Modeling Expectation Formation Involving Several Sources of Information

Otwin Becker, University of Heidelberg
Johannes Leitner*, University of Graz
Ulrike Leopold-Wildburger, University of Graz

Abstract: Experimental studies of expectation formation of subjects are predominantly limited to the prediction of one single time series despite the practical relevance of expectations in situations with multiple sources of information. In this paper we report on an experiment in which subjects are given time series (indicators) as additional information for the judgmental forecast of a stationary time series. The quality and the number of these indicators are varied in three versions of a forecasting experiment. We explore the effects on forecasting accuracy and we test the average forecasts of the subjects for consistency with the rational expectations hypothesis. A simple heuristic is presented that explains the average forecasting behavior better than the rational expectations if indicators are presented to the subjects. It is demonstrated by a simulation study that this result is representative for the considered stationary stochastic processes.

JEL-classification: C91, C92, C53.
Keywords: Expectation formation, forecasting, experimental economics, rational expectations, heuristics.

*Department of Statistics and Operations Research, Karl-Franzens-University of Graz, Universitätsstraße 15/E3, 8010 Graz, Austria, Tel: +43(316)380-7245, Fax: +43(316)380-9560, email: johannesleitner@gmx.de

Acknowledgements

The authors would like to thank Marlies Ahlert, Friedl Bolle, Tony Hall, Helmut Lütkepohl and Reinhard Selten and two anonymous referees for helpful comments on draft versions of this paper.

The support of the Austrian Science Foundation (FWF) for the project *"Heuristics for the Subjective Prediction of Time Series"* P17156-N12 is gratefully acknowledged.

Modeling Expectation Formation Involving Several Sources of Information

1 Introduction

Expectation formation of economic agents is an important issue in economic theory and its implications for economic behavior are reflected by the large number of researchers who have analyzed forecasts in order to test hypotheses about expectation formation. Although most forecasts are made by individuals, many economically relevant forecasts are aggregated measurements, i.e. statistically combined individual forecasts.

The Delphi method was one of the first conjectural forecasting techniques. Delphi is an multi round (several times repeated) forecasting procedure based on independent inputs of selected experts (see Sackman, 1974). It was developed in the 1950's at the Rand corporation in order to forecast future developments in areas of lacking theoretical background and/or insufficient data for the application of quantitative forecasting models. In such situations aggregated judgment provides valuable information.

Other examples of such procedures comprise the annual economic surveys by Livingston and the publication of consensus forecasts of financial analysts or economists for future earnings of corporations or macroeconomic data. Consensus forecasts are a popular information and research shows that this popularity is justified. By combining individual forecasts accuracy can be increased (see e.g. Clemen, 1989; Granger, 1989).

We are interested in explaining such average forecasts. In consideration of the practical relevance this is a well-founded intention but it is impossible to explain empirically generated forecasts. The information sets and forecasting methods used by the participants remain unknown. In the laboratory, the information set of the subjects can be controlled and varied systematically. The scientist knows the data generating process on which the expectations are based. This is an advantage when expectations hypotheses are tested, e.g. the rational expectations values can be calculated exactly. In this paper we present such a laboratory experiment focused on average forecasting behavior.

The first experiments on expectation formation were independently carried out by Fisher (1962) and Becker (1967). Their experimental setting is still relevant for most of the succeeding studies: A time series describing the development of an economic variable is presented to subjects in the laboratory. At the beginning of each period the subjects have to forecast the next value of the time series. The time series is not influenced by the individuals' forecasts and its past values serve as the only utilizable information for the forecasts in each period. To mention only a few authors, Schmalensee (1976), Mason (1987), Bolle (1988), Dwyer et al. (1993), Hey (1994), Becker and Bolle (1996) and Becker and Leopold-Wildburger (1996, 2000) tested hypotheses on expectation formation. The results of these studies can generally be interpreted as rejecting the rational expectations hypothesis. However, Dwyer et al. and Mason find support for rational expectations. In their experiments, the time series to be predicted was a pure random walk, while in the others autoregressive or empirical time series were applied.

These studies are unexceptionally limited to the judgmental forecast of a single time series without any additional information. In reality, forecasters consider several sources of information. This fact is taken into account in our experiment: subjects forecast an unlabelled stationary time series (called base series) judgmentally with two other, partially contradictory and differently reliable, time series (indicators) to support their decision. The indicators have a stable lead period of one and the subjects are restricted to forecast the base series by pure chart-reading all time series.

The presentation of additional information in a time series forecasting task has only been studied by a few authors. In Lim and O'Connor (1996) subjects were given low-correlated and high-correlated causal information for the judgmental forecast of sales figures. Subjects were able to differentiate the reliability of causal information and to improve their forecasting accuracy by using this information. However, they used this information inefficiently. Similar results were also found in more recent studies by Sanders (1997) and Goodwin and Fildes (1999). In the context of testing expectation hypotheses, so far only Garner (1982) and Brennscheidt (1993) reported on experiments with more than a single time series. In Garner's experiment, subjects had to predict the price of a fictitious agricultural commodity. The price was determined by an autoregressive model with three exogenous variables. In each period t the subjects had to forecast the endogenous variable for periods $t + 1$ and $t + 2$. The information set of the subjects consisted of the three exogenous variables up to period $t + 1$ and the history of the endogenous variable up to period t. Garner had to reject the hypothesis of rational expectations for almost all individuals and the average forecasts. Brennscheidt tested various hypotheses of expectation formation on a large sample of data. He experimented with 14 different settings, of which all included exogenous variables. He found biases in individual and group forecasts which caused the rejection of the rational expectations hypothesis.

