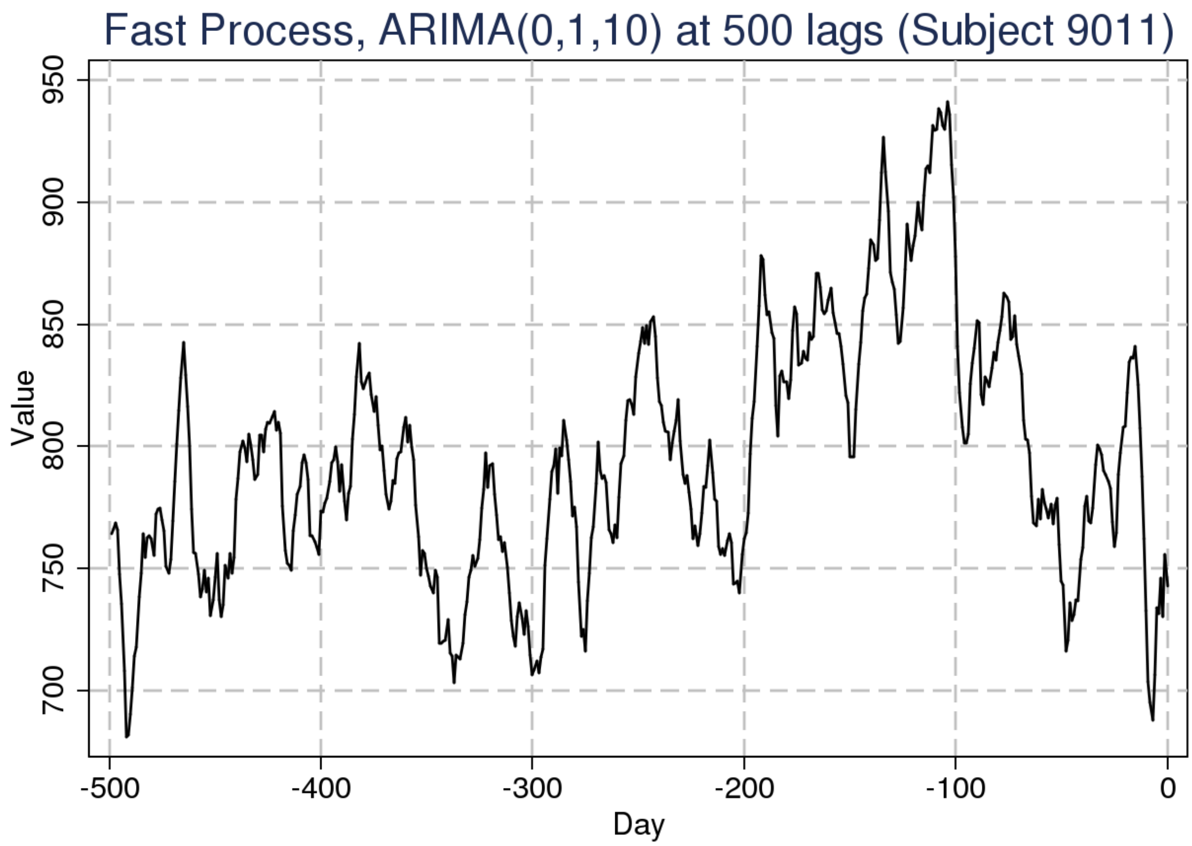
MA Term	Slow Process	Fast Process	MA Term	Slow Process	Fast Process
1	0.050	0.224	26	-0.021	
2	0.048	0.158	27	-0.025	
3	0.045	0.092	28	-0.028	
4	0.042	0.027	29	-0.031	
5	0.040	-0.053	30	-0.035	
6	0.037	-0.141	31	-0.039	
7	0.034	-0.228	32	-0.042	
8	0.032	-0.315	33	-0.045	
9	0.029	-0.219	34	-0.049	
10	0.027	-0.044	35	-0.053	
11	0.024		36	-0.056	
12	0.021		37	-0.059	
13	0.019		38	-0.063	
14	0.016		39	-0.067	
15	0.013		40	-0.070	
16	0.011		41	-0.061	
17	0.008		42	-0.053	
18	0.006		43	-0.044	
19	0.003		44	-0.035	
20	0.000		45	-0.026	
21	-0.003		46	-0.017	
22	-0.007		47	-0.011	
23	-0.011		48	-0.008	
24	-0.014		49	-0.005	
25	-0.017		50	-0.002	

[1] Note: Rounded to 3 digits.

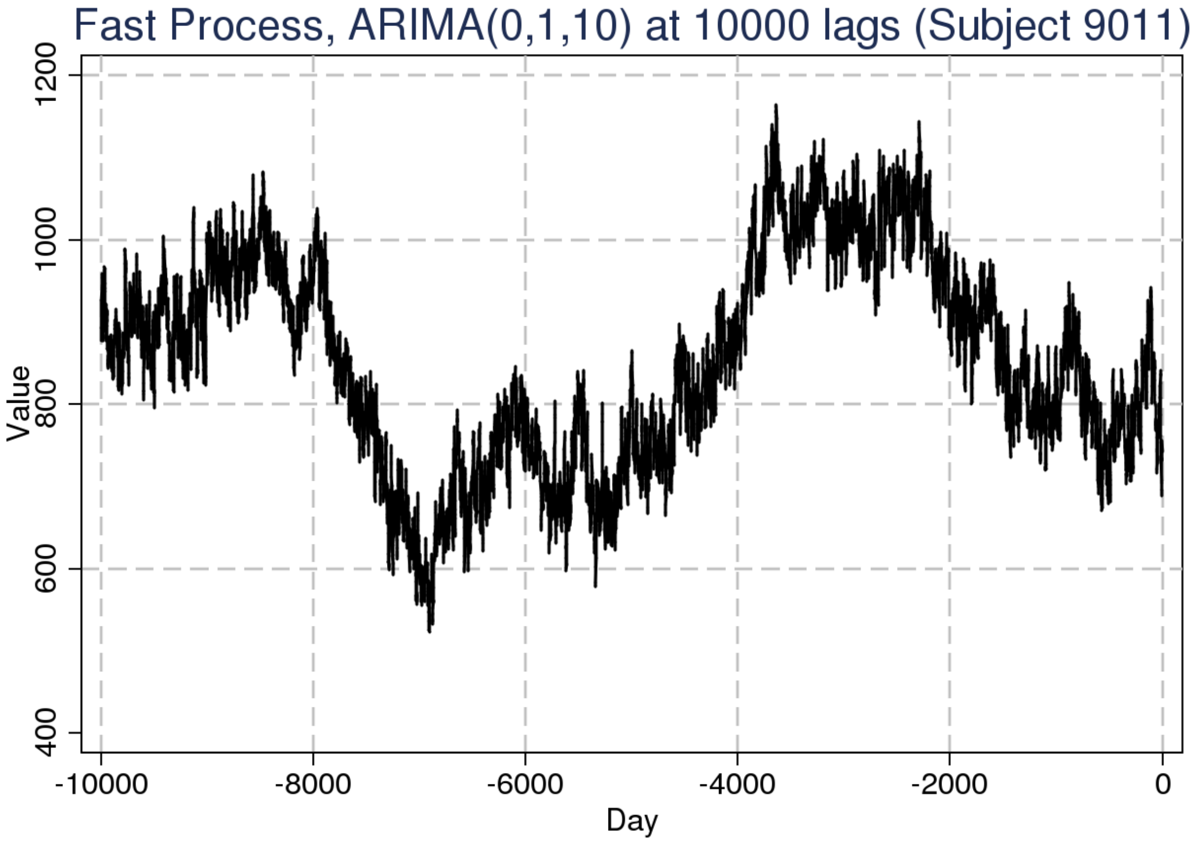
Appendix Figure A2: Sample Plot of Simulated Process



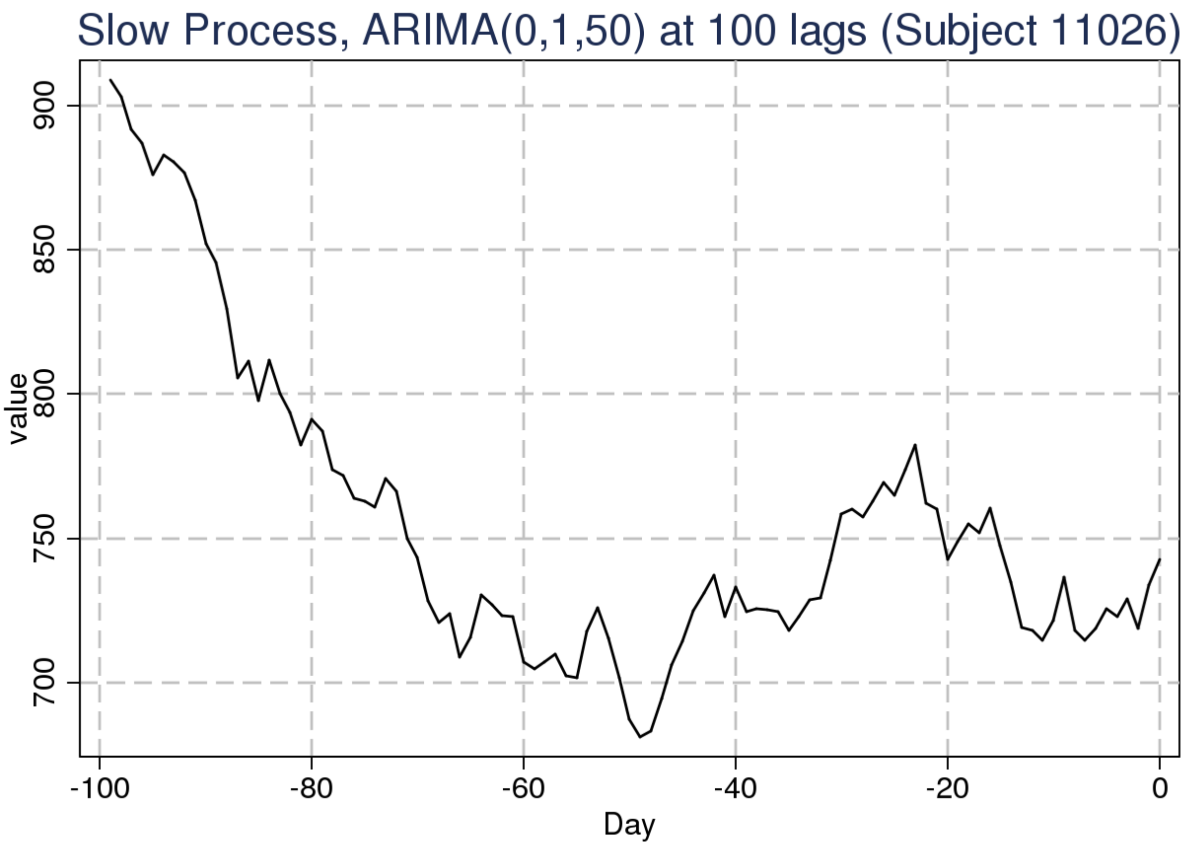
Appendix Figure A3: Sample Plot of Simulated Process



