An Experimental Study of the Nature of Consumer Expectations

Fergus Bolger

Abstract

Although important both theoretically and practically, the nature of consumer economic expectation formation has been little studied, particularly by psychologists. The most relevant previous research suggests that expectations are based on a heuristic that results in them being significantly biased. Further, relevant indicator series are poorly utilized. However, this earlier research used a task lacking in potentially important features of the real world, and this may have impaired performance. In the current experiment, participants received a more ecologically-valid task. Although there was still evidence of heuristic use, leading to suboptimal performance and bias, this performance was significantly better than anticipated from previous research, particularly regarding use of indicator series. However, when a strong trend in the criterion series allowed accurate forecasting without consideration of indicators, they were little used. I conclude that expectations are formed by first extrapolating the criterion series and only if that works poorly is other relevant information considered. Thus consumers appear to trade-off accuracy against effort, such that more effort is expended only when some threshold of acceptable performance fails to be reached.

The recent financial crisis and subsequent recession has led to a parallel crisis in Economics. How did so many economists fail to foresee the crisis? And what of the prevailing assumptions of rational economic agents, and consequently efficient markets, which justified the practices that largely caused the crisis? However, what is bad for economists turns out to be good for psychologists. Many of those who did not accept the rational-and-efficient assumption are either Keynesians, who acknowledge the important role of psychological variables in economic activity, or are those working on the margins of economics and psychology: decision scientists, behavioural economists, and economic psychologists (see e.g., Krugman, 2009).

The first of these latter, decision scientists, have the longest history, dating back to such critics of the rationality assumption as Herb Simon and Ward Edwards whose work began in the 1950's (Edwards, 1954; 1961; Simon, 1956, 1979). These researchers proposed that people are unable to conform to standards of rationality encapsulated in normative theories, such as Expected Value Maximization and Bayes' Theorem, due to psychological limitations: memory span, processing speed, channel capacity and the like. They then began to catalogue a number of biases in judgment and decision making that resulted from such restrictions, and proposed descriptive alternatives to the normative theories – the most notable being Prospect Theory (Tversky and Kahneman, 1979). More recently a growing number of economists have become interested in psychological aspects of economic behaviour, largely in response to the growing number of observed 'anomalies' – economic phenomena not accounted for by (neo) classical economic theory. This has resulted in the burgeoning field of behavioural economics/finance. Psychologists have also entered the fray, treating economic behaviour as a subject of

psychological inquiry without necessarily any reference to, or even knowledge of, economists' theories of this behaviour. Krugman (2009), for one, sees the current crisis as an opportunity for Keynesians, and economists with a psychological bent, to gain the upper hand in both economic theory and practice.

In this paper I wish to focus on a small, but important, part of this 'debate' between psychologists and psychologically-inclined economists, and rational-and-efficient economists – economic expectations. In any economy operating above subsistence level people have the ability to regulate their own economic behaviour, in other words, consumption becomes discretionary. Discretionary consumption means that people must *decide* how to spend, save, invest and so forth. As Katona (1972) points out this decision making depends upon various psychological constructs, particularly preferences between various consumption experiences, and expectations about certain future states being realized. For example, a decision to buy a car with your lottery winnings, or instead put the money into treasury bills, will be influenced by a combination of expectations about your future financial security (e.g. will you have a job in a year's time?), and the desirability of the alternatives (e.g. how much do you feel a new car will improve your life?). While preference formation is an area of great theoretical and practical significance (particularly to those wishing to influence tastes, such as marketers and politicians), it is not my concern here, rather, I wish to examine the basis of economic expectations.

A number of economists, and others, have sought to determine the nature of economic expectations. On of the earliest was Keynes himself who drew attention to the important role of expectations in the functioning of the economy in his 'General Theory

of Employment, Interest and Money' (1936). Keynes argued that the need to take action forces people to form expectations, and that these expectations are based on various 'judgment conventions'. An example of a judgment convention might be that the current state of affairs will continue indefinitely, unless there is strong evidence to the contrary. Judgment conventions, like Tversky and Kahneman's (1974) heuristics, provide reasonable, but imperfect, ways of reaching decisions – decisions that are often error-prone and sometimes even biased. The imperfect nature of both heuristics and judgment conventions permits them to account for various real-world behaviours that are not accountable for by models assuming rationality on the part of agents.

Keynes did not think that expectations could be measured or modelled accurately so did not specify the process of expectation formation in detail. However, Hicks (1939) suggested that expectations could be formed by simply extrapolating a trend in a data series into the future – Meeks (1991) took this to be an example of a Keynesian judgment convention, such as the one I gave as an example in the previous paragraph. The essence of Hicks' view is that people predict the future value of a variable by observing how that variable behaved in the past – I will refer to this as the Extrapolative Expectations model (EE). While this is, of course, a perfectly sensible means of proceeding – and the essence of many statistical forecasting models – it is not the only information that can be used for forecasting.

After Keynes, economics followed an increasingly rationalistic trajectory with regard to the assumed nature of *homo economicus* (Krugman, 2009). In this rationalistic view, if there is information that an economic agent can use then he or she should use it. One important source of information that can be used by a forecaster is his or her

previous forecast errors. This insight forms the basis of the Adaptive Expectations (AE) approach whereby an agent will revise his or her forecasts on the basis of the accuracy of previous forecasts (Cagan, 1956; Nerlove, 1958). So, for example, if the last forecast was too high then the next will be adjusted down, and so forth. Further, if previous forecast error is low then forecasts will be revised little, whereas large forecast errors will result in large revisions.

But still there is further information that the rational forecaster could use – in addition to past values of the to-be-forecast data series and previous forecast error there may be indicator series and contextual information that could improve forecasting. For instance, when forecasting sales of a product, one would think that marketing spend will be a leading indicator for sales (i.e. increased spending will precede increased sales), and that the launch of a related product by a competitor would be relevant contextual information. Indicators are of particular use when the target variable is difficult, expensive and/or slow to measure and thus values may be unavailable.

