

Inferential Expectations

Gordon Douglas Menzies¹
University of Technology, Sydney
and Australian National University

Daniel John Zizzo²
BREB Unit, Department of Economics,
University of Oxford, Oxford OX1 3UQ, UK.

March 2004

Abstract

We propose that the formation of beliefs be treated as statistical hypothesis tests, and we label such beliefs *inferential expectations*. If a belief is overturned through the build-up of evidence, agents are assumed to switch to the rational expectation. Thus, rational expectations is a special case of inferential expectations if agents are unconcerned about mistakenly changing their beliefs (the test size α equals unity), or if there is so much information available about a parameter that it is known with certainty (the sampling distribution of the estimator collapses to a point). We present the results of an individual choice experiment showing preliminary support for inferential expectations in comparison to either rational expectations, or adaptive expectations with one degree of freedom. Least-squares estimates for α are less than unity for over 40% or over 60% of experimental participants. The impact of inferential expectations is illustrated by showing how it alters a simple model of the exchange rate and a Lucas supply function.

Keywords: expectations, macroeconomics, rationality.

JEL Classification Codes: C91, D84, E50, F31.

1. Introduction

Rational Expectations (RE) applies the principle of rational behavior to the acquisition and processing of information and to the formation of expectations (Maddock and Carter, 1982). Economic modellers and policy-makers who use RE as a working hypothesis bestow upon their representative agents the statistical prowess necessary to behave as if they were able to calculate mathematical expectations, and, when information is limited, to calculate unbiased and efficient parameter estimators.³

This theory has had a central role in macroeconomics since the 1970s. It appears in Lucas' (1972) 'Islands' model, Dornbusch's (1976) 'Overshooting' model, Hall's RE Permanent Income Hypothesis (1978), Real Business Cycle theory (e.g., Kydland and Prescott, 1982), and a host of new Keynesian models (e.g., Woodford, 1991). It continues to appear in models of recent financial crises (eg. Agenor et al., 1999).

Despite its influence, the number of alleged empirical failures of RE has built up over the passage of time. In the context of consumption behavior, 'near rational' departures from RE were used to explain the excess sensitivity of consumption to anticipated changes in income (Cochrane, 1989, following Akerlof and Yellen, 1985). Similarly, evidence against Uncovered Interest rate Parity (Frankel and Rose, 1995) led some writers to suggest near rational departures from RE (Gruen and Menzies, 1995). More recently, Mankiw (2000) has argued that it is not possible to reconcile RE with the combination of the prediction of a disinflationary boom in case of a fully credible disinflatory announcement, of inflation persistence and of the observed output dynamic in response to monetary policy shocks. His later work has tried to replace sticky prices with 'sticky information' adjustments (Mankiw and Reis, 2002, 2003).

Other negative evidence comes from variable-forecasting experimental studies. Although RE predictions are not rejected as null hypotheses in some contexts (see Dwyer et al., 1993), the most common outcome is that individuals do not hold RE (e.g., Schmalensee, 1976; Blomqvist, 1989; Camerer, 1995; Beckman and Downs, 1997; Swenson, 1997). In addition, experimental research often finds either under-utilization or over-utilization of priors (Camerer, 1995). Some writers, perplexed by the rapid change of beliefs in the 1997 Asian financial crises, also wandered from the hypothesis of RE, speculating instead that the

³ The name *rational expectations* emphasises the use of mathematical expectations. But any realistic theory of 'rational' belief formation must take account of parameter estimation. Over the 1990s there were 137 papers in the *American Economic Review* that undertook estimation (Ziliak and McCloskey, in press).

¹ Tel: +61-2-95147728; fax +61-2-95147711. Email address: gordon.menzies@uts.edu.au

² Tel: +44-1865-281481; fax: +44-1865-286581. Email address: daniel.zizzo@economics.ox.ac.uk

markets did not fully understand the mechanisms operating in their environment until unfolding events made them look more closely (Corbett and Vines, 1999).

In this paper we propose a simple alternative to RE for those environments where its validity is doubted. We suggest that belief formation be treated as a Neyman-Pearson hypothesis test, dubbed *Inferential Expectations* (IE). If a belief is overturned through the build-up of evidence, agents are assumed to switch to the RE of the variable. Thus, RE is a special case of IE if agents are unconcerned about mistakenly changing their beliefs (the test size α equals unity), or, if there is so much information available about a parameter that it is known with certainty (the sampling distribution of the estimator collapses to a point) leading to the rejection of any incorrect null.

The intuition is that economic agents hold beliefs that are subject to falsification by new information, in much the same way that they are in conventional statistical hypothesis testing. A change in beliefs thus requires new information that exceeds a threshold, modelled here by statistical significance. We assert that, at the individual level, beliefs about economic variables tend to be more subject to periods of inertia interspersed with occasional discrete shifts than what would be implied by rational expectations.