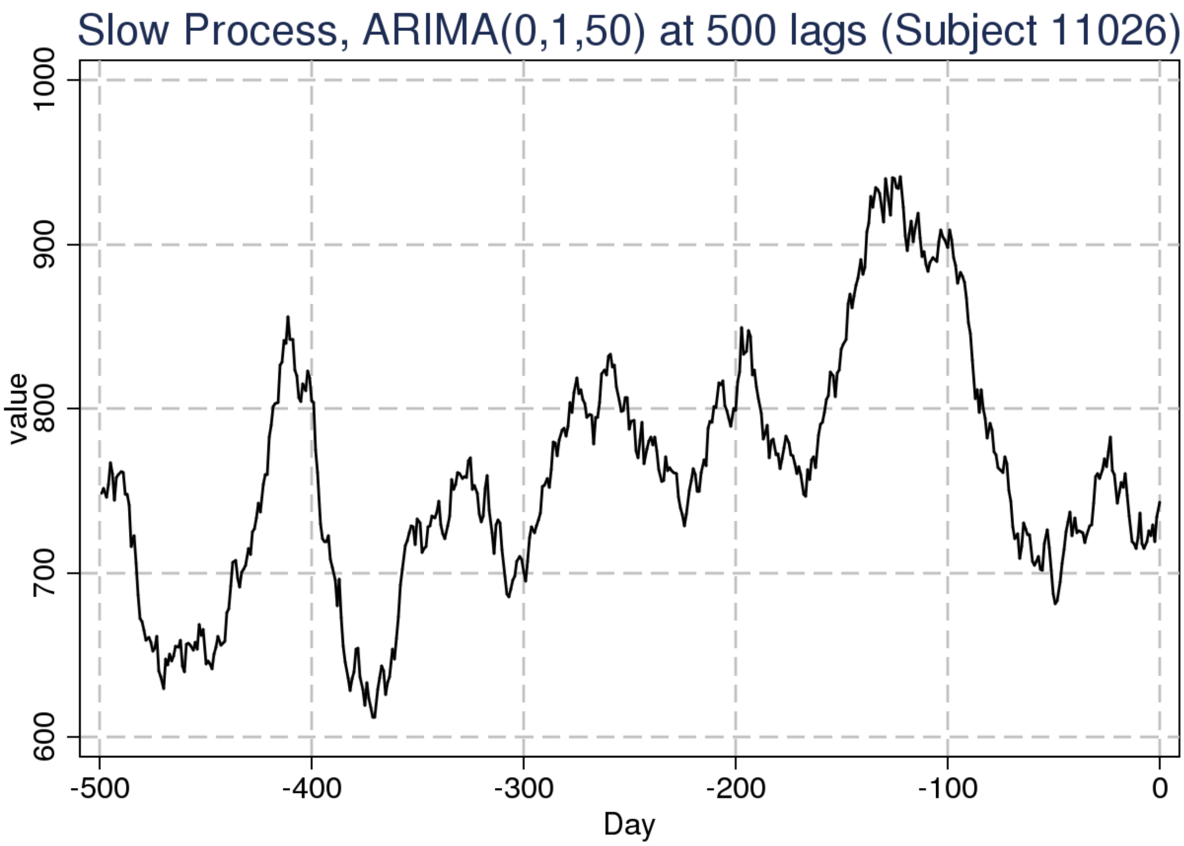
Appendix Figure A4: Sample Plot of Simulated Process



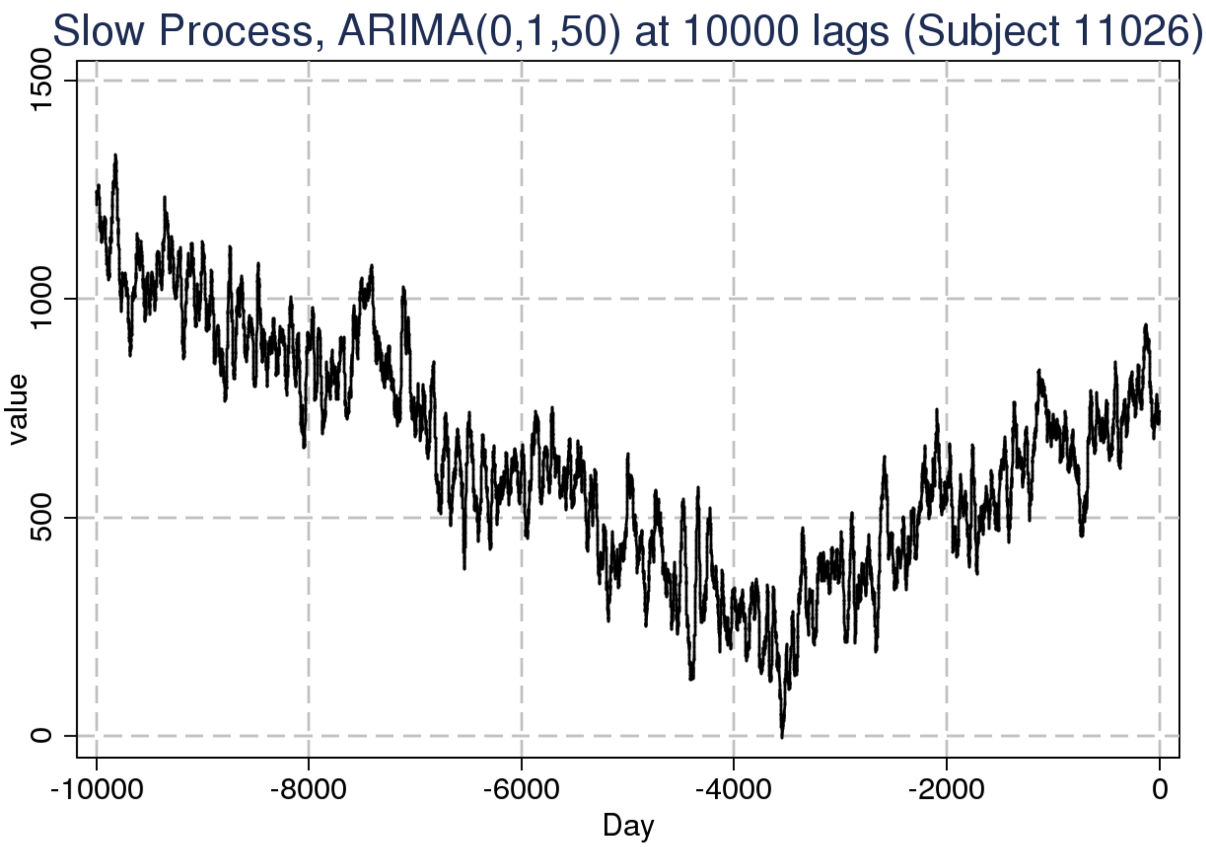
Appendix Figure A5: Sample Plot of Simulated Process



Appendix Figure A6: Sample Plot of Simulated Process



Appendix Figure A7: Sample Plot of Simulated Process



Appendix B: Experimental Instructions

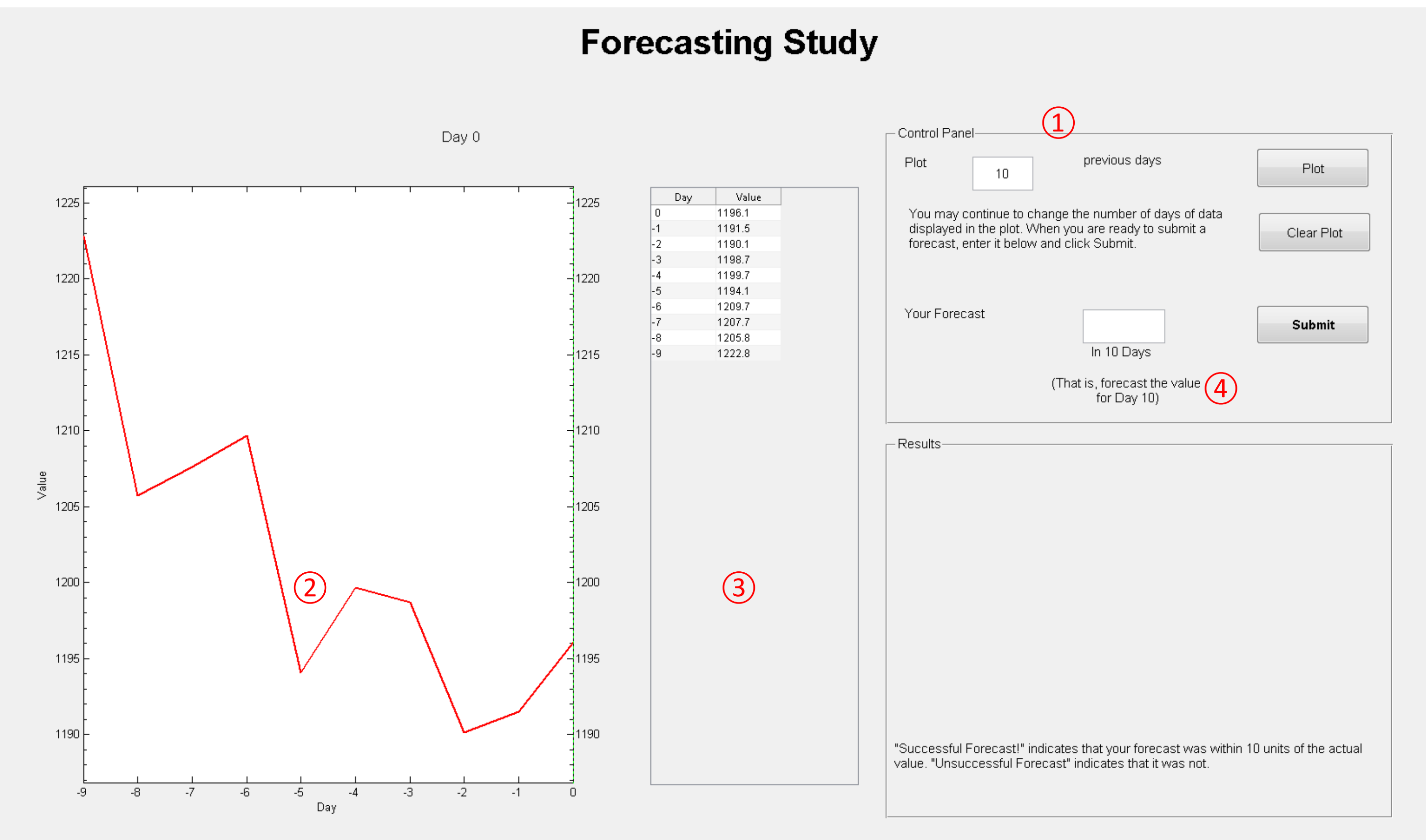
The instructions given to subjects on paper and read aloud at the beginning of each session are shown below.