So, as economic agents became more rational, so did the means by which they formed their expectations. According to the Rational Expectations model (RE: Minford, 1992; Munth, 1961), people take *all* relevant information into account when forming their expectations. In its purest form, the RE model also requires that information is used in an optimal way, meaning that predictions are also optimal – no other model can do better – which implies that although forecasts might not be perfect they should be unbiased (because then a model correcting for this bias could do better).

Although these models have been around for quite some time we are only aware of one instance where they are tested against each other in the laboratory – an experiment

conducted by myself with colleagues 15 years ago (Harvey, Bolger & McClelland, 1994; henceforth HBM). In this experiment regression analyses and lack of performance improvement suggested that the AE model was inappropriate – and the RE model was rejected because of sub-optimal within-series forecasts and biased cross-series forecasts. We concluded that within series forecasts were made by extrapolation, specifically by using the most recent value of the to-be-forecast series as an anchor and adjusting it by the most recent change in the series so as to accommodate any trend – this is essentially the extrapolative heuristic suggested by Jones (1979), but never previously tested empirically. Cross-series forecasting – equivalent to using an indicator series – was very poor and seemed to be based on a faulty conditional rule.

To my knowledge HBM has not to date been followed-up so we are left only with the tentative conclusion that EE's are probably the best candidate for economic expectations. However, HBM used a task that was not very representative of what people might be called upon to do in everyday life, so the external validity of our findings is questionable – in the current research we wish to be able to generalize to expectation formation in consumers, specifically, as their expectations are crucial to predicting and understanding the business cycle. In HBM we presented participants with data based on a predator-prey relationship, thus there were two out-of-phase sine waves, one for the predator and one for the prey. As levels of prey increase so do levels of predators until the prey levels asymptote, due to the existence of too many predators, leading to the predator levels to fall too. Eventually predator numbers fall to an extent that the numbers of prey start to rise again, and so on. Now these series bear some similarity to economic variables such as GDP and related indicators like unemployment and interest rates, but

the series we used were much more regular and also free of any noise. We also gave participants a history of five previous data points (for both predator and prey series in the first half of the study, but just the predator series in the second half), and they received regular and accurate feedback about what actually occurred – neither of these conditions is likely to pertain for non-professional forecasters such as consumers. What would happen if participants attempted a forecasting task that was more representative of the sort of situation where they might need to form economic expectations in their daily lives?

To answer this question I first need to specify a more realistic sort of task. Let's return to the problem faced by a consumer that I posed earlier – 'should I spend or should I save'? As I discussed above, this decision is in part going to be driven by the consumer's expectations about the future state of the economy and, relatedly, his or her personal financial situation. Now it can be assumed that these sorts of expectations are being formed on a regular basis, so we have realism, but in order to conduct a controlled experiment we need to find an analogous situation that is amenable to investigation in the lab. Fortunately there is such an analogue of the consumer expectation situation – this is the collection of data used to construct the Index of Consumer Sentiment (ICS: aka 'consumer confidence'). Every month, in many countries, a panel of consumers is asked to predict whether the economy generally, and his or her financial situation specifically, will get better or worse, and by how much – their answers are then used to compose the ICS for the respondents' home country. Of course, this situation is not exactly the same as the one faced by ordinary consumers because panellists are being asked to express their expectations in a formal way, and (potentially) on a regular basis – this means that

respondents are generally going to attend more closely to information relevant to the questions they know that they are going to be asked, and monitor the accuracy of their predictions. However, I believe that it is a significant step in the right direction, and has the added advantage of allowing us further insight into the psychology behind the ICS itself. Further, if judgments are elicited regarding real data for a period during which the ICS is measured – which I do – we can compare the predictions of our forecasters to those of the panel, which could be informative.

So how do ICS panellists form their expectations? Assuming that they are not just guessing then their expectations must be based on their impressions of the current economic situation. I am proposing that these impressions depend on cues such as current unemployment rates, the state of the stock market, inflation levels, interest rates etcetera. Of course, this information is not going to be nicely packaged for panellists – it will come instead from a number of sources varying in precision and reliability: news reports, conversations with friends, and direct experience of prices, people being laid-off and so forth. Again, for the purposes of this experiment, I will not try to fully replicate all the noise and imprecision inherent in the real-world situation. Instead I will give participants several cues – in a precise and non-noisy way – and see how they form expectations on the basis of these cues.

An obvious way to model the expectation formation process that I have just described is with Brunswick's (1952) 'lens model'. In this model there is a distal variable, or criterion, that cannot be observed directly, such as the presence of a disease in a patient, but which manifests through several cues, for instance symptoms such as a high temperature, rash, and sweating. The cues are stochastically related to the criterion

and some cues are better predictors of the criterion than others. Note that the criterion, or cues, or both, can be either continuous or discrete variables. The cues are perceived by a judge and integrated into a judgment regarding the criterion (e.g. whether or not the patient has a disease) – this is the proximal variable. In the current context the distal variable is the state of the economy – past, present, and in particular, future – and the proximal variable is consumer confidence; the cues I have already described.

The lens model has been adopted as the central construct in Social Judgment Theory (Hammond, Stewart, Brehmer & Steinmann, 1975) where judgments, social or otherwise, are made on the basis of learning cue-criterion relationships. In a typical study to investigate such 'multiple cue probability learning' (MCPL) there are two phases: a learning and a test phase. In the learning phase, participants engage in a number of trials where they see the values of cues and then are told the value of the criterion – sometimes participants are asked to predict the criterion before its value is revealed. In the test phase, participants receive several trials where they are again shown the values of the same cues that they saw in the learning phase and are asked to make predictions regarding the value of the criterion – generally no outcome feedback (OFB) is given in this phase.