Our experimental design strongly follows and Becker and Leopold-Wildburger (1996, 2000), who report on an experiment without any indicators. The authors present the bounds&likelihood heuristic as a simple model for the explanation of average forecasts.

The paper at hand is concerned with three major issues: First, we extend the bounds&likelihood heuristic for the modeling of the average forecasts when indicators are available to the subjects. We test whether the rationale of this simple model is better able to explain the average behavior of the subjects than the rational expectations hypothesis. We also test the model-consistent learning behavior of the subjects. Second, we analyze the effects of providing additional information on the quality of the forecasts. It could be hypothesised that the accuracy decreases when subjects are confronted with too much information. We will answer this question by comparing the results with the experiment without indicators. Third, as a normative issue, we will analyze the performance of both models on forecasting the time series

itself. For this purpose, we will present the results of a simulation study to check for the efficiency of the bounds&likelihood heuristic. In the experiment only one realisation of artificially generated time series is used. The simulation study illustrates that the experimental results are not only caused by knowingly chosen series but that they are representative for the stochastic process considered.
The paper is organized as follows: In the next section we describe the experimental procedure. We present the bounds&likelihood heuristic and the rational expectations hypothesis in sections 3 and 4. We address the three raised major issues in section 5.

2 The experimental setting

The subjects' task was the prediction of a time series x_t judgmentally, i.e. predicting the value of the next period by observing the past observations without any help from statistical or econometrical models. To support their forecast the subjects were given time series (y_t and z_t) as indicators. These gave a certain insight into the future value of x_t. The time series x_t (hereafter called "base series", the indicators are called "leading series") is a realization of the stochastic difference equation

$$x_t = x_{t-1} - int(\frac{1}{2} \cdot x_{t-2}) + u_t \tag{1}$$

with $x_1 = 7, x_2 = 12$, the endogenous variable x_t and the white noise u_t. The uniformly distributed variable u_t represents the realisations drawn from a six-sided dice, i.e. $u_t \in \{1, 2, 3, 4, 5, 6\}$ and $f_u(u) = \frac{1}{6}$. The integer function int ensures that all values of the base series and the indicators are integer. The two leading series were generated as follows:

$$y_t = x_t + 15 + v_t \tag{2}$$

$$z_t = x_t + 15 + w_t \tag{3}$$

The random variables v_t and w_t are discretely, triangularly distributed, i.e. $v_t \in \{-3, -2, \ldots, 3\}$ with $f_v(v) = \frac{1}{41}(11 - 3|v_t|)$ and $w_t \in \{-5, -4, \ldots, 5\}$ with $f_w(w) = \frac{1}{36}(6 - |w_t|)$ for all v_t and w_t respectively. This characteristic of the random variables makes the leading series y_t more reliable. Without the white noise, the leading series would equal the base series plus 15 units. All time series are stationary.
The experiment was carried out with paper and pencil. The subjects drew a graph of the values of the base series and the two leading series manually into the same graph and connected them period by period. The subjects were told in the instructions that the leading series give them a certain insight into the change of the base series but no contextual information was given to the subjects and the time series were unlabelled in order to avoid a priori biases of forecasts (see Chapman and Chapman, 1969). All subjects were presented the same realisations drawn from (1), (2) and (3).

Figure 1 shows these three time series.

Insert Figure 1 about here.

The number and the quality of the indicators was varied between three experimental versions:

Version 1: with the base series x_{t+1} to be forecasted in period t and the leading indicator y_{t+1}, i.e. information set $\Omega^1_{t+1} = \{x_t, x_{t-1}, \ldots, x_1; y_{t+1}, y_t, \ldots, y_1\}$.

Version 2: with the base series x_{t+1} to be forecasted in period t and the leading indicator z_{t+1}, i.e. information set $\Omega^2_{t+1} = \{x_t, x_{t-1}, \ldots, x_1; z_{t+1}, z_t, \ldots, z_1\}$.

Version 3: with the base series x_{t+1} to be forecasted in period t and the leading indicators y_{t+1}, z_{t+1}, i.e. information set $\Omega^3_{t+1} = \{x_t, x_{t-1}, \ldots, x_1; y_{t+1}, y_t, \ldots, y_1; z_{t+1}, z_t, \ldots, z_1\}$.

The participants did not see a history of past values at the beginning of the experiment. In period 1 of version 3 (and analogously in versions 1 and 2), y_1 and z_1 were given to the subjects and they were asked to make their first forecast. After they completed their forecast, they were told the value of x_1. Then y_2 and z_2 were presented and the subjects were asked for their forecast for the second period. This sequence was repeated for all 42 periods.
The subjects were paid for their forecasts with the function $p_t = 20 \cdot max\{3 - |x_t - f_t|; 0\}$, i.e. in each period they received 60 cents for an exact prediction and 40 (20) cents for a deviation of one (two) unit(s). The average payment was 6.3 Euros for a duration of about 50 minutes. Altogether 232 subjects participated in the experiment: 150 in version 1, 62 in version 2 and 30 in version 3.