Instructions

In this study, you will examine a series of numbers, and then forecast what numbers will come later in the series. You will make a forecast on each "day" from Day 0 through Day 59. For each "day" that passes, a new number is appended to the end of the series.

The series of numbers that you will see is determined according to some statistical rules. These rules do not change during the study and are unaffected by the forecasts that you make.

Here is a screenshot of the computer program you will use in this study:



To see previous values in the series, use the Control Panel on the right (labeled ①). If you would like to see a graph of previous days' values, enter the number of previous days you would like to see in the Plot box (②) and then click the Plot button. The values will also appear in a scrollable list (③) next to the graph. You may display anywhere from 1 to 100,000 previous days of data. You may use the Clear Plot button to clear any current plot, although it is not necessary to clear the plot before displaying a new one.

You are welcome to change the number of days of past values you are viewing at any point during the study.

You will be asked to make a forecast for the value of the series on a future day – either 1, 5, 10, 20, 35, or 50 days away (it will be different for each forecast you are asked to make). The text under the box (labeled ④) tells you how many days into the future you are being asked to forecast (as well as what day that is).

You will receive $0.50 for your forecast when that day comes if your forecast is within 10.0 of the true value for that day. You will be asked to make forecasts on Days 0 through 59, but the computer program will also determine the values in the series for days beyond Day 59. If the future day you're forecasting will occur after Day 59, and your forecast is within 10.0 of the true value, you will still receive $0.50 for the forecast.

When you are ready to make your forecast, enter it in the box next to Your Forecast and click the Submit button (or hit Enter on your keyboard). Each time you click Submit, the next value in the series will be revealed in the Results box, along with any previous forecasts relevant to that day and whether or not your forecasts were successful ("Successful Forecast!" means your forecast was within 10.0 of the true value, while "Unsuccessful Forecast" indicates that it was not). If you were at no time asked to provide a forecast for that day, then you will still see the value for that day in addition to text saying you were not asked to provide a forecast for that day's value. If you have chosen to display a graph of previous values, the graph will automatically update to include the new value at the end of each day, although you can change the number of previous days plotted at any time during the study.

Here is an example: If you are currently at Day 14 and the text under the box says "In 10 Days (That is, forecast the value for Day 24)", then you need to submit your forecast for the value of the series on Day 24. Imagine that the series is currently at 2182.2 and you believe that on Day 24 it will be at 2189.5. Then you would enter 2189.5 in the box and click Submit. Once you reach Day 24, you will see the true value of the series on that day. If the true value is between 2179.5 and 2199.5, then your forecast on Day 14 was successful (and you just earned an additional $0.50); if not, your forecast on Day 14 was unsuccessful (and you did not earn an additional $0.50).

At the end of the study, your total payment will be shown on the bottom-right of the screen. Your total payment includes your $10 show-up fee in addition to the amount you made from correct forecasts. When you finish, please come to the examiner station in the front of the room to receive your payment.

If you have any questions about these instructions, please feel free to ask.

Now, please hit the "Begin" button. It may take a few seconds to finish, so please do not hit it more than once.

To get started, type a number of days into the Plot box and click the Plot button, and you will see the previous values graphed and listed. To help familiarize yourself with the series of values, we ask that you plot three different numbers of days of past values before starting to submit forecasts. After doing this, the fields for entering and submitting forecasts will become available.

Appendix C: Complete Protocol

This section of the appendix describes the complete protocol for the experiment.

Subjects were recruited from the subject pool of the Harvard Decision Science Lab, restricting only to current undergraduate and graduate students (from any university).

Before the start of the experiment, the interface was initialized on each computer, and a station number from 1-12 was entered into each computer.[2] The password was also entered and hidden, but the station was not unlocked.[3] After signing informed consent forms, subjects were brought to the computer lab and told to sit down at any computer (and that the choice of computer did not matter). Subjects had a printed set of instructions at their stations, but they were not provided with any other materials and did not have access to any software besides the interface. Instructions were read aloud to the subjects, who were encouraged to ask questions if anything was unclear. While the instructions were being read, a randomly chosen natural number n was entered into the interface by another examiner. The number n identified the data generating process. The Matlab randomization seed was also set to n, so it also determined the realization of the process the subject received. Thus, no two distinct numbers yielded the same process realization. To minimize human error, the number was entered on behalf of the subjects, who were not informed of its purpose.[4]

After going through the instructions, subjects began the forecasting task (outlined in the paper text and in the instructions (Appendix B)). Aside from responding to questions the subjects had, the examiners had no interaction with the subjects during the forecasting, and subjects had no interaction with other subjects. Examiners provided help with task comprehension and the use of the interface as needed, but no advice was given on the actual forecasting, the amount of data to view, etc.

Upon completing the forecasting task, subjects were asked (but not required) to respond to a set of questions within the interface (see Appendix D for the list of questions). After finishing, subjects were paid in cash and signed a receipt verifying receipt of payment. This ended the experiment. Subjects were allowed as much time as desired to finish the task.

---

[2] Corresponding to the 12 computer stations in a room in the Harvard Decision Science Lab. The station number (combined with the subject number at that station) is used to identify subjects.

[3] The password prompt reappears after subjects have finished the experiment and prevents them from accidentally continuing to another session of the experiment. Between sessions, examiners reenter the password into computers that have been used.

[4] Unfortunately, human error was not completely eliminated, as two numbers were repeated while 12 were skipped. The repeated numbers were dropped from the sample. The 12 skipped numbers were not reinserted into the list. Due to an initial programming error in the seed setting, about 60 subjects originally received one of only a few different shock series realizations. These subjects were all discarded from the sample. The error was fixed, and the subjects in the final sample are not affected by it.

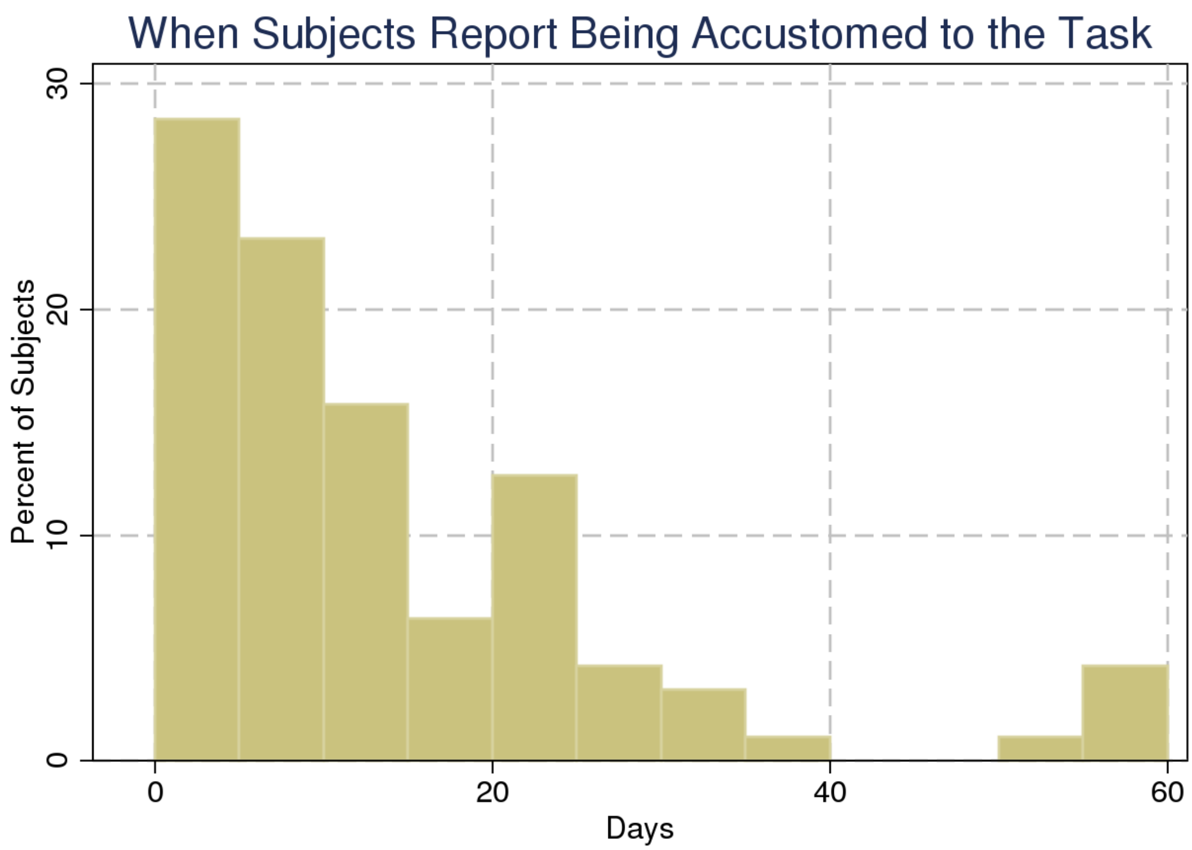
Appendix D: Post-Experiment Questions and Summary Statistics

After completing the forecasting component of the experiment, subjects were asked (but not required) to respond to the following questions:

1) Please provide the following demographic information (enter D or Decline in the corresponding box for any information you decline to provide).
    a. Gender
    b. Age
    c. University
    d. Major/Concentration
2) How much statistical training do you have, on a scale of 1 (no training) to 7 (advanced training)?
3) Were the instructions clear? Did you ever feel uncertain about what you were expected to do? How do you think the instructions could be improved to make the task clearer?
4) Did you have any problems using the interface? If so, what were they? Did it seem to function correctly?
5) Some subjects tell us it took them some time to familiarize themselves with the task and the number series. Around what Day did you feel you were familiar with the task (and series)?
6) How enjoyable did you find this experiment, on a scale from 1 (not enjoyable at all) to 7 (very enjoyable)?

The sample consists of 98 subjects, with 50 receiving the 'fast' ARIMA(0,1,10) process and 48 receiving the 'slow' ARIMA(0,1,50) process. The sample is 47% male and 53% female. The median age is 20, and the average age is 21 (standard deviation 2.5). Harvard affiliates (mostly undergraduate students, who report a variety of different concentrations) make up 87.8% of the sample. Median self-reported statistical training is 2, with an average of 2.47 and a standard deviation of 1.38. Median reported time it took to become familiar with task and series is 7 days, and the mean is 12.48 days. Appendix Figure D1 shows a histogram of when subjects reported becoming familiarized with the task and series. Subjects at 59 reported never being accustomed to the task and series (although it is ambiguous whether they were reporting not being familiar with the task or not understanding the data series).