In MCPL studies researchers seek answers to two key questions: how good are people at predicting the criterion on the basis of the cues given? And how, and how well, are the cues used in order to make judgments? The first of these is referred to as 'achievement' and can be examined by the correlation between the criterion and the judgments of the criterion. The second is termed 'cue utilization' and can be examined through comparison of the correlations between cues and judgments with correlations

between cues and criterion (e.g. if cue utilization is good then people should place more weight in their judgments on those cues that better predict the criterion). Brunswick stressed that the correlations between cues and criterion are an extremely important feature of the task in that cue-criterion correlations place an upper-limit on judges' performance. Thus, for instance, if a set of cues can only explain 50% of the variance in the criterion then that is also the best that a judge can do. Further, if one is interested in looking at the learning and judgment of anyone who might loosely be described as 'expert' (i.e. performing a judgment task where they have prior experience with the cue-criterion relationships) then performance will almost certainly be impaired if the task cue-criterion correlations, or 'validities', do not correspond to the cue-criterion correlations experienced in the real-world (the 'ecological validities'). In the study to be described here I wish to give participants the opportunity to use their previous experience about the economy when making their predictions, or at very least, I do not want to make the cue-criterion relationships such that they are contrary to what participants might expect. To this end I use real economic data that participants will have been exposed to in their recent past.

In the current experiment I use the same basic design as described above, but with one important difference. In a typical MCPL set-up each coupling of cues and criterion (trial) is independent and the order of presentation of trials is randomized (or at least, counter-balanced). However, a member of an ICS panel observes cue-criterion relations for a particular month (and possibly for several months in the past) and then makes a prediction. The next month the panellist does the same thing, and this is the next 'trial'. Thus in this situation the cues and criterion at each trial are most likely *not* independent –

economic time-series, which are what we have, typically display trends and cycles that introduce significant amounts of serial dependency or 'autocorrelation'. As I mentioned in the previous paragraph, I am using real economic data that indeed contains substantial autocorrelation. One solution to this would be to randomly select the data for each trial from the time-series of the cues and criterion, of course ensuring that all data at each trial comes from the same month for the cues, and *n* months into the future for the criterion. There are some problems with this approach, though. First, the autocorrelation itself can be regarded as an important and useful cue – removing it will almost certainly make the task more difficult for participants. Second, and relatedly, studies of expectation formation and judgmental forecasting, such as HBM, use time-series regression analyses to identify what information from the past is used by forecasters to make their predictions – with each month's data removed from its historical context such analysis could not be performed. Third, as discussed above, I wish to maintain the ecological validities of the cue-criterion relationships as far as possible, and disordering the time-series in this manner will be detrimental to this goal (and again negatively impact on performance). For these reasons I have opted *not* to disorder the time-series used in this study.

Although, I will be facilitating judgment as far as possible by providing ecologically valid cues, presented as they would be in reality (as consecutive observations in time-series), *and* giving data in a regular manner, with no omissions, and free of any degradation (three conditions unlikely to pertain for ICS panellists), I do not anticipate either good achievement or cue utilization on the part of my participants. The reasons for this are that HBM found such poor cross-series forecasting with a much simpler task (only one cue, noiseless series) and because the typical finding of MCPL

studies is that, unless the situation is very simple (i.e. very few cues that are uncorrelated and positively related to the criterion), learning from outcome feedback is slow or non-existent (e.g. Deane, Hammond & Summers, 1972; Klayman, 1988). My task is relatively complex, requiring more computational effort on the part of participants than a typical MCPL task or the one used by HBM, thus the heuristic strategies of extrapolative expectations should be more prevalent.

Method

Participants

27 undergraduates (11 M, 16 F) aged 19-24 participated for course credit. All the participants were Turkish.

Materials

The stimuli were quarterly Turkish economic series for the 6-year period 2003-Q3 to 2009-Q2; the most recent data being for the quarter immediately preceding the study. The four indictor series were the current unemployment rate, annual interest and inflation rates, and stock-index change from the last quarter – all as percentages. There were two criterion series, one for each experimental group: GDP expressed as percentage change or absolute value (constant-price index).

If these participants use the within-series forecasting strategy found by HBM – estimating trend from the last change in the series and adding this to the most recently observed value of the series – then those receiving GDP-change series should find the

task easier than those receiving absolute-GDP series because no trend computation is needed. However, much autocorrelation that is useful for forecasting is absent from change series so the opposite result might be observed.

To disguise the source of the data, and facilitate presentation, the series were rescaled to be within similar ranges. The data covered about one cycle from trough-to-trough of the business cycle and were presented on 23 PowerPoint slides in each phase, one for each trial. There was a 2-Quarter moving window so the participants always saw the most recent past values of cues and the current values (see Figure 1). Note that in Phase 2 *exactly the same data series were presented to participants as in Phase 1* (and in the same order), but no OFB was provided.

[Figure 1 about here]

Whereas ICS panellists normally make predictions for the next quarter to a year ahead, my participants did not really forecast at all, but merely report the *current* value of the criterion given the *current* values of cues. The reason for this is to maintain comparability with MCPL studies and HBM's cross-series forecasting task. Lagging the OFB by 1-4 quarters is an additional complication that can be investigated in future studies.

Procedure

Participants were first supplied with general written instructions and a response sheet for Phase 1, where predictions for each quarter could be made by putting an X in one of 10 boxes covering the range the series fell within.

Participants were tested in two experimental groups. In addition to the written instructions the experimenter explained that the experiment required economic forecasting but that economic expertise was not required. Further, there would be two parts to the experiment and that details of the second part would be given after the first was completed. After the participants wrote their name, age and sex on the response sheets, and any questions were answered by the experimenter, the first data slide of Phase 1 was presented. The imminent presentation of each new slide was announced by the query 'ready?' when additional time for responses could be requested. After Phase 1 the response sheets were collected in, the response sheets for Phase 2 distributed, and further instructions given. A second set of response sheets identical to those at Phase 1 were handed out. The slides for Phase 2 were presented in the same way as for Phase 1, then the second response sheets were collected and the participants debriefed.

Results

Figure 2 shows the forecasts for the four different conditions of the experiment averaged over the 13 participants in each of these four conditions[1]. It can be seen that Phase-1 Change forecasts tend to lag behind the true series by approximately a Quarter, except for the last 2 Quarters where participants prematurely forecast a recovery. Phase-2 Change forecasts are very similar to Phase 1 – forecasts are of lower amplitude than true,

14

probably due to averaging over participants' forecasts. Phase-1 Absolute forecasts quite closely track the true series, but they too mostly lag behind by about a Quarter. At Phase 2 there is much less trend and consequently the true series is tracked poorly.