3 The bounds&likelihood heuristic

The bounds&likelihood heuristic (b&l heuristic) was developed by Becker and Leopold-Wildburger (1996, 2000, in press) for the modeling of average forecasts. The calculation of the b&l heuristic is based on the same information that is available to the subjects, namely the time series' gestalt characteristics. One predictor is calculated for every time series and all predictors are aggregated to a combined forecast. In our version 1 and 2 (3) two (three) predictors are calculated, one for the base series and one for each leading series. In principle, the heuristic can be extended to any number of leading series in order to incorporate further information. Below, the procedure is described for the information set in version 3 but analogous calculations are performed for the other versions of the experiment.
For the calculation of the predictors, the base series x_t and the indicators y_{t+1} and z_{t+1} are distinguished. It is assumed that two characteristics of the base series are

essential for the forecasts: the average variation and the turning points. The average absolute variation of the time series $b_t = \frac{1}{t-1}\sum_{j=2}^{t}|x_j - x_{j-1}|$ are the bounds for the predicted change based on the actual time series value x_t. The maximum predicted change is supposed to vary between the interval [-b_t, b_t]. The actually predicted change depends on the likelihood that x_t is a turning point. For $x_t > x_{t-1}$, an upswing case, $l_{t(peak)}$ is the probability that x_t is a local maximum. If in period t all local maxima are above x_t, the probability that x_t is a turning point is low. If all local maxima are below x_t, i.e. x_t is the highest time series value, it is very likely to be a turning point. For a downswing case ($x_t < x_{t-1}$), the local minima are considered and $l_{t(trough)}$ is calculated. Both likelihoods are defined in (4).

$$
\begin{aligned}
l_{t(peak)} &= \frac{1 + \text{number of local maxima} \leq x_t}{2 + \text{number of local maxima}} \\
l_{t(trough)} &= \frac{1 + \text{number of local minima} \geq x_t}{2 + \text{number of local minima}}
\end{aligned}
\tag{4}
$$

In the case of no change ($x_t = x_{t-1}$) it is assumed that the upswing and downswing have the same probabilities and both cases are combined linearly. At a high level of the time series, subjects will forecast a downswing; at a low level, an upswing. Based on these assumptions, the predictor of the base series $f^x_{t,b\&l}$ is defined in (5):

$$
f^x_{t+1,b\&l} = \begin{cases} x_t + b_t(1 - 2l_{t(peak)}) & for\ x_t > x_{t-1} \\ x_t + b_t(l_{t(trough)} - l_{t(peak)}) & for\ x_t = x_{t-1} \\ x_t - b_t(1 - 2l_{t(trough)}) & for\ x_t < x_{t-1} \end{cases}
\tag{5}
$$

The stationarity of the base series is crucial for the heuristic. In trended time series the effect of local extrema would have to be considered differently. In our case the distribution of turning points are considered as estimators of the turning point probabilities and the bounds are taken as average changes. This can be presumed for the collective of the subjects and makes $f^x_{b\&l}$ an estimator for the rational expectations. In addition to $f^x_{t,b\&l}$, two predictors are generated from the leading series, $f^y_{t,b\&l}$ and $f^z_{t,b\&l}$, that are finally combined to the forecast $f_{t,b\&l}$. The calculation of the leading series predictors is based on the fact that their values for $t+1$ are available. The subjects observe the change of an indicator, e.g. $y_{t+1} - y_t$, and have to relate it to the base series in order to make a forecast of x_{t+1}. This change has a certain dimension (absolute value) and direction (sign). For the calculation of $f^y_{t,b\&l}$ it is assumed that the absolute value is adapted and the sign is unchanged. Generally, a modification is necessary because the base series and the leading series have different dimensions and average variations. In order to relate smaller (or larger) average variations of the base series to the indicator, we define the relation factor r_t. With a linear relation and a leading period of one, the predictors for the leading series can be described as follows in (6).

$$f^s_{t+1,b\&l} = x_t + r^s_t(s_{t+1} - s_t) \qquad \text{with } r^s_t = \frac{\sum_{i=2}^t |x_i - x_{i-1}|}{\sum_{i=2}^t |s_i - s_{i-1}|}, \text{ for } s = y, z. \tag{6}$$

With (5) and (6), three predictors are available. If these values differ from each other a solution has to be found that gives more weight to the predictor with the lower prediction error so far. The predictors are weighted with α_t reciprocally proportional to their absolute prediction error D_t for each time series, given in (7):

$$D^s_t = \sum_{i=2}^t |f^s_{i,t,b\&l} - x_i| \qquad \text{for } s = x, y, z \tag{7}$$

These are calculated on ex post basis which is indicated by the parameter i in $f^s_{i,t,b\&l}$. The calculation of the weights is described in equation (8):

$$\alpha^s_t = \frac{\frac{1}{D^s_t}}{\frac{1}{D^x_t} + \frac{1}{D^y_t} + \frac{1}{D^z_t}} \qquad \text{for } s = x, y, z \tag{8}$$

Definition (8) implies that the relationships of the weighting factors are

$$\frac{\alpha^x_t}{\alpha^y_t} = \frac{D^y_t}{D^x_t}, \quad \frac{\alpha^x_t}{\alpha^z_t} = \frac{D^z_t}{D^x_t}, \text{ and } \quad \frac{\alpha^y_t}{\alpha^z_t} = \frac{D^z_t}{D^y_t} \tag{9}$$

Using these terms the combined forecast, $f_{t+1,b\&l}$, of the b&l heuristic is generated. The forecast values of the next period consist of a linear combination of all predictors:

$$f_{t+1,b\&l} = \alpha^x_t \cdot f^x_{t+1,b\&l} + \alpha^y_t \cdot f^y_{t+1,b\&l} + \alpha^z_t \cdot f^z_{t+1,b\&l} \tag{10}$$

The values of (10) are compared to the subjects' average opinion in section 5.

