Chapter 35

LEARNING TO FORECAST RATIONALLY

HUGH KELLEY and DANIEL FRIEDMAN

Economics Department, University of Santa Cruz, Santa Cruz, CA 95064, USA

1. Introduction

Economists routinely assume that all participants in the economy are rational forecasters who can correctly incorporate all available information when they form expectations of price and other variables that matter to them. Economists such as Marcet and Sargent (1989) point out that, when assessing the relevance of rational expectations models, researchers must ask whether or not repeat experience allows people to closely approximate rational forecasts. If people exhibit systematic departures from rational forecasts, then much of economic theory needs reconstruction.

The empirical literature on forecast rationality is surprisingly thin. Surveys of consumers, professional economists, and other market participants generally find that forecast errors have a non-zero mean, are correlated with other observable information, and follow an adaptive process (Camerer, 1995, pp. 609–611). Laboratory experiments with discrete forecasting tasks often indicate persistent biases (e.g., Grether, 1990). The most relevant previous experiment, Williams (1987, pp. 1–18) finds autocorrelated and adaptive errors when laboratory market participants forecast next period's market price. Other ties to existing literature can be found in Kelley (1998) and Kitzis et al. (1997).

The experiment described in this chapter isolates the forecasting process in two different stochastic individual choice tasks. The first task is based on Roll (1984), who finds that even in a very simple field financial market (Florida Orange Juice futures) where only two news variables are relevant (Florida weather hazard and competing supply, mainly from Brazil), the news can only account for a small fraction of the price variability. The second task is a variant of psychologists' standard discrete Medical Diagnosis task, e.g., Gluck and Bower (1988).

2. The Tasks

2.1. Orange Juice Forecasting (OJ)

In each trial, a subject views two continuous variables on her monitor: x_1 (called weather hazard) and x_2 (called Brazil supply), as in the upper-left corner of Figure 1. The values of x_1 and x_2 are independent random draws from the uniform distribution on (0, 100). The task is to forecast the dependent variable y (called orange juice futures price). The

Handbook of Experimental Economics Results, Volume 1

DOI: 10.1016/S1574-0722(07)00035-2

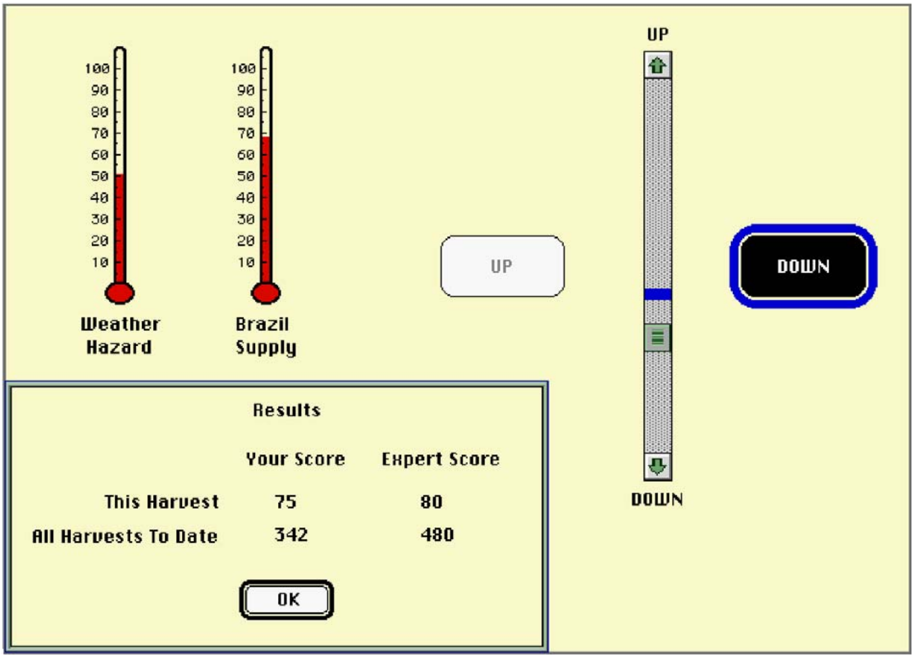


Figure 1. The values of x_1 (Weather Hazard) and x_2 (Brazil Supply) are random draws from the uniform distribution on (0, 100). The subject uses the slide bar on the right side of the display to enter her forecast. She uses the mouse to slide the small box up or down to enter her forecast. Then the realized value of the price appears on the same slide bar as a blue (dark) bar. Finally, in some treatments a score box is provided in the lower left of the screen. Score calculations are described in the text.