[Figure 2 about here]

Achievement: How Good Were the Forecasts?

Two aspects of the accuracy were analyzed: mean signed error (True - Forecast), representing bias, and mean absolute percent error (MAPE), which is comparable across different measurement scales, were computed for each participant. Excepting Phase-1 Change forecasts there was significant positive bias indicating that forecasts were too low – they lagged behind a mostly positively trended series. The MAPE's for the 2 conditions x 2 phases, and t-tests comparing these performances, are shown in Table 1. The best the participants could have done using the cues available was a MAPE of 48.1% for the Change series, and 1.24% for the Absolute (using the best-fitting regression models on the cue and criterion series provided). It is clear the Absolute series is much easier to forecast than the Change series, presumably due to the relatively strong trend, and thus autocorrelation. It can also be seen that Phase-1 Change accuracy was not significantly different from the benchmark, whereas Absolute forecasts at both phases were clearly and significantly worse. Phase-2 Change forecasts were slightly worse than benchmark (but still not significantly), whereas Absolute forecasts were hugely (and again significantly) worse.

15

[Table 1 about here]

Cue Utilization

The cues used for forecasting were compared with the cues that *should* have been used according to their ecological validities. Cue utilization is good for the Change series – participants paid attention to the most relevant cues: previous GDP change (where available); and unemployment rates and stock prices; although stock-price change was given too much weight, particularly at Phase 2 (see Table 2). Cue utilization was also fair for the Absolute series. The best predictor by far was previous GDP – reflecting the strong trend in the series – which was preferred when available. Otherwise indicators of interest rates, inflation and stock prices were the most highly correlated with the criterion, and these were the cues used.

[Table 2 about here]

How Were Forecasts Made?

I conducted stepwise regressions with average forecasts as the criterion and with 9 (Phase 1) or 8 (Phase 2) predictors (since previous values of the criterion were not given at Phase 2, this was not included as a predictor at Phase 2) using standardized data. The remaining 8 predictors were the current and previous values of the four cues. The best-fitting models for the Change series were:

Phase-1 Forecasts = 0.85 Previous GDP + 0.41 Stock Prices, R^2 (adjusted) = 86%

Phase-2 Forecasts = 0.76 Stock Prices - .35 Unemployment rate, R^2 (adjusted) =

73%

The best-fitting models for the Absolute series were:

Phase-1 Forecasts = 0.93 Previous GDP, R^2 (adjusted) = 87%

Phase-2 Forecasts = 0.63 Previous Interest Rates, R^2 (adjusted) = 36%

All β's are significant p = .016 or less. Note that the residuals displayed no significant

autocorrelation suggesting that the R^2 values are not overestimates.

Phase-1 forecasts in both conditions were apparently made using a *naïve*

forecasting strategy (i.e. giving the last true value of the criterion as the forecast) with

some adjustment based on current stock-price change in the case of the Change series.

Naïve forecasting tends to lead to forecasts lagging behind the true series when there is

trend, this accounts for the observed bias and consequent suboptimal performance

described above – the adjustment in the Change series was not sufficient to overcome this

bias. Phase-2 forecasts could not be made using a naïve strategy because previous GDP

values were no longer available. Forecasts appear instead to have been made using

current unemployment rates and stock-price change (Change series), or solely using prior

interest rates (Absolute series).

I also conducted similar analyses to those described above for the average

forecasts to discover the determinants of each individual participant's forecasts. In this

case I added an additional predictor, the participant's own prior forecast, this being

available to each participant on his or her response sheet[2]. For Phase-1 Change series the

stepwise procedure found a significant model for all but two of the participants. In

accordance with the analysis of average forecasts, previous values of GDP, and current

and previous values of stock-price change, were the predictors of forecasts appearing most commonly in the models (see Figure 3). The analysis of Phase-1 Absolute series produced significant models for all participants and a similar pattern of cue usage to that for the Change series, with prior values of GDP and prior and current stock-price change appearing most often as predictors in the models. Overall cue usage appears to have been greater by participants in the Absolute condition than the Change, though, and reliance on prior GDP was significantly greater in the former condition than the latter (13 vs. 8 of the participants respectively used this cue, chi-squared = 6.19, p = .013). Comparison of cue-usage between the Change and Absolute conditions at Phase 2 also shows some similarities and some differences (see Figure 3). In both conditions previous and current stock-price change appeared frequently in the models, and to a lesser extent current and prior unemployment rates. However, although for the Change condition it was stock-price change that most commonly best predicted forecasts, for the Absolute condition it was the prior forecast – I will attempt to account for this second, and somewhat puzzling, finding in the Discussion section.

[Figure 3 about here]

Are the Forecasts like Consumer Confidence?

The Central Bank of Turkey started to construct a consumer confidence index (CBCCI) monthly starting in December 2003. The index is based on the responses to a short survey of around 2000 individuals over the age of 15 who are in employment – the sample is designed to be representative of the overall population in terms of age, employment status

18

and income groups. Each respondent has a 50% chance of being surveyed a second time (after 3 months) so opportunities for learning are very limited, particularly in comparison to my experimental setup. Four tendencies are measured in the survey: personal financial standing (4 questions); general economic situation (5 questions); expenditure plans (4 questions); and price expectations (1 question). I compare the forecasts in my study to responses to two questions designed to measure attitudes regarding the general economic situation to the forecasts of my participants: "compared to the past 3 months the current economic situation is much worse (a little worse, about the same, a little better, much better)"; "in the next 3 months the economic situation will get much worse, get a little worse (stay about the same, get a little better, get much worse)". The first of these questions seems the most directly comparable to the task in my experiment because, as I already pointed out, participants are more accurately 'nowcasting' than forecasting. In other words, the task was to estimate the current value of GDP given its previous value and previous and current values of the indicator cues. However, we can also compare the nowcasts to consumer confidence regarding the next quarter by lagging the latter by one quarter. Thus, for example, estimates of the GDP for Q1 of 2008 made by my participants should be equivalent to consumer confidence for the next quarter measured in Q4 of 2007.