4 The Rational Expectations Hypothesis

The rational expectations hypothesis (REH) was proposed by Muth (1961). He suggested that agents form their expectations consistently with economic theory. They should derive their forecasts from the true economic model that generates the variable to be predicted. The subjective distributions about future realizations should be the same as the actual distributions, conditional on the available information. With the knowledge of the true model that generated the time series in the experiment we are able to calculate the values of rational expectations. The information set of a rational forecaster contains the true model and its parameters (equations (1), (2) and (3)) and all the realizations of the time series observed so far. In the experiment without any indicators ($\Omega_t = x_t, x_{t-1} \dots x_1$) the REH values can simply be calculated by replacing u_t in (1) with its expected value 3.5. In versions 1,2 and 3 two auxiliary variables g_t and h_t need to be introduced.[1] The stochastic independence of the

[1] Again, the procedure is only described for version 3.

random variables allows the calculation of their simultaneous distribution $f(u, v, w)$ by the product of the distributions $f_u(u), f_v(v)$ and $f_w(w)$. From the equations (1), (2) and (3) we obtain (11) and (12).

$$u_t + v_t = y_t - 15 - x_{t-1} + int(\frac{1}{2} \cdot x_{t-2}) \tag{11}$$

$$u_t + w_t = z_t - 15 - x_{t-1} + int(\frac{1}{2} \cdot x_{t-2}) \tag{12}$$

We introduce the function δ_{uvw} with the values 0 and 1. δ_{uvw} has the value 1 if equations (11) and (12) are fullfilled simultaneously and has the value 0 otherwise. The auxiliary variables g_t and h_t can be calculated by summing up all discrete values of the three random variables.

$$g_t = \sum_{u=1}^{6} \sum_{v=-3}^{3} \sum_{w=-5}^{5} u \cdot f_u(u) \cdot f_v(v) \cdot f_w(w) \cdot \delta_{uvw} \tag{13}$$

$$h_t = \sum_{u=1}^{6} \sum_{v=-3}^{3} \sum_{w=-5}^{5} f_u(u) \cdot f_v(v) \cdot f_w(w) \cdot \delta_{uvw} \tag{14}$$

With these auxiliary variables and $E(u_t|x_{t-1}, x_{t-2}, y_t, z_t) = \frac{g_t}{h_t}$, the values of the REH $f_{t,REH}$ are calculated as follows:

$$f_{t,REH} = x_{t-1} - int(\frac{1}{2} \cdot x_{t-2}) + \frac{g_t}{h_t} \tag{15}$$

With these values it can be tested whether the rational expectations hypothesis gives a valid explanation of the subjects' forecasting behavior.

5 Results

Our results are an extension of prior experiments by Becker and Leopold-Wildburger (1996, 2000) without any indicators, the available information set being $\Omega_t = \{x_t, x_{t-1}, \ldots, x_1\}$. The comparison to the new versions allows a direct test of the effects of additionally presented information. Besides the performance of the subjects and both models at forecasting the base series in all four versions, our main interest is whether the group forecasts are consistent with the rational expectations and our bounds &likelihood heuristic.

The first six periods serve as a phase for familiarization and practice in all experiments. Hence, periods 1 to 6 are not taken into account within the statistical analysis. We only consider periods 7-42.

5.1 The forecasts of the subjects

The crucial question is to which degree the average forecasts can be explained by the b&l heuristic and the REH. For this purpose, a simple linear regression is estimated with the average forecast $f_{t,avg} = \frac{1}{N}\sum_{i=1}^{N} f_t^i$ (with the individual forecast f^i of N subjects) as a predicted variable and the two models as predictors:

$$f_{t,avg} = \alpha + \beta f_{t,\theta} \qquad \text{with } \theta = b\&l, REH \tag{16}$$

Both models are tested over the forecasting horizon of 36 periods. The results of version 3 are presented in Table 1. In version 1 (2) [3] the heuristic explains 95.3% (90.9%) [95.6%] of the variance of the average forecasts over time, while the REH explains only 79.7% (63.9%)[66.4%]. This result holds at insignificant autocorrelation of the residuals as indicated by the Durbin-Watson statistics.[2]

Considering the total sample of 36 periods does not tell us enough about potential learning processes or time invariance of the results. If the subjects learn to behave more rationally in the sense of the REH, then the coefficients of determination of regression (16) should be increasing over time. Analogous results should be observed for the b&l heuristic. In order to test for learning behavior, we divide the forecasting horizon into three sections, namely periods 7-18, 19-30 and 31-42. The results of these regressions of session 3 are reported in Table 1. For b&l, the coefficients do not vary in any of the 3 versions. At REH, the coefficients drop in the final periods which is especially obvious in versions 2 and 3 at values of 0.505 and 0.52 respectively. In all three versions the coefficients of determination are lowest in the final 12 periods. Thus, the subjects do not learn to behave more rationally, quite the contrary. The b&l heuristic performs strongly from the very beginning of the considered periods to the end.