subject uses the slide bar on the right side of the monitor display to enter the forecast. The realized value of y appears on the same vertical display as the slide bar, as indicated in Figure 1. The realized value on trial t is

$$y_t = a_o + a_1 x_{1t} + a_2 x_{2t} + v e_t, \tag{1}$$

where the coefficients a_1 and a_2 are unknown to the subject (implicitly – they are the objects of learning) with baseline values 0.4 and -0.4, v is the noise amplitude (typically $v = 10$), e_t is an independent random variable drawn each trial from the uniform distribution on $(-1, 1)$, and a_o is the intercept used to center the data so that y_t falls at the middle of the vertical scale when $e_t = 0$ and $x_{1t} = x_{2t} = 50$.

In one treatment (called History), before entering her forecast the subject can view a summary of price outcomes from previous trials where values of x_1 and x_2 are close to current values. In another treatment (called Score), at the end of each trial the monitor displays a score S as shown at the bottom of Figure 1, computed from the continuous forecast c and the actual price p according to the quadratic scoring rule $S(c, p) =$

$80 - 280(c - p)^2$. [The "expert score" in Figure 1 uses the forecast c obtained from Equation (1) with e_t set to 0.]

Sessions last 480 trials, and so far 57 subjects have been tested. Most subjects' cumulative scores fall between 36,000 and 37,000, with a theoretical maximum score of $480 * 80 = 38{,}400$. Additional treatments include high noise ($v = 20$ instead of 10), asymmetric weights (e.g., $(a_1, a_2) = (0.24, -0.56)$ instead of $(0.40, -0.40)$) and structural breaks (e.g., weights shift from symmetric to asymmetric in trial 241). See Kelley (1998) for a more complete description of the task and treatments.

2.2. The Medical Diagnosis Task (MD)

The medical diagnosis task also is a stochastic individual choice task with 480 trials for each subject, two independent variables (the symptoms) and one dependent variable (the disease). The user interface is quite similar to that in the OJ task. However, the independent variables (temperature and blood pressure) are discrete with four possible values (high, medium high, medium low and low), and the dependent variable is binary (the disease is either Autochus or Burlosis). Conditional on the realized disease, the symptoms are independently drawn according to likelihoods unknown to the subject.

Subjects' continuous response c_t in trial t consists of naming the disease deemed more likely and indicating (with the slide bar) the degree of confidence. The response is coded as a continuous variable between 0 (completely confident that the disease is B) and 1 (completely confident that the disease is A). Kitzis et al. (1997) show that the true Bayesian relationship between symptoms and diseases can be very closely approximated by the linear equation

$$y_t = a_o + a_1 x_{1t} + a_2 x_{2t}, \tag{2}$$

where now y_t is the posterior log odds of disease A over disease B, and x_{it} is the discrete variable with values 1, 0.3, -0.3 and -1 respectively for high, medium high, medium low and low values of symptom $i = 1, 2$. The unknown coefficients a_1 and a_2 again are the implicit objects of learning; the true values are 1.39 and -2.30.

We tested 123 subjects in a 2×3 factorial design with the treatments History (vs No History) and Score (vs No Score and vs Score + Pay) with 20+ subjects in each cell.

3. Results

3.1. Rolling Regressions

Although a subject may think of the task in various idiosyncratic ways, the analyst can summarize the subject's beliefs by seeing how he responds to the current stimuli x_{it}. Moreover, the analyst can summarize the learning process by seeing how the subject's response to stimuli changes with experience.

Given Equations (1) and (2), learning thus can be seen in the changes over time in a subject's implicit subjective values of the coefficients a_1 and a_2. The data analysis reconstructs the implicit values of these coefficients and tracks their changes over time.

The reconstruction proceeds as follows. For the OJ task, take the subject's actual forecast c_t in trial t as the dependent variable, and take the actual values of x_{it} as independent variables. Then, run a rolling regression of c_t on the two independent variables over a moving window of 160 consecutive trials, incrementing the last trial T from 160 to 480. The procedure in the MD task is the same except that the dependent variable is the log odds of the continuous choice, $L(c_t) = \ln[(c_t + .01)/1.01 - c_t)]$; c_t is shifted by .01 away from 0 and 1 to avoid taking the log of zero. The intercept coefficient is constrained to its objective value in the results shown below in order to reduce clutter and to improve statistical efficiency.