The task of respondents to the CBCCI survey is to forecast *change* in the economic situation – this is equivalent to what my participants in the Change condition had to do. In order to make the task equivalent in the Absolute condition, I computed the difference between each consecutive average forecast. These 'first differences', as they are known, represent the amount of change in GDP each quarter forecast by participants

in the Absolute condition and result in series that appear quite similar to the corresponding Change average forecasts for the two phases. The correlations of forecasts in each condition and at each phase of the experiment with the two measures of consumer confidence are shown in Table 3. It can be seen that correspondence between Change forecasts at both phases are moderately and significantly correlated with both consumer confidence measures whereas the correlations between Absolute forecasts and each of the two consumer confidence measures at both phases are weak and non-significant despite being rendered equivalent.

[Table 3 about here]

Discussion

The findings suggest a much more optimistic picture regarding the quality of consumer expectations than did those of HBM. Although in Phase 1 (both conditions) the MAPE's were huge compared to those obtained in this earlier study the task was much harder. Further, at Phase 2 the MAPE's for both conditions were substantially *lower* then the cross-series MAPE of 73% observed by HBM. Also, in the Change condition at both Phase 1 and 2 the MAPE was not significantly larger than that for the benchmark regression model. True, the MAPE's for the Absolute condition are very poor compared to what could be achieved even by a naïve forecaster, and Change forecast MAPE's at Phase 1 are hardly any better than the naïve benchmark, but then Change forecasts hold up pretty well when outcome feedback is removed (when naïve forecasting is no longer possible). Cue utilization was also pretty good – participants in both conditions tended towards the best cues. Taken together the evidence is that, in this more realistic task than

that used by HBM, and under certain conditions (i.e. no strong within-series trend),

people may be able to make a reasonable attempt at cross-series forecasting after all.

My interpretation of the results is that, when confronted with a set of cues for a

novel prediction task the participants quickly discovered if there was a cue that was

highly predictive of the criterion – as was the case with previous GDP in the Absolute

condition – and used this cue while ignoring other cues. If the participants could not find

a clearly dominant cue – as was the case for the Change series – then they sought

something that would work and, mostly, seemed to find it. By the end of Phase 1 more

than 50% of participants in the Change condition were using either stocks or

unemployment (or both) and these were indeed the best cues to use. When it came to

Phase 2 the fact that those in the Absolute condition had an easy time at Phase 1 may

have meant that not all the available cues were properly appraised. Thus when the

dominant cue was removed – the one that had been relied on – these participants were in

trouble. Hence the big increase in error at Phase 2 for these forecasters. In contrast, those

in the Change condition had to attend to the cues at Phase 1 that, fortuitously, were still

available at Phase 2, and as a consequence they suffered little when outcome feedback

was removed.

Thus in both conditions participants quickly found the best cues to use for

forecasting at Phase 1 and used them – this seems clear from the data – but how did they

find these cues? In a standard MCPL experiment, good cue utilization can only be the

result of learning, however, in my study there is another possibility. As a consequence of

using real (and recent) economic data that was correctly labelled, participants could have

been using their domain knowledge to guide cue selection rather than inducing which

21

cues to use purely from the data given. Thus participants may have simply inferred that stock prices and unemployment rates are the best predictors of GDP. The participant-by-participant analysis shown in Figure 3 provides some support for this view because we can see that stocks and unemployment were the most highly used cues at Phase 1 after previous GDP in *both* conditions despite the fact that interest rates and inflation had higher ecological validities as cues in the Absolute condition (see Table 2). If people *are* selecting cues on the basis of their past experience, and the cues they choose happen to coincide with those with the highest ecological validity, as in the Change condition, then this might explain why cue utilization is so good despite relatively little time to learn (in MCPL time – of course 23 quarters is nearly 6 years in real time, and much much longer than the time available to consumer-confidence panellists to learn on the job).

It is important to note, though, that it may not be entirely accidental that participants chose the cues with the highest ecological validities on the basis of their experience, this is what Brunswick would have expected. Incidental learning of ecological validities is also claimed by neo-Brunswikians, most notably Gerd Gigerenzer (e.g., Gigerenzer, Hoffrage & Kleinbölting, 1991), if not yet reliably demonstrated. Cue selection at Phase 1 cannot be solely top-down, though – the greater reliance on the previous value of GDP in the Absolute condition than in the Change condition suggests that the participants were also attending to the feedback regarding cue validities in the task at hand, or were at least sensitive to features of the stimulus series when formulating their forecasting strategy (cf. Bolger & Harvey, 1993). Further, even if participants were using prior experience to determine which cues to attend to in the Change condition this does not preclude them also monitoring the success of these cues by attending to

feedback (in the Absolute condition use of stocks and unemployment would not be negatively reinforced for those who also used previous GDP because their overall strategy would work reasonably well). I therefore conclude that a mixture of top-down and bottom-up processes were used for forecasting at Phase 1. At Phase 2 participants in both conditions continued to favour stocks and unemployment for forecasting, which was fine in the case of the Change series but not so good in the case of the Absolute, thus contributing to the relative collapse in performance in the latter condition at Phase 2.

So were expectations rational, adaptive or extrapolative? In the Absolute condition the forecasts were clearly not rational because the MAPE's were too high, even at Phase 1, and there was a clear bias due to under-adjusting for trend. However, there is evidence that Absolute forecasts may have been adaptive for some participants, previous forecast error being a significant predictor of 4 out of the 13 participants' Phase 1 forecasts. For the majority, though, Phase 1 Absolute forecasts, being a proportion of the previous value of the criterion series, were clearly extrapolative. Phase 2 Absolute forecasts could not be adaptive, as without OFB there was no way for participants to judge their forecast error. They could not be extrapolative either, at least in the sense that they merely extended the past values of the to-be-forecast series into the future since this series was no longer available to them. However, I wish to propose loosening the definition of EE's that I gave in the Introduction to include forecasts based on observation of *any* relevant historic information, not just that relating to the variable to be forecast, and indeed, information about the present situation too – I do not think this looser definition goes against the spirit of either Katona's or Keynes' understanding of extrapolative expectations. By this definition, the Absolute forecasts at Phase 2 could be

23

regarded as extrapolative if they relied on some information from the past or current situation excepting their own forecast error (in which case they would be AE's), or everything relevant (in which case they would be RE's). Thus, in my view, the aggregate Absolute Phase 2 forecasts, that mainly seemed to depend on previous interest rates, should be considered extrapolative.