Appendix Figure D1



Appendix E: Other Processes

Appendix Table E1 contains the coefficients of the other four processes used to generate the data viewed by the other subjects (those not analyzed in this paper) participating in the experiment. Two of these processes are 'bookend' ARIMA(0,1,q) processes, with a constant impulse response function of 1 until period 5 (for the fast process) or 25 (for the slow process), at which point the impulse response function drops to one-fourth of its original magnitude. The other two processes are 'momentum' ARIMA(0,1,q) processes, with impulse response functions that increase gradually to 1.5 times[5] the original shock magnitude over 5 periods (for the fast process) and 25 periods (for the slow process). Shock terms were drawn independently from a N(0,100) distribution. All process realizations were modified by an additive constant that set the "Day 0" value to 742.8.

A total of 198 subjects received one of the four processes below. Their forecasts will be analyzed in detail in future work. As mentioned in the main text, the qualitative results from the bookend processes are consistent with those in this paper, as subjects are more prone to detect mean reversion in the fast version of the process than in the slow one.

Appendix Table E1: Coefficients of Other Processes

MA Term	Bookend (Slow)	Bookend (Fast)	Momentum (Slow)	Momentum (Fast)
1	0	0	0.05	0.19
2	0	0	0.04	0.11
3	0	0	0.04	0.1
4	0	0	0.03	0.05
5	0	-0.75	0.03	0.06
6	0		0.03	
7	0		0.02	
8	0		0.02	
9	0		0.02	
10	0		0.02	
11	0		0.02	
12	0		0.02	
13	0		0.02	
14	0		0.02	
15	0		0.02	
16	0		0.01	
17	0		0.01	
18	0		0.01	
19	0		0.01	
20	0		0.01	
21	0		0.01	
22	0		0.01	
23	0		0.01	
24	0		0.01	
25	-0.75		0.01	

[5] Due to a typo in the code, the fast momentum process actually increased to 1.51 times its original magnitude.

Appendix F: Lower Order Model Fits

Appendix Tables F1-F2 provide some characteristics of the models used to generate the model forecasts for $0 \leq q \leq q^*$. All coefficients for models $0 < q < q^*$ were estimated from a series of 100,000 data points, and the same coefficients were used for each subject. In each table, column 1 shows the degree of the ARIMA(0,1,q) model (0 corresponds to random walk, while 10 (50) corresponds to the true process for the fast (slow) process). Column 2 shows the maximum value attained in the impulse response function for that model. Column 3 shows the long-run asymptote of the impulse response function for that model. For example, the model of the true process has a maximum of 1.5 and an asymptote of 0.5, while the random walk model has a maximum of 1.0 and an asymptote of 1.0. Full coefficient lists can be found in Excel files available from the authors.

It is important to note that adding additional MA terms to the ARIMA(0,1,q) models has a non-monotonic effect on the implied long-run persistence of an impulse. As a consequence, moving from q to q+1 does *not* necessarily improve forecast performance for horizons more than one period out in our setting.

Appendix Table F1: Fast Process Model Descriptions

	Impulse Response Function	
q	*max*	*asymptote*
0	1.000	1.000
1	1.281	1.281
2	1.475	1.475
3	1.627	1.627
4	1.721	1.721
5	1.859	1.859
6	1.929	1.929
7	1.629	1.494
8	1.356	0.702
9	1.468	0.532
10	1.500	0.500

Appendix Table F2: Slow Process Model Descriptions

	Impulse Response Function			Impulse Response Function	
q	*max*	*asymptote*	*q*	*max*	*asymptote*
0	1.000	1.000	26	1.956	1.956
1	1.095	1.095	27	1.952	1.951
2	1.166	1.166	28	1.957	1.957
3	1.233	1.233	29	1.935	1.928
4	1.291	1.291	30	1.918	1.905
5	1.339	1.339	31	1.877	1.844
6	1.387	1.387	32	1.832	1.774
7	1.440	1.440	33	1.759	1.655
8	1.482	1.482	34	1.687	1.530
9	1.527	1.527	35	1.594	1.344
10	1.558	1.558	36	1.513	1.162
11	1.597	1.597	37	1.460	1.021
12	1.628	1.628	38	1.417	0.883
13	1.650	1.650	39	1.396	0.781
14	1.674	1.674	40	1.392	0.687
15	1.693	1.693	41	1.403	0.625
16	1.723	1.723	42	1.422	0.584
17	1.745	1.745	43	1.448	0.549
18	1.764	1.764	44	1.473	0.526
19	1.792	1.792	45	1.492	0.512
20	1.798	1.798	46	1.501	0.506
21	1.823	1.823	47	1.506	0.503
22	1.855	1.855	48	1.514	0.498
23	1.876	1.876	49	1.513	0.499
24	1.929	1.929	50	1.500	0.500
25	1.952	1.952			

Appendix G: Additional Analyses

I) Subject Performance

Appendix Table G1 shows the median absolute deviations of the subjects' forecasts from the true model (i.e., the 'rational' forecast), for each forecast horizon separately and also when pooling all horizons ('All'):

Appendix Table G1

	All	1	5	10	20	35	50
Fast (ARIMA(0,1,10))	13.173	4.170	13.878	14.446	15.717	18.138	21.673
Slow (ARIMA(0,1,50))	18.930	2.646	10.880	20.933	27.706	36.845	36.220

To facilitate the interpretation of these numbers, Appendix Table G2 shows for each process and horizon separately the expected absolute change in the series under the true model, $E\left(\left|x_{t+\tau}^{ARIMA(0,1,q^*)} - x_t\right|\right)$, and also the expected absolute deviation of realized values from the true-model forecast that arises due to unforecastable noise, $E\left(\left|x_{t+\tau} - x_{t+\tau}^{ARIMA(0,1,q^*)}\right|\right)$

Appendix Table G2

		1	5	10	20	35	50
Fast	Expected absolute change	4.428	15.290	18.639	18.639	18.639	18.639
	Expected absolute deviation from E(X)	7.977	23.712	30.575	33.070	36.510	39.641
Slow	Expected absolute change	2.008	9.536	17.665	29.595	39.131	41.367
	Expected absolute deviation from E(X)	7.977	19.571	30.255	47.592	63.798	67.746

II) Goodness of Fit of the Assigned Models

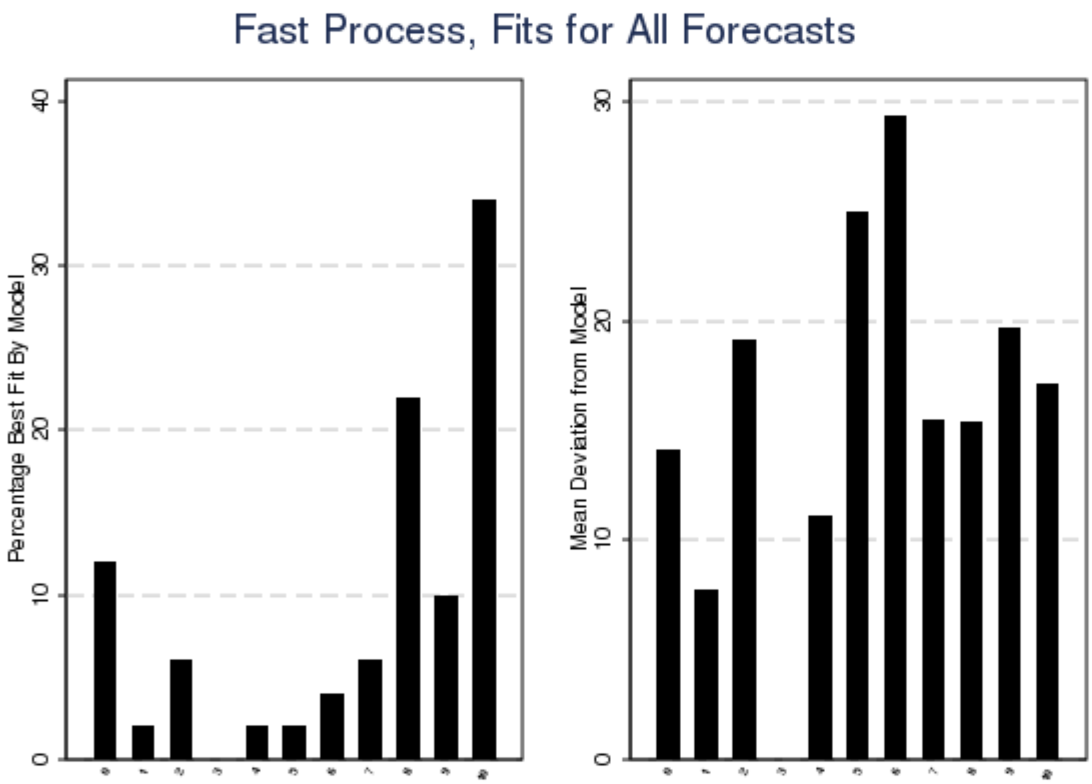
Appendix Figures G3-G8 expand upon Figures 2 and 3 from the paper by also showing the mean absolute deviation of subject forecasts from the models that were the best individual fits. Furthermore, we break up forecasts into short (1, 5, 10 periods) and long (20, 35, 50 periods) horizon forecasts, and we fit models to subjects separately for short and long horizons.

Focus first on the 'fast' process and long horizon forecasts (Appendix Figure G5), which we are most interested in. The left panel shows that a high number of subjects are assigned to models of relatively high order when looking at only long horizon forecasts. Indeed, the distribution is very similar to the distribution of model fits that use forecasts from all horizons (Figure 2 in the main text; left panel of Appendix Figure G3). The right panel shows that for those subjects who are assigned models of order 8 or higher, the model fit is relatively good.