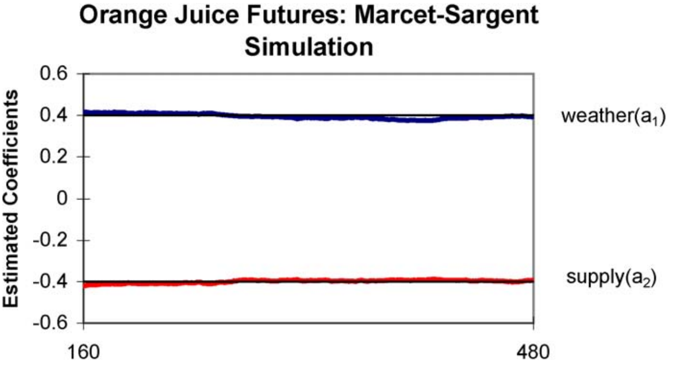
Effective learning is indicated by rapid convergence of the coefficient estimates a_{iT} (as T increases) to the objective values a_i. Obstacles to learning are suggested by slow convergence, convergence to some other value, or divergence of the coefficient estimates. This empirical approach embodies some of the theoretical ideas on learning in Marcet and Sargent (1989) as explained in Kelley and Friedman (2002).

3.2. OJ Learning Curves

Figure 2 presents two examples. Top panel shows the simulated performance of a Marcet–Sargent econometrician who uses realized prices for all trials observed so far to estimate the coefficients a_1 and a_2 and then uses these coefficients in Equation (1) with $e_t = 0$ to forecast the current price. Learning seems immediate (within 160 trials). The R^2 for the first 160 trial window of data was 0.93 and ended at the same level, 0.93, for the last window.

Bottom panel shows that the actual subject who earned the top score came fairly close to the Marcet–Sargent ideal. The coefficient estimates indicate that he slightly overresponded to current symptoms throughout the session, but the overresponse was negligible by the last 160 trials. His R^2 for the first 160 trial window of data was 0.94 and increased to 0.96 by the last window. This high scoring subject is fairly representative; coefficient estimates for other subjects in most treatments sometimes indicate overresponse and sometimes underresponse, but on average are quite close to or slightly beyond objective values.

Table 1 summarizes the main departures from effective learning detected so far. Coefficient estimates indicating "Significant" under and overresponse by the end of the session ($T = 480$) are about equal in the baseline and asymmetric treatments, but underresponse is much more prevalent than overresponse in the structural break treatments. In the high noise treatment (amplitude $v = 20$ instead of $v = 10$), overresponse is much more common than underresponse. Figure 3 presents the corresponding histograms, which clearly show that coefficient estimates for subjects in the high noise environment tend strongly toward overresponse.



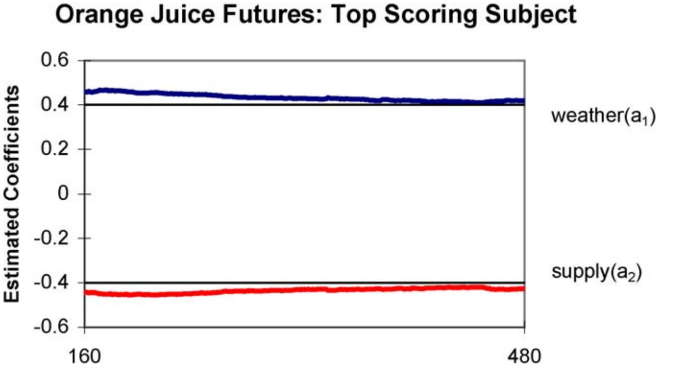


Figure 2. (Top) Coefficient estimates from rolling regression $y_t = a_1 x_{1t} + a_2 x_{2t} + v e_t$ on trials $\{T-159, T-158, \ldots, T\}$ for $T = 160$ to 480. The equation is estimated using the Marcet–Sargent Model to generate the forecasts y_t. (Bottom) Coefficient estimates from rolling regression $c_t = a_1 x_{1t} + a_2 x_{2t} + v e_t$ on trials $\{T-159, T-158, \ldots, T\}$ for $T = 160$ to 480. The equation is estimated using forecasts c_t from Subject 44.