Analysis of the individual forecasting strategies shows that more than half (9/13) of the participants in the Absolute condition made their Phase 2 forecasts using their previous forecast, either alone or in conjunction with some other cue. How can we explain this? One possibility is that, since at Phase 2 they were forecasting exactly the same series as at Phase 1, they used their memory of the starting point and/or trend to make their forecasts and either ignored the cues altogether, or used them to make some adjustment to their remembered 'forecast'. Note that due to the strong trend in the Absolute series any forecasts that try to replicate this trend will tend to be highly autocorrelated thereby increasing their probability of appearing as a significant predictor in the regression model. I will return to discuss the effect of memory of Phase 1 on Phase 2 forecasting shortly.

Turning now to the Change forecasts, they were again suboptimal and biased so the RE hypothesis can be rejected (although not quite as wholeheartedly as in most previous studies, or indeed for the Absolute condition of this study). AE's have some support in that at the individual level 5 out of 13 participants made use of their previous forecast error directly. The majority, though, made their forecasts at Phase 1 using current and past information about the criterion and indicator series and at Phase 2 using current

and past information about the indicator series. Thus it appears that the Change forecasts are mostly extrapolative in my new, looser, sense of the term.

There are, of course, several limitations to the current study, but none I think is major. The sample size is rather small, but there is quite a lot of agreement between the strategies used by participants – and the effects of these strategies such as extent of error – within each of the two conditions leading to clear (and statistically reliable) differences. It does not appear to me that a larger sample size would have produced further, or different, information.

The use of economic data from just one country for a particular period of time could also be criticised for limiting the generalizability of the findings. For example, it could have been that participants recognized the pattern of rise-and-fall in the economic data as part of recent history – but none admitted that they did in debriefing – future studies might benefit from using data from further in the past or, better, simulated data that maintains the ecological validity of cue-criterion correlations from the recent past, but permits sampling from any part of the business cycle. A related criticism is that, since my findings are somewhat different from HBM where we used different stimulus series, it seems that we need to use a wide range of series in order to fully understand how expectations are formed. Clearly the forecasting strategies or heuristics used are rather 'context specific', as we previously observed (Bolger and Harvey, 1993). However, it would not be practical to explore fully the implications of changing all possible parameters of the stimuli in one study – a series of experiments is needed, preferably using simulated data to allow systematic manipulation of series' characteristics.

Perhaps, the most serious criticism, and one that I alluded to above, is that, since exactly the same series, in the same order, were used twice, people may have used their memories of their forecasts at Phase 1 to inform their Forecasts at Phase 2. Now this is plausible, and would represent a feature not available to the usual ICS panellist, but I argue that it is not a serious flaw. Participants were not told that they would have to forecast the same series twice, or that the second time they would have no outcome feedback, so presumably they would not have tried to memorize anything at Phase 1. I therefore suspect that, if anything, they only had the vague impression of the fact that the criterion series rose and then fell to work with at Phase 2. However, in the Absolute condition, where the trend was fairly salient, and where participants were presumably clutching at straws, remembered trend may have been used by some participants, as I discussed above. In the Change condition where the trend was much less salient, and where many of the participants had identified useful cues for forecasting, it seems unlikely that the remembered trend would be a significant influence on the forecasts. However, if future studies used simulated data that maintained ecologically valid cue-criterion correlations, as suggested above, then training and test series could be created that were different in appearance but were identical in every other aspect, thereby eliminating the possibility of basing forecast on memory.

A further problem with using real data is that one has no control over the size of the cue-criterion correlations (except by selecting indicator variables that are related in the way that you want to the criterion, if they exist), and also no control over the inter-cue correlations. Both of these factors could affect the cue utilization and achievement of

26

participants and in future studies it would be interesting to systematically manipulate them, again by using simulated data series.

Finally, it would be interesting to investigate the interaction between top-down and bottom-up processes. This could be done by, for example, providing stimulus series where the best cues were not those that the participants expected to be the best on the basis of past experience. This could be done by, for example, changing the labels on the cues – longer series than those used in this study should probably be used, though; this would give participants the chance to revise their theories on the basis of the outcome feedback received.

Conclusions

In contrast to HBM's findings, it seems that people *can* use indicator series for forecasting, under the right circumstances – these circumstances might include having naturalistic series with ecologically valid cue-criterion correlations, and *not* having a dominant cue that 'blind' forecasters to other important information. The evidence from this experiment again does not favour the RE approach – too much bias, although the Phase-1 Change series aggregate forecast were pretty close to being as good as the least-squares regression model. The AE model is not very-well supported by the data either, although a minority of the participants may have used it at Phase 1. In real life, the situation would be more like Phase 2, with no (explicit) outcome feedback, so an AE strategy could not be used (or would be difficult to use – ICS panellists are asked to make forecasts as much as a year into the future so that feedback, if they received it at all, would usually be too lagged to be of much use). Having rejected RE and AE, we

conclude, in agreement with HBM, that expectations were extrapolative, but not in the purest sense of a simple extrapolation of the criterion series, rather in a more general sense of being based on observations of some (but not all) past and current values of the criterion *and* relevant cues.

To summarize my view of the expectation-formation processes used by the majority in this study, first, participants try to extrapolate using within series data but if this does not lead to satisfactory forecasting, as for Phase 1 Change series, then indicator series will be used (and possibly their relationships with the criterion series learned). If this analysis is correct then it supports my developing view that although people might not use the optimal strategy when making forecasts, they find one that offers a good accuracy-effort trade off (Bolger & Harvey, 1993; HBM). It also raises the interesting question of how people determine a satisfactory level of forecasting performance against which to trade the effort of searching for additional cues and learning their validities. Plausibly each individual has his or her own thresholds for acceptable performance and these thresholds could be different for different task domains according to past experience and/or knowledge of those domains – a future line of research would therefore be to ascertain if such performance thresholds exist in a forecasting context and, if so, how they are set.