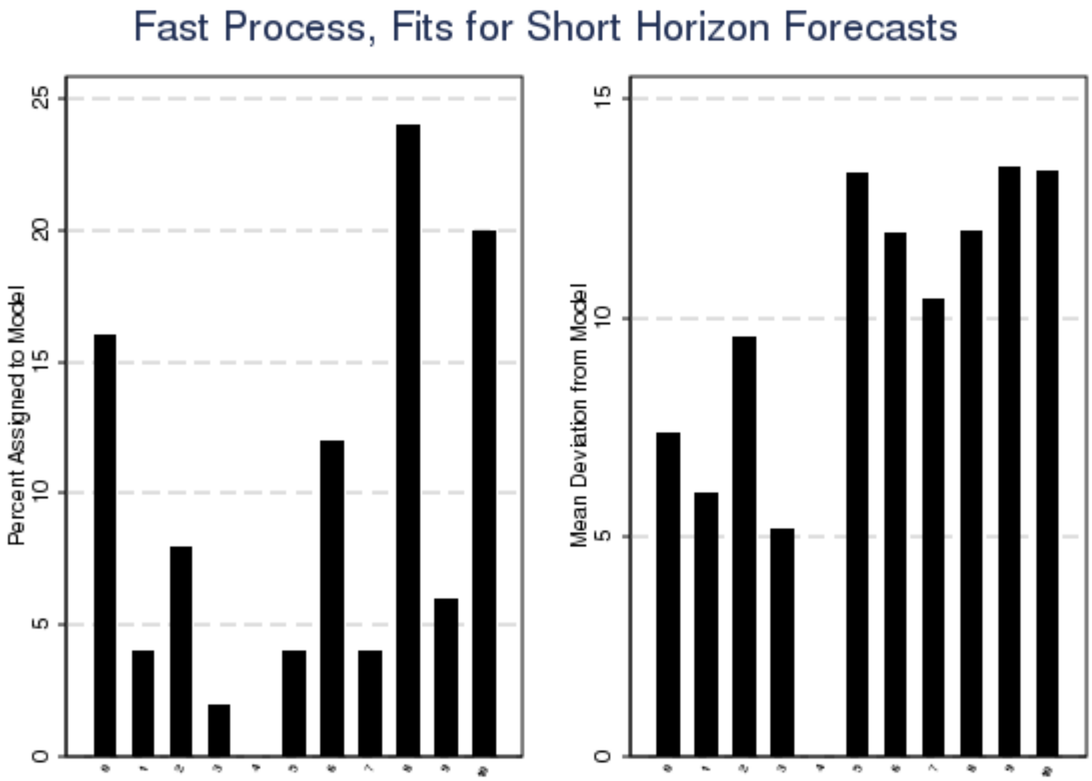
For the 'slow' process and long horizon forecasts (Appendix Figure G8), we see that the mean deviation for subjects who get assigned the random walk (ARIMA(0,1,0)) model is relatively low, while for those who get assigned high-order models, the deviations of their forecasts from the assigned modes are more substantial. This evidence suggests that for at least some of the subjects who got assigned high-order models, the assignment may have happened by chance, further strengthening the conclusions from the main text.

Appendix Tables G9-G10 attempt to look at the goodness of fit of the models that subjects were assigned to, relative to the benchmarks of the random walk and the true process. This analysis is important for understanding how well the assigned models are identified. Subjects are grouped by best-fitting model, and for each group the tables show the mean absolute deviation of the subjects' forecasts from those of the best-fitting model (column 2), from those of the random walk process (column 3), and from those of the true model (column 4). Note that model 0 is the random walk. Model 10 (50) is the true model for the fast (slow) process. Overall, these tables suggest that there are usually substantive differences in how well different models fit a subject's forecasts.