Table 1
Over and under-response in Orange Juice forecasting

	Under response	Objective	Over response
Symmetric weights	20	11	21
Asymmetric weights	10	1	9
High noise	5	3	12
Structural break	13	2	7

Note: Coefficients a_1 and a_2 are estimated at $T = 480$ for all Ss for the equation $c_t = a_1 x_{1t} + a_2 x_{2t}$. Responses for each subject are classified as over or underresponse if the estimate differs from the objective value by more than $1.96 *$ std error.

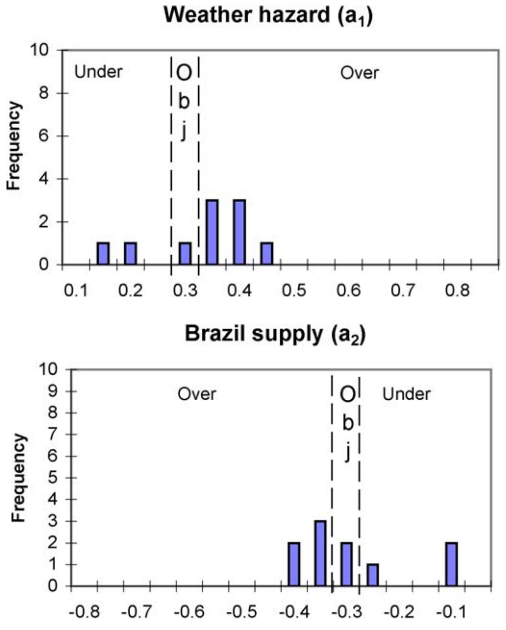


Figure 3. Distribution of final ($T = 480$) coefficient estimates in high noise treatment for the Orange Juice Futures experiment. "Obj" indicates estimates near objective values of (0.33, −0.33) respectively. "Under" (and "Over") refer to cases where the absolute value of the estimate is less (and more) than the objective value.

3.3. MD Learning Curves

Figure 4 provides two examples from the second experiment, Medical Diagnosis. Top panel simulates a Bayesian econometrician (the MD counterpart of Marcet–Sargent) who uses realized disease outcomes for all trials observed so far to estimate the coefficients a_1 and a_2 and then uses these coefficients in Equation (2) with $e_t = 0$ to predict the current disease. Ideal learning is a bit slower and more erratic than in the OJ task, but it still converges to the true values $a_1 = 1.39$ and $a_2 = -2.3$ quite rapidly.

Bottom panel shows that the actual subject who earned the top score differs noticeably from the Bayesian ideal. The coefficient estimates for this subject indicate persistent overresponse and are quite representative of the subject pool. The histograms in Figure 5 confirm that overresponse is indeed the prevailing bias in our MD data.

4. Discussion

Kelley (1998) reports several robustness checks. OJ specifications designed to capture prior beliefs and non-linear responses to news detected some transient effects in many subjects, but for the most part the final regression is indistinguishable from the basic