Insert Table 1 about here.

It can be observed that the subjects forecast only three or four different values in some periods, while in other periods the individual forecasts deviate strongly. The variances of the forecasts are not constant over time. One reason for this phenomenon are the directions of change of the base series and the indicators. When the leading series indicate different directions for period $t+1$ than the base series in period t, the subjects are confronted with divergent signals. They have to decide whether to follow the trend of the base series or to follow the trend of the indicators. Three categories of divergent signals are defined:

1. Strong divergence (D_s): The signs[3] of both indicated directions are equal and differ from the base series: $sig_x \neq sig_y \wedge sig_x \neq sig_z \wedge sig_y = sig_z$
2. Medium divergence (D_m): The signs of all indicators differ from the base series but they are not required to be equal: $sig_x \neq sig_y \wedge sig_x \neq sig_z$

[2] The Durbin-Watson value $d = \sum_{t=2}^{T}(\widehat{u}_t - \widehat{u}_{t-1})^2 / \sum_{t=1}^{T}\widehat{u}_t^2$ detects first order autocorrelation of the estimated residuals $\widehat{u}_t$. $0 \leq d \leq 4$, d =2 indicates absence of autocorrelation, $d = 0$ strong positive autocorrelation and $d = 4$ strong negative autocorrelation.

[3] The signs are defined as $sig_x := sig(x_{t-1} - x_{t-2}), sig_y := sig(y_t - y_{t-1}), sig_z := sig(z_t - z_{t-1})$ with the standard definition of sig thus $sig \in \{-1, 0, 1\}$.

3. Weak divergence (D_w): Only one of the leading series indicates a different direction than the base series: $sig_x \neq sig_y \vee sig_x \neq sig_z$.

D_s, D_m and D_w are index sets. We define the convergent periods C_s, C_m and C_w as complementary to these index subsets: $C_\phi = [\{7, 8, ..., 42\} \backslash D_\phi]$ for $\phi = s, m$, and w. These definitions are only applicable to version 3 of the experiment. In versions 1 and 2 only one divergent signal is defined, namely $sig_x \neq sig_y$ and $sig_x \neq sig_z$ respectively. This corresponds to the definition of strongly divergent signals, when all available leading series indicate different directions than the base series.
In Table 2 the average variance of the forecasts are reported. The number of the concerned periods is reported in parentheses. The differences of the average variances is highest in strongly divergent periods and smallest when the weak definition is applied. Therefore these differences are significant on a 10%-level at least with exception of version 2 which is inconsistent with this pattern. The reason can be found in two outlying periods in which two subjects unreproducibly forecasted the maximum value of 30. Without these subjects the average variance would be lower in the convergent periods, but the differences in version 2 are not significant according to a Wilcoxon signed rank test (sig. = 0.65).

Insert Table 2 about here.

In order to test the stability of the performance of both models in periods which are more difficult to predict, we divide the dataset according to the definitions of divergent periods and convergent periods. The results of estimating equation (16) at these data subsets are presented in Table 1. Two phenomena can be observed. First, the b&l heuristic explains more of the variance of average forecasts in all versions, except the (strongly) convergent periods of version 2. Second, independent of the definition of divergence, both models explain more of the variance of average forecasts in convergent periods of almost all versions of the experiment. This finding provides substantial support for the b&l heuristic to be superior to REH even in these critical periods.
The excellent performance of the b&l-heuristic in explaining the average forecasts of the subjects seems to increase with the number of indicators: The values of the REH and the b&l model in version 0 are calculated and a linear regression of the corresponding average forecasts of 264 subjects are estimated as in (16). Without presenting the results in detail, the coefficients of determination are almost equal (94.5%) but the Durbin-Watson statistics indicates autocorrelation of the residuals in both cases. Hence, the heuristic is very well able to explain the forecasts also when no indicators are included.
The b&l heuristic models average forecasts, i.e. forecasts of representative agents. It is not intended to explain the forecasts of individuals. However, the model can be applied to individual forecasts and can be compared to the REH. In Table 3 the results of a simple linear regression are presented. A vector of all available individual forecasts in each version is explained by the values of the heuristic and the REH. Both

models explain far less of the variance of the forecasts, but the b&l heuristic explains still more than the REH. The intercept and the slope coefficients are identical to the estimates in Table 1, but the standard errors are lower. This causes the rejection of the null hypotheses of $H_0 : \alpha = 0$ and $H_0 : \beta = 1$. The coefficients of determination are lower than in any of the regressions of average forecasts, but at least in version 1 and 3 they are still remarkably high. The b&l heuristic explains more than 65% of the variance of the individual forecasts.

Insert Table 3 about here.

Furthermore the average absolute and squared deviations of individual forecasts from the two models are reported in Table 4. The b&l heuristic performs slightly worse than REH in version 0 but it performs better in all other versions. According to Wilcoxon sign ranked tests these differences are all highly significant, even in version 0. Hence the application of the heuristic to individual data brings similar results to the application to average data.

Insert Table 4 about here.