This is not a new idea in the philosophy of science (Kuhn, 1970), and it is consistent with the standard practice of most scientists, including many economists, who revise their theoretical beliefs only when classical hypotheses tests achieve statistical significance at conventional levels such as 0.05 or 0.01.⁴ That being so, beliefs of economists about economics tend to be subject to periods of inertia interspersed with occasional discrete shifts. IE potentially explains deviations from strict rationality while still imposing a plausible structure on expectations which is consistent with economists' own practices in statistical data analysis.⁵

One possible explanation for behavior consistent with IE is that, in many contexts, gathering and processing information to adjust beliefs may be costly relative to the marginal incentives to collect and process such information, in a way which is not very different from the adjustment of nominal prices.⁶ IE may provide a 'near rational' state-dependent rule to decide whether to make the effort. In the language of Gigerenzer et al. (1999), it may embody a 'fast and

⁴ This occurs despite the availability of an alternative approach based on Bayesian inference (e.g., Zellner, 1988). One of the authors has once had a paper rejected by a top tier journal because one of the main results was significant only at the 0.052 level.

⁵ It is also consistent with modelling practice in the Markov Switching literature, where passing a threshold, however tentatively, leads to a new regime (Hamilton 1989).

⁶ Romer (1996) makes this point in relation to the Lucas (1972) 'islands' model.

frugal heuristic' which may be rational in the presence of information-gathering and information-processing costs.

IE could be relevant for more sophisticated hypothesis testing, for example to detect the presence of autocorrelation, as in Rotheli (1998). Our approach is in spirit similar to Frydman and Goldberg's (1996, 2003) approach to expectations based on hypothesis testing over models.⁷ Their research program, tracing its roots back at least to an informal discussion by Rappaport (1985), is more radical than ours (they allow for multiple models, and imperfect information), but at the cost of structural indeterminacy and complexity. Another way of viewing IE modelling is as complementary to Mankiw and Reis (2002, 2003) 'sticky information' approach: IE at the micro level might imply sticky information at a macro level.

We intend to explore these research avenues in the future. However, this paper has the narrow remit of explaining the idea of IE, finding some preliminary experimental evidence for it, and showing its impact on theorizing with two stylized models.

The paper is organized as follows. In section 2 we provide a general framework for IE, and demonstrate the conditions under which IE becomes RE; we illustrate our analysis with a Bernoulli data generation process, which we apply to a stylized account of currency collapse in section 3. In section 4 we outline an individual choice experiment that provides preliminary evidence for IE. In section 5 we illustrate the simplicity of using IE as a modelling tool by introducing IE into the Lucas (1972) 'islands' model. Section 6 concludes.

2. Inferential Expectations

In all IE models, there is a *cognitive target* (the variable or parameter that is believed to be in one of two states, described by the null hypothesis H_0 and the alternative hypothesis H_1), a *signal* (a model variable that provides information about the cognitive target), and a *test statistic* and *rejection region* which are defined conventionally.

Let x be a parameter or random variable related in some way to a random variable Y . Granted some economic significance to x , agents form beliefs about it. Suppose that a data generating process for Y emits n stochastic signals S_i (for $i = 1, \dots, n$) which provide information about x . We assume that stochastic signals are independent random draws from Y . The expectational task facing agents is to arrive at beliefs about x based on the signals.

⁷ Our work can also be seen as related to Foster and Peyton Young's (in press) game-theoretical work on hypothesis testing by bounded-rational agents on their opponents' repeated games strategies.

The rational expectation is the mathematically best guess for x . The inferential expectation is the mathematically best guess for x , subject to a concern about changing beliefs (made operational by a Neyman-Pearson hypothesis test of size α), and incorporating any testing shortcuts that qualify as a ‘fast and frugal’ heuristic. When the concern about changing beliefs becomes vanishingly small ($\alpha \rightarrow 1$), IE and RE coincide.⁸ The *cognitive target* is x , the *signal* S_i , the *test statistic* some function of S_i , and the *rejection region* are the values of the test statistic that lead to a rejection of a Neyman-Pearson hypothesis test of size α .

The simplest application is when x is the true mean of Y . In this case, the Maximum Likelihood (and Least Squares) estimator, \bar{S} , is the RE, since it is the mathematically best guess for x . Under RE, beliefs about x evolve continuously as the estimate \bar{S} updates for every new signal. For example, a series of signals higher than the current sample mean will shift up the RE of x : see Figure 1.

(Insert Figure 1 about here)

If x is the true mean of Y , and IE hold, x is the *cognitive target*, and the draws S_i (for $i = 1, \dots, n$) are the *signals*. The *test statistic* is the sample mean \bar{S} of all the S_i , which forms the basis for hypothesis tests about x .