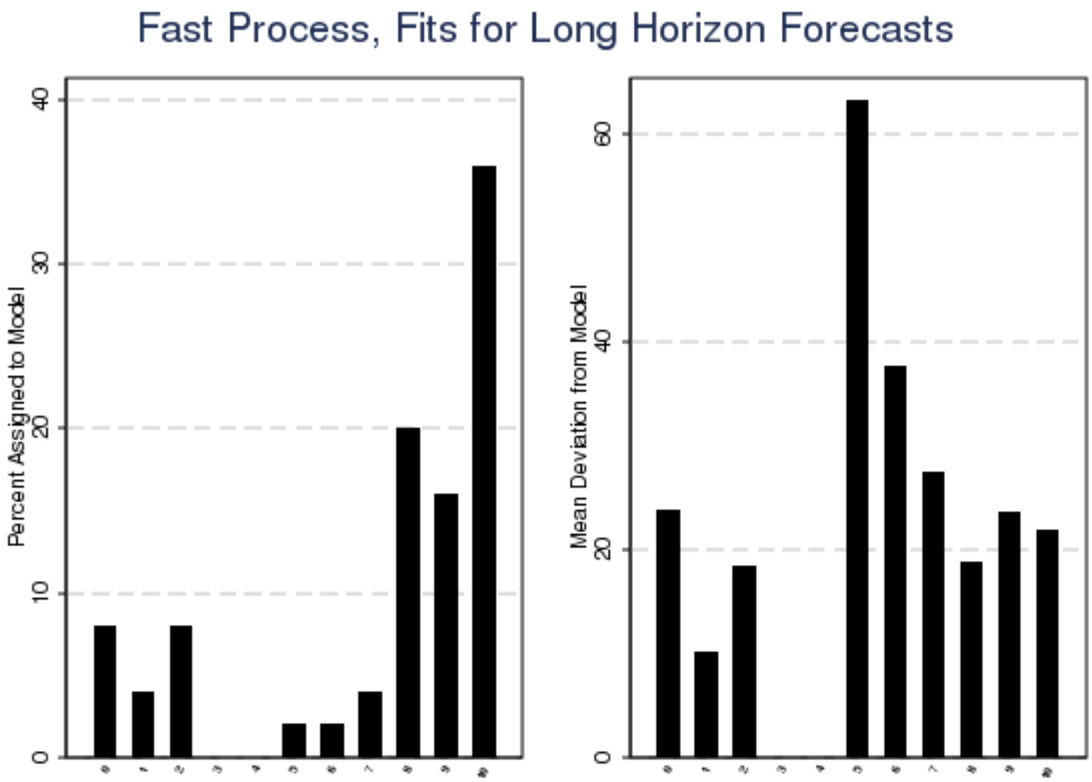
Appendix Figure G3



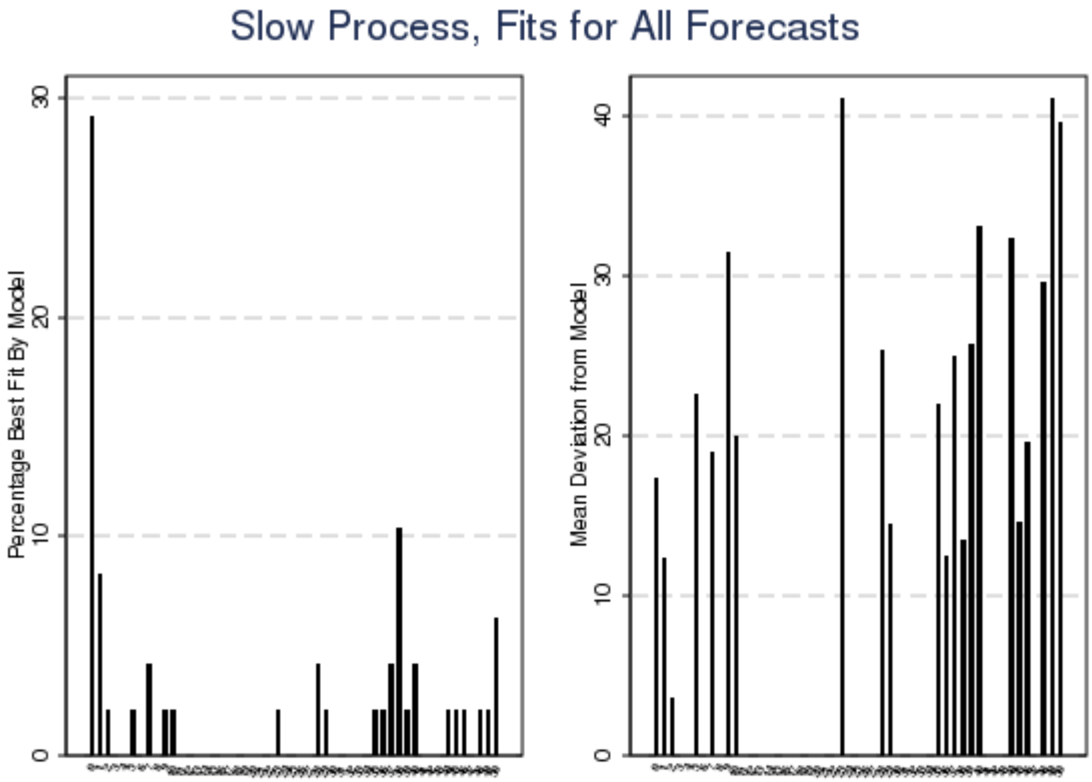
Appendix Figure G4



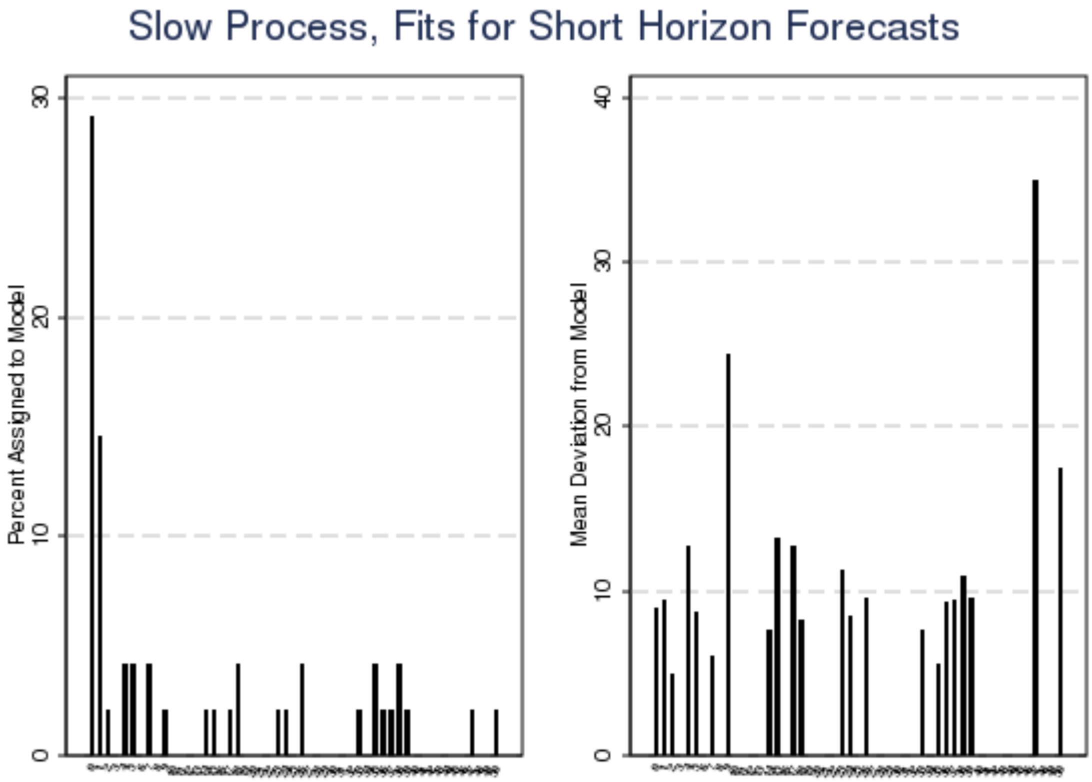
Appendix Figure G5



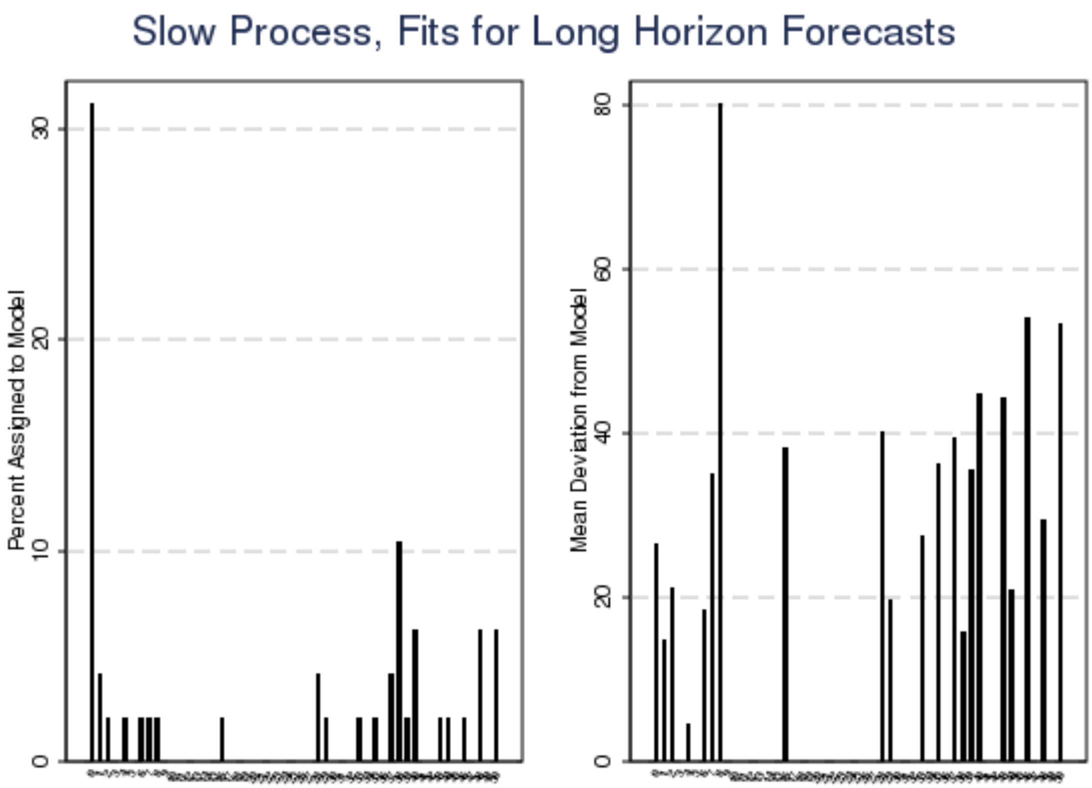
Appendix Figure G6



Appendix Figure G7



Appendix Figure G8



Appendix Table G9: Relative Goodness of Fit for Fast Process

Model	# Subjects Fit to Model	Mean Absolute Deviation from Fitted Model	Mean Absolute Deviation from Random Walk	Mean Absolute Deviation from True Model
0	6	14.092	14.092	21.465
1	1	7.738	8.547	12.215
2	3	19.164	19.833	24.619
4	1	11.100	15.176	19.611
5	1	25.004	27.152	33.782
6	2	29.331	31.498	36.275
7	3	15.499	16.765	17.471
8	11	15.387	18.174	16.417
9	5	19.690	24.136	19.887
10	17	17.123	22.874	17.123

Appendix Table G10: Relative Goodness of Fit for Slow Process

Model	# Subjects Fit to Model	Mean Absolute Deviation from Fitted Model	Mean Absolute Deviation from Random Walk	Mean Absolute Deviation from True Model
0	14	17.353	17.353	30.165
1	4	12.332	12.443	21.484
2	1	3.555	3.841	11.171
5	1	22.682	23.056	40.355
7	2	19.050	20.115	35.130
9	1	31.520	31.917	65.293
10	1	20.022	22.349	31.427
23	1	41.159	42.207	53.543
28	2	25.457	33.258	43.423
29	1	14.528	19.835	38.085
35	1	21.986	24.857	27.359
36	1	12.536	15.408	15.995
37	2	25.046	27.646	29.504
38	5	13.449	17.364	19.010
39	1	25.809	27.991	29.449
40	2	33.209	42.582	35.750
44	1	32.456	35.444	32.675
45	1	14.658	17.425	14.690
46	1	19.602	32.854	19.739
48	1	29.613	32.142	29.716
49	1	41.129	55.142	41.134
50	3	39.636	49.941	39.636

III) Reduced Form Analysis

Appendix Table G11 shows the full results from the median regression of form $\hat{x}_{i,t} - c_{i,t} = \alpha + \beta\left(x_{i,t}^{RE} - c_{i,t}\right) + \eta_{i,t}$, discussed in the text:

Appendix Table G11[6]

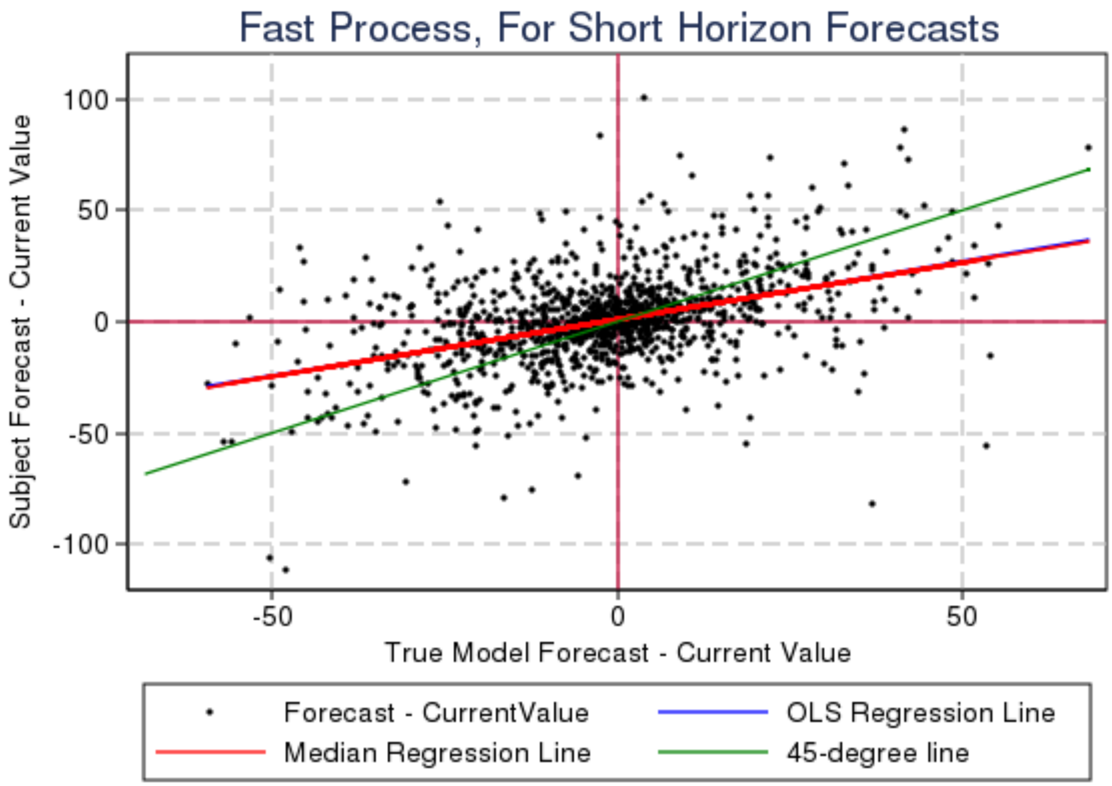
	Fast Process Subject Forecast - Current Value	Slow Process Subject Forecast - Current Value
True Model Forecast - Current Value	0.600***	0.0895*
	(0.0313)	(0.0386)
Constant	1.894***	0.905**
	(0.372)	(0.296)
N	2475	2376

We extend this analysis by looking at different forecasting horizons. The following plots (Appendix Figures G12-G15) present scatter plots by process and by grouped forecasting horizons. Each process has two separate plots, one for short forecasting horizons (1, 5, 10) and one for long forecasting horizons (20, 35, 50). The scatter plots put the difference between the true model forecast and the current value on the horizontal axis and the difference between the subject's forecast and the current value on the vertical axis. In addition to the scatter plot, the figures show an ordinary least squares regression line, a median regression line, and a 45-degree line.

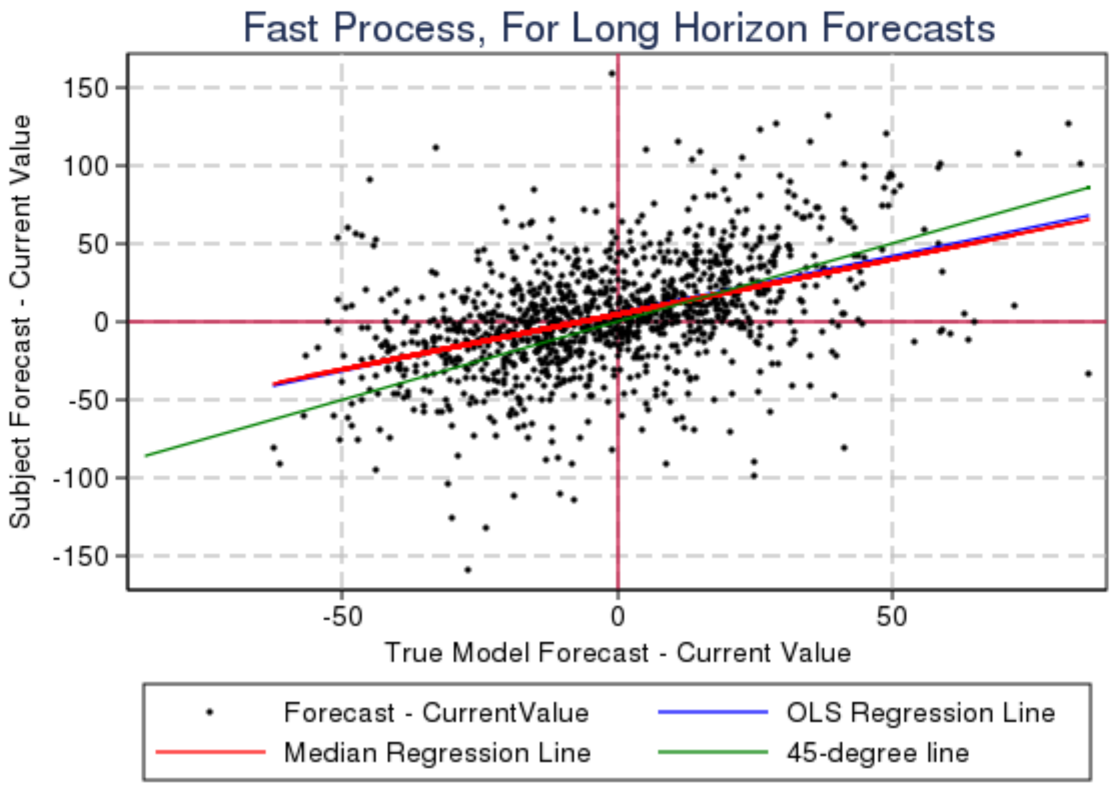
These regression lines are much closer to the 45-degree line (the rational benchmark) for the fast process than for the slow process, at both short and long horizons.