5.2 Forecasting accuracy

The forecasting accuracy of the subjects of the four versions is directly compared. From this comparison, conclusions on how the presentation of different indicators affects the quality of the forecasts can be drawn. In Table 5 the mean errors (ME), the mean absolute errors (MAE), the mean squared errors (MSE) and the Theil's U[4] for the direct comparison to naive (no change) forecasts are reported. A ranking of the four average forecasts after MAE, MSE and Theil's U indicates clearly: The most accurate results are achieved by those subjects only given the reliable indicator y_t. The presentation of the indicator z_t results in even lower forecasting accuracy than in the experiment without any indicators. In our experiment with two leading indicators presented to the subjects simultaneously, the accuracy improves but is still lower than in the case of indicator y_t. It can be concluded that only considering the more reliable indicator gives better forecasting accuracy than the consideration of both indicators. Furthermore, the weak indicator z_t does not improve the quality of forecasts, it actually seems to worsen the results.

Insert Table 5 about here.

[4]Naive forecasts are the most simple forecasting model. Models unable to beat naive forecasts are of no value. In our case, for all forecasts f_t, Theil's U = $\sqrt{\frac{1}{36}\sum_{t=7}^{42}(x_t - f_t)^2} / \sqrt{\frac{1}{36}\sum_{t=7}^{42}(x_t - x_{t-1})^2}$. Theil's U $>$ 1 indicates that the forecasts are less accurate than the naive forecast.

5.3 A simulation study

As expected, the analysis reveals that the REH performs better than the b&l heuristic in forecasting the base series x_t. Its forecasts are more accurate as indicated by the mean absolute error and the mean squared error. Although the REH is superior to the heuristic the accuracy of both models is higher than the naive forecast as indicated by low values of the Theil's U.
The experimental dataset of x_t, y_t and z_t represents only one of many results drawn from the distributions of u_t, v_t and w_t. To face the crucial question whether our results hold for other realisations, a Monte Carlo simulation is performed independently of our experimental results in order to generalize our results for other realizations of the time series. The three vectors of x, y and z are drawn according to (1), (2) and (3) over 42 periods. This is replicated over 100,000 times. The relative efficiency of the b&l heuristic is demonstrated by the following measure

$$\eta = \frac{\sum_{t \in T} |f_{t,REH} - x_t|}{\sum_{t \in T} |f_{t,b\&l} - x_t|} \tag{17}$$

with T being the number of predictions used. This simulation is performed for all four versions. The results of the simulation are presented in Table 6. Although the average deviation is lowest in version 3, its relative efficiency is only 69.0%, while it is 90.6% in the case of no additional information at a deviation of 1.656. We find the error to be lowest in version 4 and highest in version 1. With both indicators, the b&l heuristic performs better than in version 2 but the advantage of adding another indicator improves the accuracy only slightly. The results from the experimental dataset are given in Table 6. Both models predict the experimental time series less accurately. Almost all mean absolute errors are higher than the average values from the simulation. The relations η, however, seem to be fairly comparable. It can be concluded that the results from the experimental dataset are representative for the stochastic processes considered.

Insert Table 6 about here.

6 Conclusion

In this study we reported on a forecasting experiment with several sources of information. The rational expectations hypothesis is not able to explain the subjects' average forecasts in this abstract setting. These results are in line with Brennscheidt (1993) and Garner (1982) who had to reject the hypothesis of rationality in similar experimental settings. The simple bounds&likelihood heuristic explains average forecasting behavior very well, even in periods in which the leading series and the base series indicate divergent signals. The relative efficiency of the bounds&likelihood heuristic was successfully demonstrated by a simulation study over 42 periods with 100,000 replications. From the analysis of the simulated datasets it could be concluded that the findings with the experimentally applied time series are representative for the considered stationary stochastic processes.

However, our results are limited to some extent, namely to the number of time series in the information set. In real life decision-making situations a multitude of leading series is available. The application of the b&l-heuristic is limited to situations with few leading series and short time series. The rational expectations hypothesis is more accurate at an increasing number of leading series. At an infinite number of indicators its values equal the true realizations of the base series in all periods. The relative efficiency of the bounds&likelihood heuristic will converge to zero in this case. To be applicable to a setting with many indicators, the heuristic has to be extended by selection strategies because it cannot be assumed that subjects are able to form a linear combination of an arbitrary number of predictors. We suppose that when overloaded by additional information they will base their decisions on only a few selected indicators. In future experiments, the subjects' behavior in a setting with a multitude of indicators will be explored experimentally.

The bounds & likelihood heuristic is not intended to model individual forecasts. For individuals other models must be found. Therefore the next focus of future research will be the explanation of distributions of the forecasts. This subject was addressed by the definition of divergent signals which gave a first insight into the variance of the distributions.

7 References

Becker, O. (1967), Experimentelle Untersuchung der Erwartungsbildung für eine Zeitreihe, in: H. Sauermann (ed.): *Beiträge zur experimentellen Wirtschaftsforschung*, Tübingen.

Becker, O., Bolle, F. (1996), Expectations in Economics: Rational or not? Evidence from Experiments, *Jahrbuch Ökonomie und Gesellschaft*, Bd. 13, Campus, 88-119.

Becker, O., Leitner, J., Leopold-Wildburger, U. (in press), Heuristic modeling of expectation formation in a complex experimental information environment, *European Journal of Operations Research.*

Becker, O., Leopold-Wildburger, U. (2000), Erwartungsbildung und Prognose - Ergebnisse einer experimentellen Studie, *Austrian Journal of Statistics*, 29, 7-16.