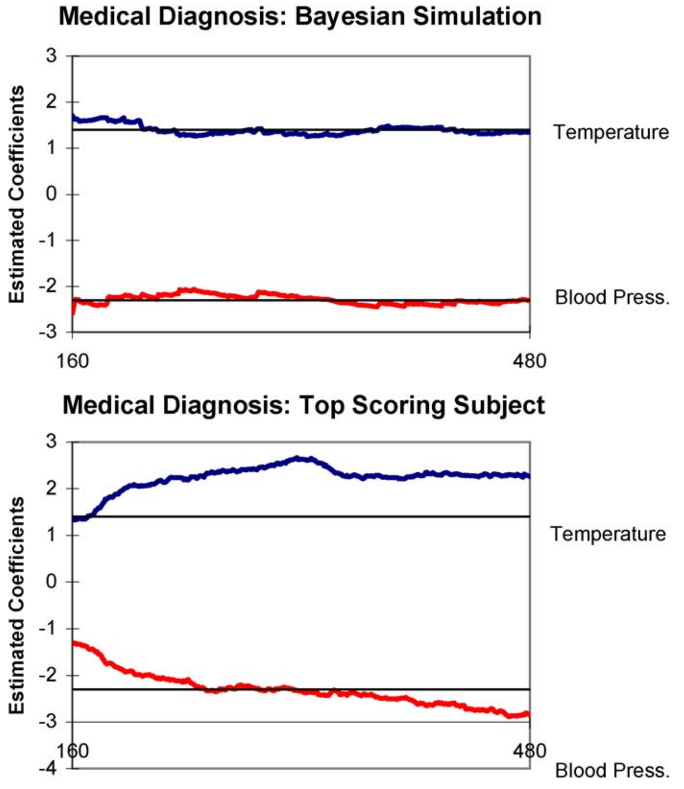


Figure 4. (Top) Coefficient estimates from rolling regression $y_t = a_1 x_{1t} + a_2 x_{2t}$ on trials $\{T - 159, T - 158, \ldots, T\}$ for $T = 160$ to 480. The equation is estimated using the Bayesian model to generate the forecasts y_t. (Bottom) Coefficient estimates from rolling regression $c_t = a_1 x_{1t} + a_2 x_{2t}$ on trials $\{T - 159, T - 158, \ldots, T\}$ for $T = 160$ to 480. The equation is estimated using the forecasts ct from Subject 28.

specification presented above. Eight parameter MD specifications that allow separate learning for each level of each symptom also converged roughly to the basic specification presented above, but we detect a general bias towards overresponding to the more informative symptom levels and underresponding to the less informative symptom levels.

We draw three conclusions from the data analysis. First, the rationality assumption is a good first approximation to subjects' forecasts at the end of 480 learning trials. Second, systematic biases towards under or overresponse can be detected in specific circumstances, e.g., overresponse in the noisier OJ environment. Third, more experiments are needed in a wider variety of tasks and environments in order to understand more fully when people can learn to forecast rationally. We anticipate that the rolling regressions and learning curves featured in this chapter will continue to be a useful tool in that research.

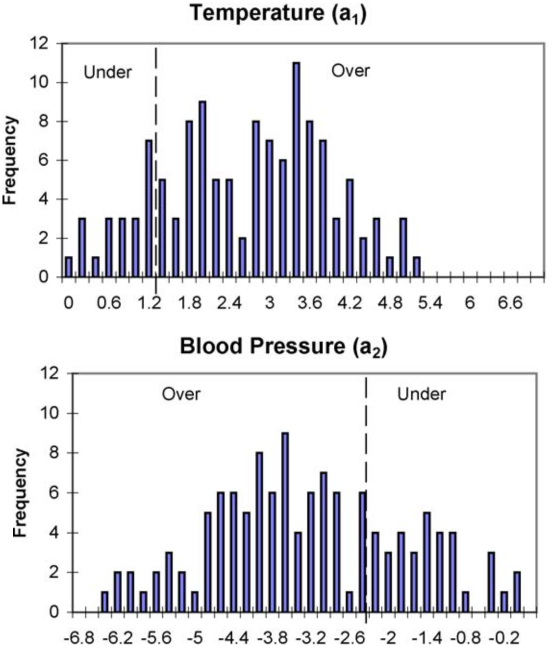


Figure 5. Distribution of final ($T = 480$) coefficient estimates for the Medical Diagnosis experiment. Objective values indicated by vertical dashed line. Objective values are (1.39) for Temperature and (−2.3) for Blood Pressure.

References

Camerer, C. (1995). "Individual decision making". In: Kagel, J.H., Roth, A.E. (Eds.), The Handbook of Experimental Economics. Princeton University Press, Princeton, NJ, pp. 587–703.

Gluck, M.A., Bower, G.H. (1988). "From conditioning to category learning: An adaptive network model". Journal of Experimental Psychology: General 117, 225–244.

Grether, D.M. (1990). "Testing Bayes rule and the representativeness heuristic: Some experimental evidence". Journal of Economic Behavior and Organization 17, 31–57.

Kelley, H. (1998). "Bounded rationality in the individual choice experiment". Unpublished thesis, Economics Department, University of California, Santa Cruz.

Kelley, H., Friedman, D. (2002). "Learning to forecast price". Economic Inquiry 40, 556–573.

Kitzis, S., Kelley, H., Berg, E., Massaro, D., Friedman, D. (1997). "Broadening the tests of learning models". Journal of Mathematical Psychology 42, 327–355.

Marcet, A., Sargent, T. (1989). "Convergence of least squares learning mechanisms in self referential linear stochastic models". Journal of Economic Theory 48, 337–368.

Roll, R. (1984). "Orange juice and weather". American Economic Review 74, 861–880.

Williams, A.W. (1987). "The formation of price forecasts in experimental markets". Journal of Money, Credit, and Banking 19, 1–18.