[6] Standard errors are obtained by bootstrapping with 1,000 repititions.

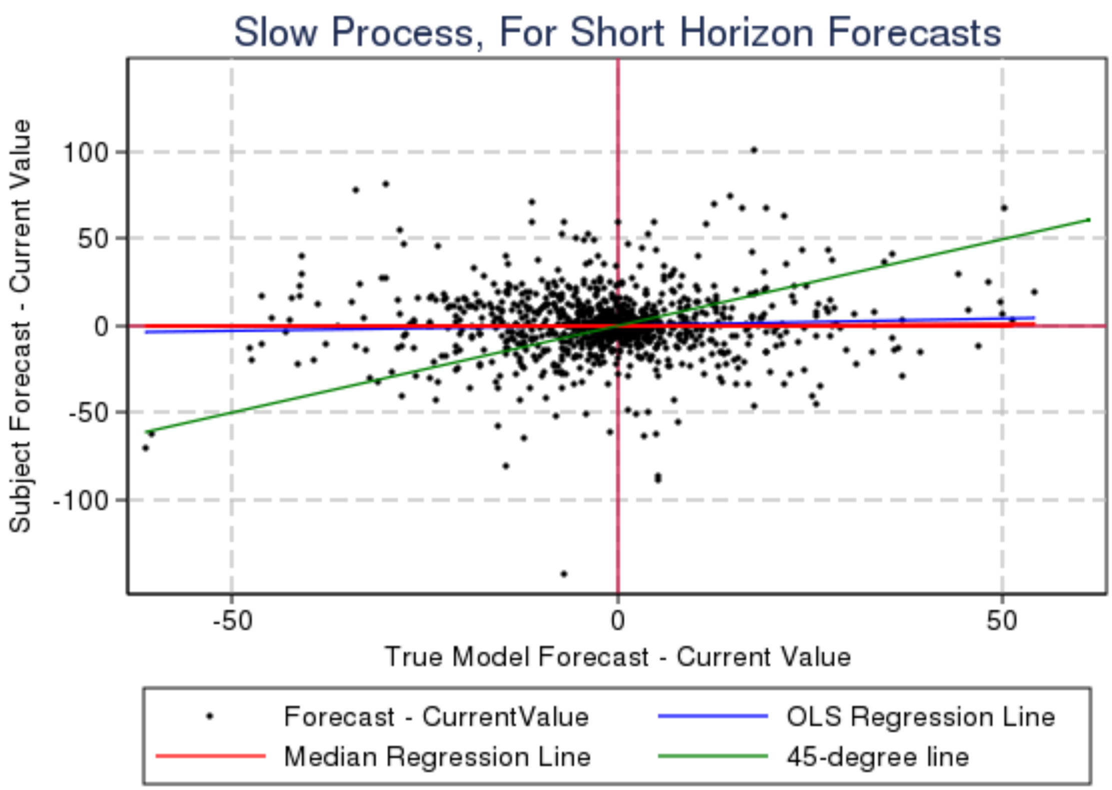
Appendix Figure G12



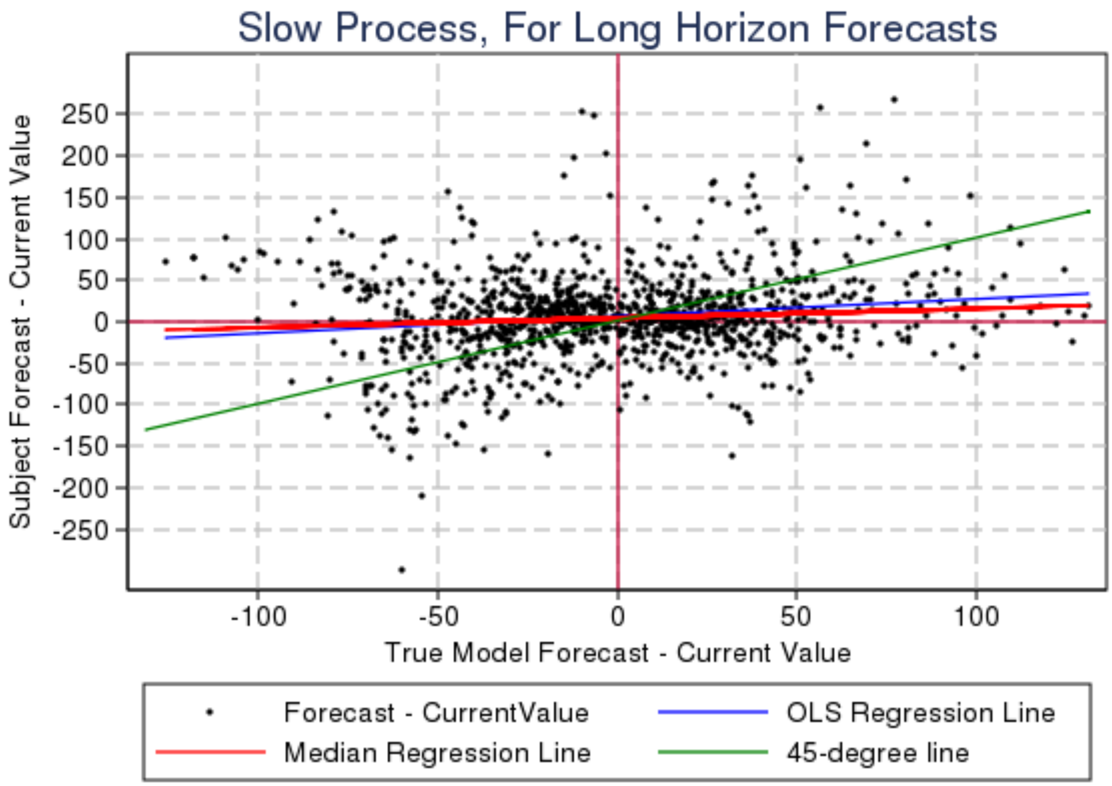
Appendix Figure G13



Appendix Figure G14



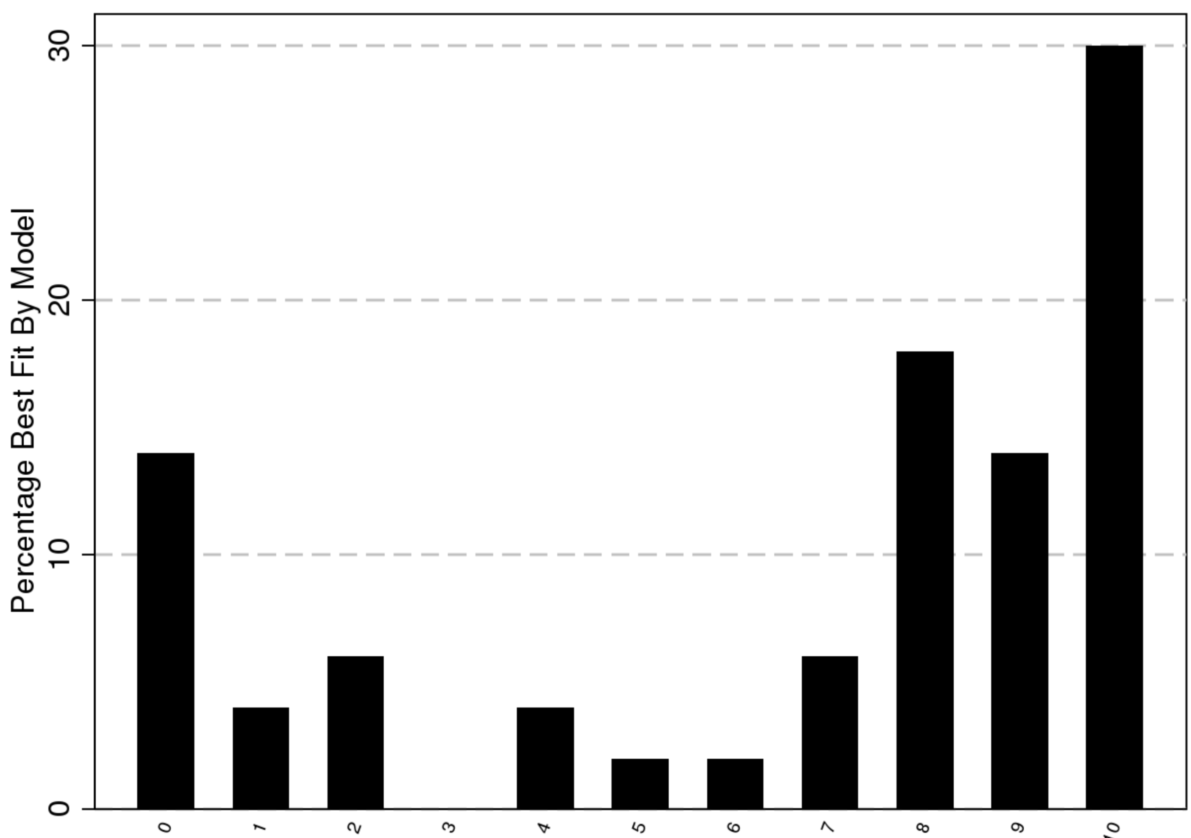
Appendix Figure G15



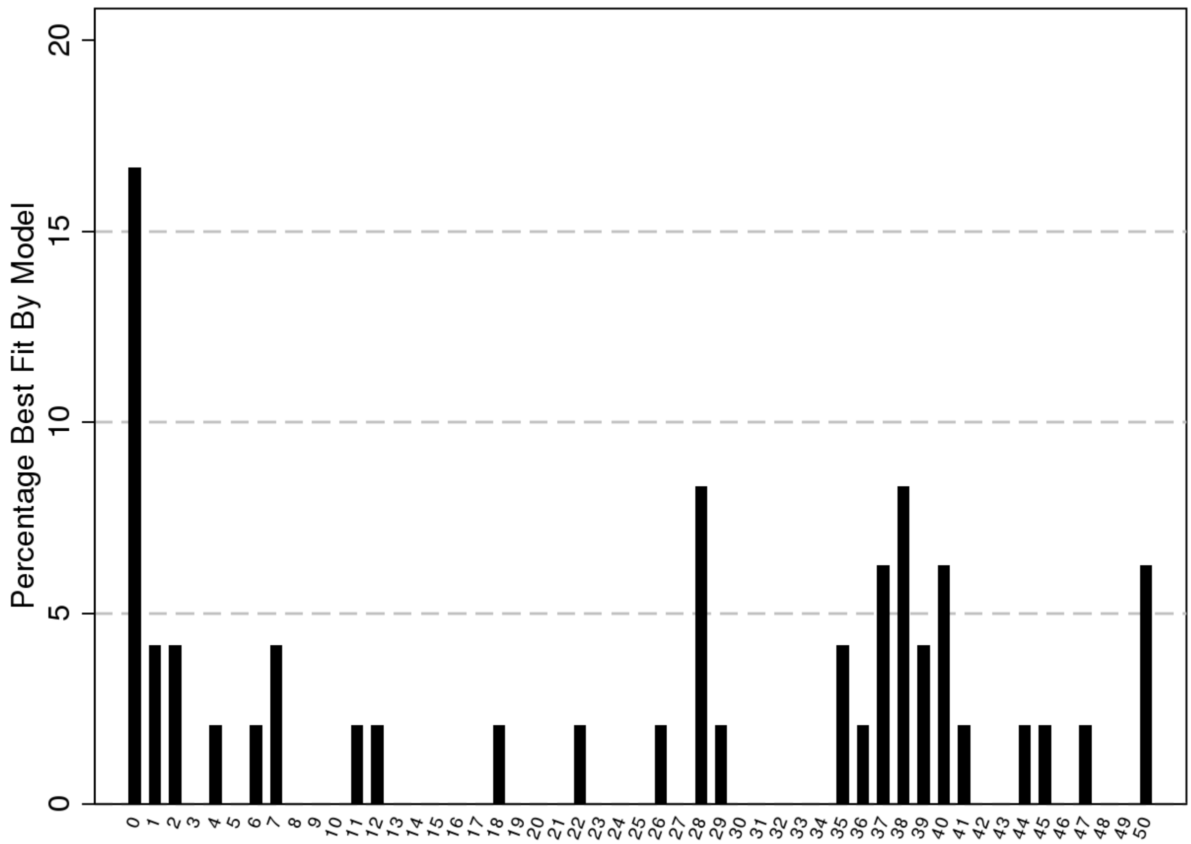
IV) Robustness Checks

In this section, we compare our main results, which are based on a sample that discards forecasts made in the first 10 days and discards the 1% of forecasts that are furthest from the realized value, with the results of: 1) the same analysis using the sample without any forecasts discarded; and 2) the same analysis with 1% outliers discarded and with a subject-by-subject rule for discarding forecasts made in the first days of the experiment (based on when a subject reported feeling accustomed to the task and series – see Appendix Figure D1). Figures 2 and 3 from the main text are recreated in each case (Appendix Figures G16-G19). The median regression results for each case are reported in Appendix Table G20. Perceived mean reversion as a fraction of actual mean reversion is unchanged from the results in the main text, except for the perceived mean reversion for the slow process when no forecasts are discarded, which becomes -2.1%). The main results are robust to these different sample definitions.