Becker, O., Leopold-Wildburger, U. (1996), The bounds and likelihood-procedure - A Simulation Study Concerning the Efficiency of Visual Forecasting Techniques. *Central European Journal of Operations Research and Economics*, 4, 223-229.

Bolle, F. (1988), Learning to make good predictions in time series, In: R. Tietz, W. Albers and R. Selten (eds.), *Bounded Rational Behavior in Experimental Games and Markets*, Springer, Berlin.

Brennscheidt, G. (1993), *Predicitve Behavior - An Experimental Study*, Lecture Notes in Economics and Mathematical Systems Vol. 403.

Chapman, L.J. Chapman, J.P. (1969), Illusory correlation as an obstacle to the use of valid psychodiagnostic signs, *Journal of Abnormal Psychology*, 74, 271-280.

Clemen, R.T. (1989), Combining forecasts: A review and annotated bibliography, *International Journal of Forecasting*, 5, 559-583.

Dwyer, G. P., Williams, A.W., Battalio R.C., Mason T.I. (1993), Tests of rational expectations in a stark setting, *The Economic Journal*, 103, 586-601.

Fisher, F.M. (1962), *A Priori Information and Time Series Analysis*, Amsterdam.

Garner, A.C. (1982), Experimental Evidence on the Rationality of Intuitive Forecasters, In: V.L. Smith (ed.), *Research in Experimental Economics*, Greenwich. Conn: JAI Press.

Granger, C.W.J. (1989), Combining forecasts - twenty years later, *Journal of Forecasting*, 8, 167-173.

Goodwin, P., Fildes, R. (1999), Judgmental Forecasts of Time Series Affected by Special Events: Does Providing a Statistical Forecast Improve Accuracy? *Journal of Behavioral Decision Making*, 12, 37-53.

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

Lim, J.S., O'Connor, M. (1996), Judgmental forecasting with time series and causal information, *International Journal of Forecasting*, 12, 139-153.

Mason, T. (1987), *Expectation Formation in a Controlled Laboratory Environment*, PhD Thesis, Blomington, Indiana.

Muth, J.F. (1961), Rational Expectations and the Theory of Price Movements, *Econometrica*, 29, 315 - 335.

Sackman, H. (1974), *Delphi Assessment, Expert Opinion, Forecasting, and Group Process*, Santa Monica: Rand Corporation, Rand report R-1283-PR.

Sanders, N.R. (1997), The impact of task properties feedback on time series judgmental forecasting tasks, *Omega: International Journal of Managment Science*, 25, 135-144.

Schmalensee, R. (1976), An Experimental Study of Expectation Formation, *Econometrica*, 44, 17-41.

Footnotes:

1. Again, the procedure is only described for version 3.

2. The Durbin-Watson value $d = \sum_{t=2}^{T}(\widehat{u}_t - \widehat{u}_{t-1})^2 / \sum_{t=1}^{T} \widehat{u}_t^2$ detects first order autocorrelation of the estimated residuals $\widehat{u}_t$. $0 \leq d \leq 4$, d =2 indicates absence of autocorrelation, $d = 0$ strong positive autocorrelation and $d = 4$ strong negative autocorrelation.

3. The signs are defined as $sig_x := sig(x_{t-1} - x_{t-2}), sig_y := sig(y_t - y_{t-1}), sig_z := sig(z_t - z_{t-1})$ with the standard definition of sig thus $sig \in \{-1, 0, 1\}$.

4. Naive forecasts are the most simple forecasting model. Models unable to beat naive forecasts are of no value. In our case, for all forecasts f_t, Theil's U $= \sqrt{\frac{1}{36}\sum_{t=7}^{42}(x_t - f_t)^2} / \sqrt{\frac{1}{36}\sum_{t=7}^{42}(x_t - x_{t-1})^2}$. Theil's U $>$ 1 indicates that the forecasts are less accurate than the naive forecast.