To conclude, although there are some limitations to the current experiment, most of these limitations are the consequence of using real economic series as stimuli. This approach more than compensated for its limitations by showing that, although too biased and error-full to be considered strictly 'rational', the expectations of consumers are more measured than recent commentators, including myself, have suggested. Indeed my

findings are generally pretty favourable regarding the idea of using consumer-based

forecasts as indicators of the future state of the economy. For example, the average

forecasts by the Change group predicted the current recession three Quarters before it

was actually declared.

Notes

[1]Although it is, of course, of interest to examine the forecasting strategies of each

participant individually it is the average forecasts that are of central interest here. This is

because these are the equivalent of measures of consumer expectations such as the ICS

(although based on a much smaller sample, of course). Further, it is the economic

expectations of groups rather than individuals that are typically the subject of economic

theory and practice. Participant-by-participant analyses are of interest to Psychologists,

though, and I will refer to these in order to interpret the average forecasts better.

[2]It seems reasonable to focus on the cues directly available to participants, however, I

also conducted regression analyses using predictors that could be derived from the data,

namely the difference between current and prior indicators and, for individual data, the

forecast error (it is not meaningful for aggregate data). The former did not enter

significantly into either the analysis of aggregate data or that for individual participants

but the latter entered significantly for a few of the participants – I address this in the

Discussion.

References

Bolger, F., & Harvey, N. (1993). Context sensitive heuristics in statistical reasoning. *Quarterly Journal of Experimental Psychology*, *46A*, 779-811.

Brunswik, E. (1952). The conceptual framework of psychology. In *International Encyclopedia of Unified Science*, *1*. Chicago: University of Chicago Press.

Brunswick, E. (1955). Representative design and probabilistic theory in a functional psychology. *Psychological Review*, *62*, 193-217.

Deane, D., Hammond, K., & Summers, D. (1972). Acquisition and application of knowledge in complex inference tasks. *Journal of Experimental Psychology, 92,* 20-26.

Gigerenzer, G., Hoffrage, U., & Kleinbölting, H. (1991). Probabilistic mental models: A Brunswikian theory of confidence. *Psychological Review*, *98,* 506-528.

Hammond, K. R., Stewart, T. R., Brehmer, B., & Steinmann, D. O. (1975). Social judgment theory. In M. F. Kaplan & S. Schwartz (Eds.), *Human judgment and decision processes* (pp. 271-312). New York: Wiley.

Harvey, N., Bolger, F., & McClelland, A. (1994). On the nature of expectations. *British Journal of Psychology*, *85*, 203-229.

Katona, G. (1972). Theory of expectations. In B. Strumpel, J. Morgan & E. Zahn (Eds.), *Human behavior in economic affairs: Essays in honour of George Katona* (pp. 549-582). Amsterdam: North-Holland.

Katona, G. (1974). Psychology and consumer economics. *Journal of Consumer Research*, *1*, 1-8.

Keynes, J.M. (1936). *The general theory of employment, interest and money.* London: Macmillan.

Klayman, J. (1988). On the how and why (not) of learning from outcomes. In B. Brehmer & C.R.B. Joyce (Eds.), *Human judgment: The SJT view*. Amsterdam: North-Holland.

Krugman, P. (2009). How did Economists get it so wrong? *The New York Times*, Sept. 6, p. MM36 (New York Edition).

Tversky, A., & Kahneman, D. (1974). Judgment under uncertainty: Heuristics and biases. *Science, 85*, 1124-1131.

a

	Mean	SD	Comparison	*t*	*p*	Cohen's *d*
CH1	0.13	0.54	$\mu = 0$	0.85	.410	0.24
CH2	0.28	0.30	$\mu = 0$	2.87	.014	0.93
AB1	0.38	0.37	$\mu = 0$	3.67	.003	1.03
AB2	1.13	0.82	$\mu = 0$	4.97	<.001	1.38

b

	Mean	SD	Comparison	*t*	*p*	Cohen's *d*
CH1	49.3%	12.8%	CH1 v. CH2	2.01	.067	0.56
CH2	56.0%	16.9%	CH1 v. AB1	3.76	.001	1.53
AB1	31.7%	11.1%	AB1 v. AB2	2.98	.011	0.83
AB2	64.2%	40.2%	CH2 v. AB2	0.68	.505	0.28

Table 1: (a). Mean-Signed Error (MSE) and (b) Mean Absolute Percent Error (MAPE) for the two conditions at each of the two experimental phases (CH1 = Change series, Phase 1; AB2 = Absolute series, Phase 2 etc.).

	CUES	Change Series			Absolute Series		
		EV	CU-1	CU-2	EV	CU-1	CU-2
Prior	**Interest Rates**	+.08	+.00	+.14	-.51*	-.44*	-.63*
	Unemployment	-.55*	-.34	-.22	-.25	-.01	+.12
	Inflation	-.18	-.19	+.01	-.56*	-.33	-.55*
	Stock Prices	+.59*	+.32	+.32	-.24	-.42*	-.45*
	GDP	+.53*	+.83*	----	+.95*	+.95*	----
Current	**Interest Rates**	+.19	+.14	+.17	-.40	-.31	-.52*
	Unemployment	-.65*	-.58*	-.43*	-.15	+.09	-.17
	Inflation	-.15	-.20	-.07	-.31	-.17	-.46*
	Stock Prices	+.17	+.37	+.80*	-.41	-.36	-.44*

Table 2. Cue ecological validities (EV) and cue utilization for each of the two phases (CU-1 and CU-2, respectively) broken-down by series type (Change or Absolute). The validities are Pearson correlation coefficients and asterisks indicate $p < .05$.