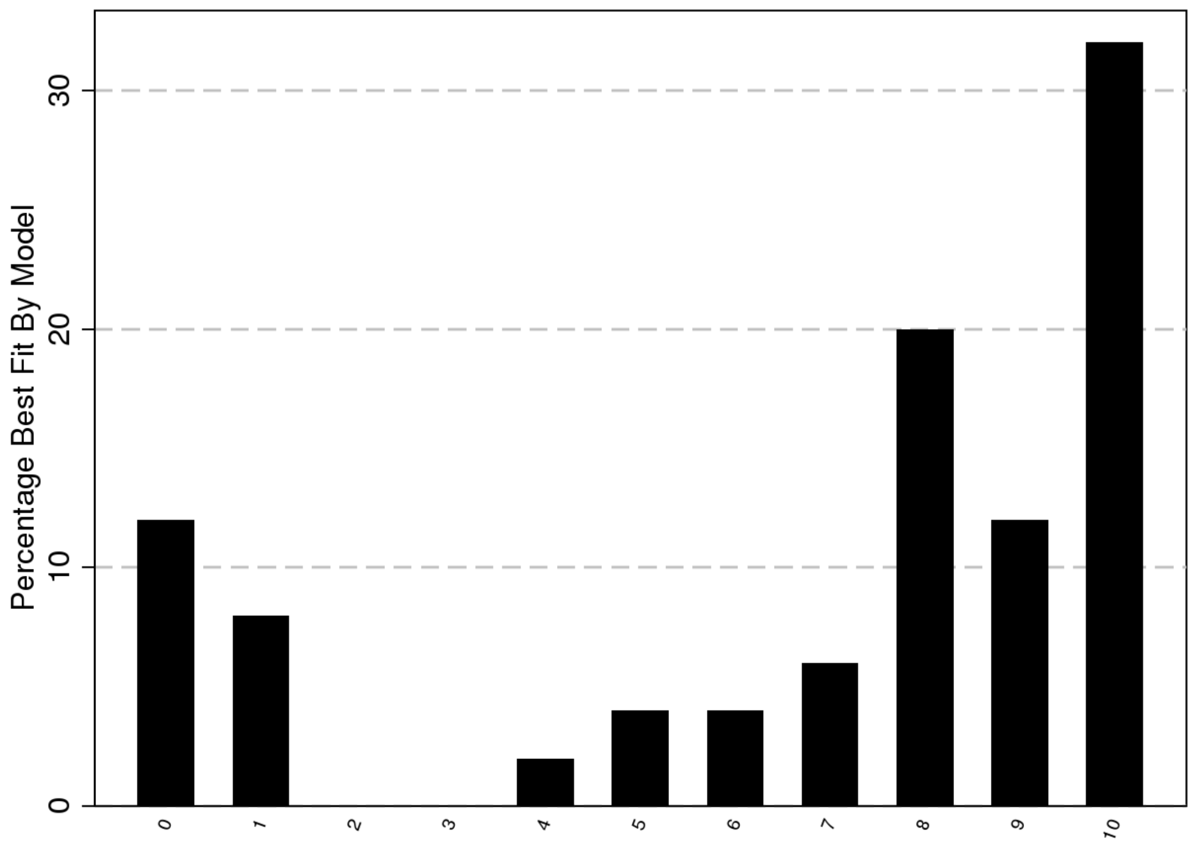
Appendix Figure G16: Fast Process Model Assignments, No Forecasts Discarded



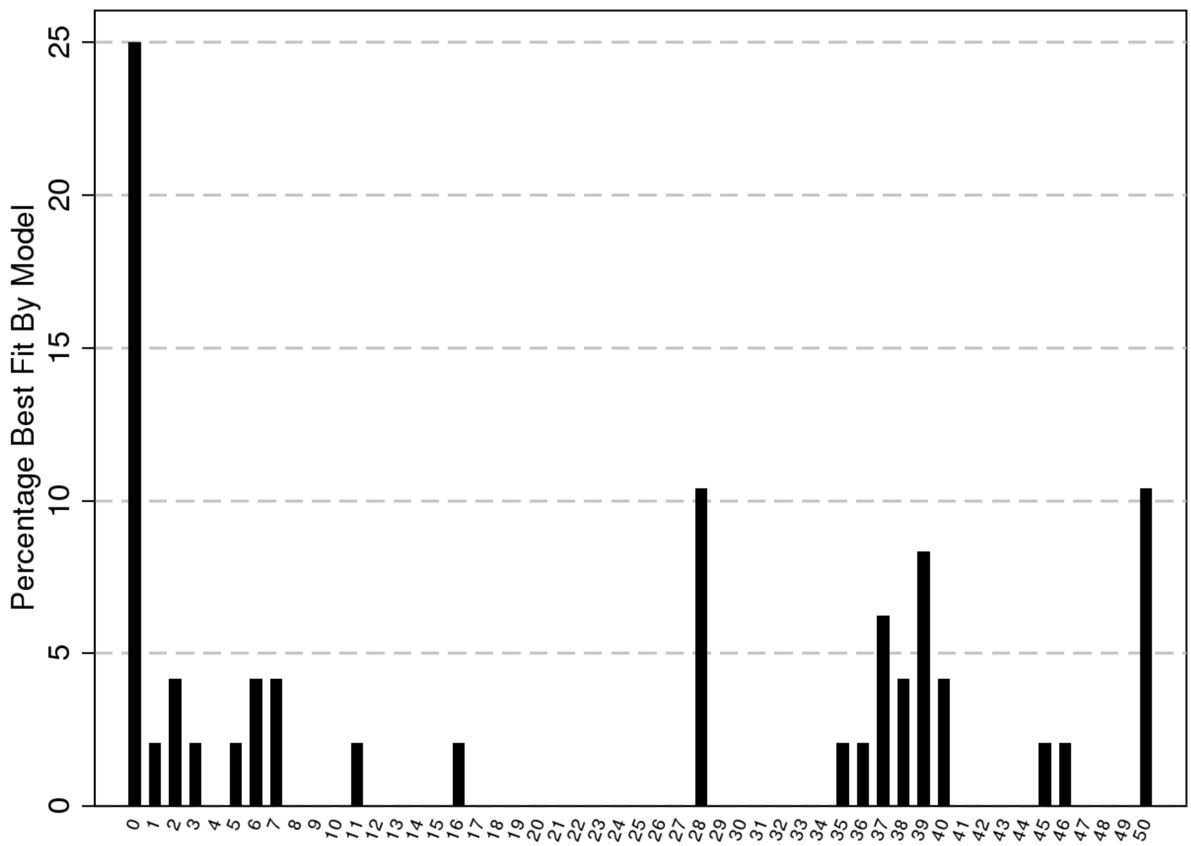
Appendix Figure G17: Slow Process Model Assignments, No Forecasts Discarded



Appendix Figure G18: Fast Process Model Assignments, 1% Furthest Absolute Deviations Discarded, Subject-by-Subject Rule for Discarding First Forecasts Applied



Appendix Figure G19: Fast Process Model Assignments, 1% Furthest Absolute Deviations Discarded, Subject-by-Subject Rule for Discarding First Forecasts Applied



Appendix Table G20[7]

	No Forecasts Discarded		1% Furthest Absolute Deviations Discarded, Subject-by-Subject Rule for Discarding First Forecasts Applied	
	Fast process	Slow process	Fast process	Slow process
True Model Forecast - Current Value	0.529***	0.0720*	0.664***	0.103*
	(0.0331)	(0.0350)	(0.0352)	(0.0409)
Constant	1.616***	0.912**	1.764***	0.697*
	(0.348)	(0.301)	(0.392)	(0.297)
N	3000	2880	2292	2356

[7] The dependent variable for all regressions is: Subject Forecast – Current Value. Standard errors are obtained from bootstrapping with 1,000 repititions.