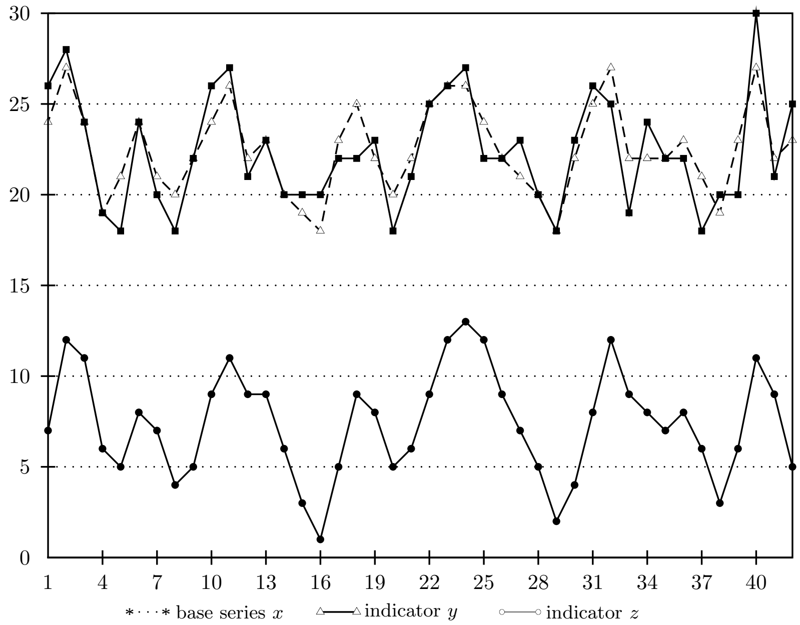

Figure 1: The base series and the indicators

Table 1: Regression results of version 3

model	**periods**	α	β	R^2	DW	df
b&l	7-42	0.542 (2.025)	0.953 (-1.343)	0.956	1.980	34
	7-18	1.065 (2.393)	0.893 (-1.623)	0.949	2.857	10
	19-30	0.064 (0.154)	0.988 (-0.240)	0.975	2.154	10
	31-42	0.110 (0.191)	1.010 (0.135)	0.950	1.967	10
	C_s	0.227 (0.847)	0.997 (-0.086)	0.975	2.281	21
	D_s	1.533 (2.596*)	0.813 (-2.332)	0.903	2.326	6
	C_m	0.107 (0.383)	1.007 (0.194)	0.978	2.427	18
	D_m	1.443 (2.909**)	0.834 (-2.447*)	0.917	2.292	14
	C_w	-0.298 (-0.870)	1.066 (1.478)	0.986	2.108	8
	D_w	0.917 (2.780**)	0.902 (-2.279**)	0.947	2.375	24
REH	7-42	1.350 (1.734)	0.824 (-1.757)	0.664	1.719	34
	7-18	2.098 (1.755)	0.682 (1.936)	0.633	1.631	10
	19-30	0.452 (0.368)	0.985 (-0.093)	0.796	1.198	10
	31-42	1.928 (1.047)	0.742 (-1.142)	0.520	2.400	10
	C_s	1.339 (1.494)	0.832 (-1.476)	0.718	1.353	21
	D_s	1.701 (0.850)	0.763 (-0.888)	0.426	2.069	11
	C_m	1.222 (1.357)	0.832 (-1.494)	0.754	0.931^{+}	18
	D_m	1.653 (0.845)	0.792 (-0.785)	0.390	2.372	14
	C_w	0.479 (0.555)	0.899 (-0.962)	0.902	1.884	8
	D_w	1.769 (1.581)	0.781 (-1.479)	0.537	1.567	24

Note: t-values of $H_0 : \alpha = 0$ and $H_0 : \beta = 1$ in parentheses. *: p<0.05; **: p<0.01. +: significant serial autocorrelation of lag 1 or higher; in these cases standard errors are enumerated with the method by Newey and West (1987).

Version	**Periods**	**Strong**	**Medium**	**Weak**
1	divergent	2.421** (19)		
	convergent	1.629 (17)		
2	divergent	5.711 (23)		
	convergent	8.599 (13)		
3	divergent	2.612* (13)	2.408* (16)	2.188 (26)
	convergent	1.833 (23)	1.938 (20)	2.040 (10)

Table 2: Average variances of forecasts in divergent and convergent periods. The differences between the subsets are significant at the 5%-level (*) or 10% level (**). The number of periods concerned is reported in parentheses.

model	Version	α	β	R^2	df
b&l	0	1.647 (0.092)	0.858 (0.012)	0.346	9502
	1	-0.020 (0.076)	1.013 (0.010)	0.653	5398
	2	1.261 (0.182)	0.877 (0.024)	0.377	2230
	3	0.542 (0.161)	0.953 (0.021)	0.655	1078
REH	0	1.225 (0.097)	0.873 (0.012)	0.346	9502
	1	0.591 (0.087)	0.904 (0.011)	0.546	5398
	2	1.306 (0.231)	0.851 (0.030)	0.265	2230
	3	1.286 (0.214)	0.831 (0.028)	0.457	1078

Note: standard errors in parentheses

Table 3: The linear regressions of individual forecasts in version 0,1,2 and 3.

	b&l		REH	
Version	**MAE**	**MSE**	**MAE**	**MSE**
0	2.137	4.568	2.1	4.403
1	1.089	1.19	1.331	1.772
2	1.777	3.157	2.096	4.395
3	1.094	1.527	1.196	2.33

Table 4: The average deviations of individual forecasts from both models

subjects	**information**	**ME**	**MAE**	**MSE**	**Theils' U**
264	x_t	0.517	1.745	3.980	0.741
150	x_t, y_t	0.060	1.031	1.597	0.469
62	x_t, z_t	0.352	1.771	4.324	0.772
30	x_t, y_t, z_t	0.217	1.393	2.971	0.641

Table 5: Forecasting accuracy in the different versions of the experiment

Dataset	Measure	Version 0	Version 1	Version 2	Version 3
Simulation	$\lvert f_{t,b\&l} - x_t \rvert$	1.656	1.211	1.433	1.144
	$\lvert f_{t,REH} - x_t \rvert$	1.500	0.896	1.147	0.789
	η	90.6%	74.0%	80.0%	69.0%
Experiment	$\lvert f_{t,b\&l} - x_t \rvert$	1.934	1.195	1.715	1.408
	$\lvert f_{t,REH} - x_t \rvert$	1.833	0.952	1.176	0.901
	η	94.8%	79.7%	68.6%	64.0%

Table 6: Relative performance (efficiency) of the b&l heuristic