Experimental	Consumer Confidence	
Condition	Current Situation	Lag 1 Next Quarter
Change Phase 1	.48*	.55**
Change Phase 2	.62**	.50*
Absolute Phase 1	.08	14
Absolute Phase 2	.27	.26

Table 3. Correspondence between forecasts in the current experiment and consumer confidence measured during the same historic period from which the experimental stimuli were drawn, broken-down by experimental condition and confidence measure. The figures are Pearson correlation coefficients and a single asterisk indicates $p < .05$, two asterisks $p < .01$.

Phase 1

INT	UNE	INF	STO	GDP
6.13	5.10	7.57	+6.3	33.0
5.20	5.10	5.99	+27.7	?

INT	UNE	INF	STO	GDP
5.20	5.10	5.99	+27.7	32.4
4.60	5.45	4.44	+15.3	?

Phase 2

INT	UNE	INF	STO	GDP
6.13	5.10	7.57	+6.3	XXXX
5.20	5.10	5.99	+27.7	?

INT	UNE	INF	STO	GDP
5.20	5.10	5.99	+27.7	XXXX
4.60	5.45	4.44	+15.3	?

Figure 1: Stimuli for two consecutive trials in the Absolute GDP condition at Phase 1 and Phase 2.
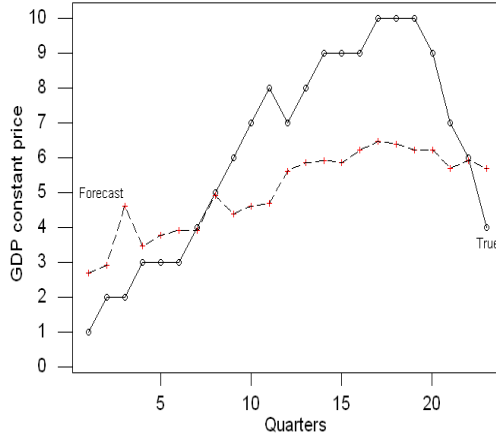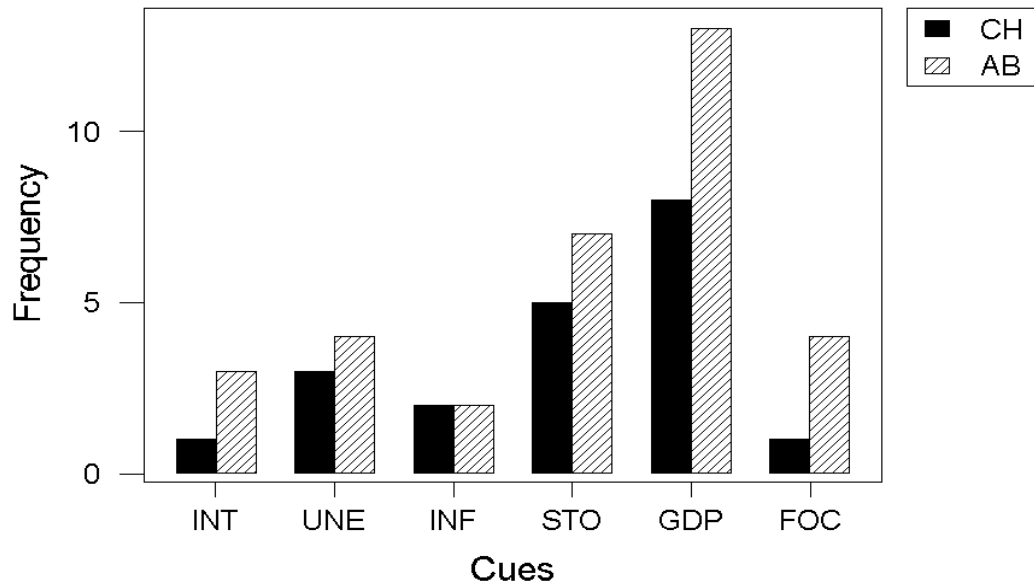
Figure 2: Average forecasts by condition and phase compared to the true values.
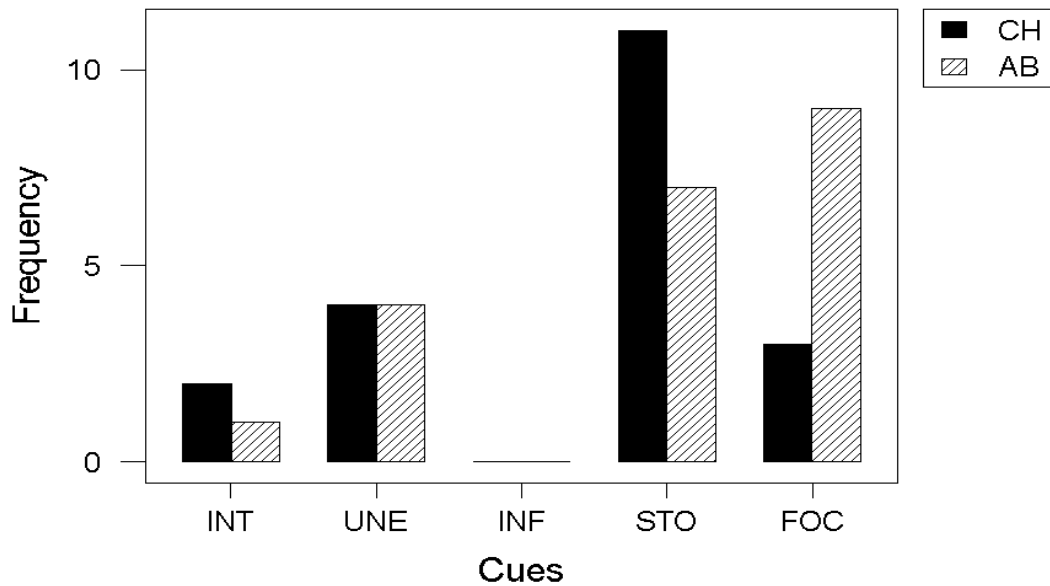
Phase 1

Phase 2

Figure 3: Frequency of cues appearing as significant predictors of individual participant's forecasts as a function of experimental condition (CH is Change and AB is Absolute) and phase. Cues entered were: previous and current interest rates (INT), unemployment (UNE), inflation (INF), stock prices (STO); and previous GDP and forecast (FOC).