To define the *rejection region*, we need to start with an initial belief about x (the null hypothesis). Assume agents first form a belief about x upon receipt of the k th signal ($1 \leq k < n$), and this belief is \bar{s}_k where the subscript refers to the number of individual signals used in the calculation. In Figure 1, this is the value a . This belief is maintained as a null hypothesis about the cognitive target. This null hypothesis is tested against the data upon the receipt of each additional signal. The null hypothesis is not rejected until the test statistic (for example, the z value if a normal approximation is used) passes a critical value determined by a standard statistical hypothesis test of size α . In Figure 1, we assume that the hypothesis test is two-sided, defining the rejection region with probability $\alpha/2$ in each tail. Contingent on α and on the variance of the sampling distribution of \bar{S} (which is decreasing in n), this could take many signals. If the critical value is reached after i signals, the mean is assumed to jump to the RE estimate \bar{s}_i . In Figure 1, the RE and IE of x are both equal to c .

⁸ This is true regardless of any shortcuts used in the testing procedure. If the test size is unity, then hypothesis testing is suspended (along with any shortcuts about distributional assumptions, etc.) because a Neyman-Pearson hypothesis test minimizes the size of the probability of a type II error given the test size (probability of a type I error). That is, given a test size of unity, the best way to minimize the chance of falsely believing the null is to always reject it.

Under the assumption of IE, there is first ‘under-use’ of information (by comparison with RE) and then ‘over-use’ (when beliefs change). As indicated in the introduction, RE is nested in IE in two ways:

1. RE is IE for the special case when α equals unity. This is clear from the fact that α equals unity implies a rejection for any value of \bar{s} . That being so, the RE belief is constantly embraced.
2. RE is IE for the special case when x is a parameter and n equals infinity (and memory is unbounded). When the number of signals is very large, the sampling distribution of \bar{S}_n collapses to a single point at the true value. All nulls will be overturned (except if the null is correct), and RE beliefs embraced. In this limiting case, the RE belief will actually be the true parameter x .

IE requires the specification of a test size α and of a rejection region for the test given a given value of α . Standard classical inference theory (e.g., Hoel, 1984), and computational convenience - on the part both of modelling economists and presumably of economic agents -, can be used to determine the rejection region. One obvious shortcut to model rejection regions is to assume that agents use a Normal approximation.⁹ Another one is to devise a statistical test based on Chebyshev’s inequality.¹⁰ Chebyshev’s inequality says that:

$$P(|z| > k) \leq \frac{1}{k^2}$$

where z is a standardized random variable (the distance from the mean in units of standard deviations), and the weak inequality is relevant for a discrete random variable. If the probability of getting an observation more than k standard deviations away from the mean is less than $1/k^2$, we may set $1/k^2$ equal to α , and make a rare event statement.

For example, suppose a test of size $\alpha = 0.25$ is required. The above inequality says that the chance of getting an observation more than 2 standard deviations away from the mean is less than 25%. Therefore, if such an observation is observed, a rare event has occurred and the belief can be changed with a chance of making a mistake (probability of a type I error) no more than 25%.

The main advantage from employing Chebyshev’s inequality is that it just requires the computation of mean and variance, sidestepping the need for distributional assumptions, albeit at a loss of statistical power. Both shortcuts

⁹ Naturally, when the Central Limit Theorem holds this is not a shortcut.

¹⁰ On Chebyshev’s inequality, see for example Davidson and MacKinnon (1993).

can be considered as consistent with a view of IE as a fast and frugal heuristic (Gigerenzer et al., 1999), and either one may be easier to work with in theoretical applications. Chebyshev's inequality will be used in our illustrative Lucas (1972) model with IE in section 5, and both IE with Chebyshev's inequality and IE with a Normal approximation will be tested experimentally in section 4.

3. Inferential Expectations and Sudden Currency Movements

We now illustrate our analysis using a Bernoulli data generating process to provide a stylized account of a sudden currency movement. There are three reasons for doing so. First, it may make the analysis more concrete. Second, while purely suggestive, the example of sudden currency movements is of intrinsic interest, given the difficulties in explaining currency movements with RE.¹¹ Third, we use an almost identical set-up in relation to the experiment described in section 4.

Let x now be the probability that a currency is worthless tomorrow. Suppose a Bernoulli Data Generating Process emits n stochastic signals $S_i = 1, 0$ (for $i = 1, \dots, n$) about the value of x .¹² These could be Bloomberg reports, where 1 indicates that the currency will be worthless and 0 indicates that it will not. The average of the signals is now the proportion of ones, \hat{s} , and $E(\hat{s}) = x$. That is, if the probability of the currency being worthless is 0.2 then, on average, twenty per cent of Bloomberg reports indicate that it will be worthless.

In this application the *cognitive target* is x , the *signal* is S_i , the *test statistic* is \hat{s} , and the *rejection region* is defined as it was in section 2, namely using either the Normal approximation or Chebyshev's inequality.¹³

The Normal approximation method requires agents to reject the expectation \hat{s} , corresponding to the null hypothesis H_0 whenever the z statistic is higher in absolute terms than the critical value of z for some given test size α , i.e. z_α :

¹¹ One problem is the vast literature debunking Uncovered Interest Parity (Frankel and Rose, 1995). Another is that models of the Asian crisis struggle to reconcile RE with the apparent dearth of new information at the onset of the crisis. One response has been to speculate that agents did not understand some important relationships in the economy, but this really amounts to 'discarding RE' (for example, Corbett and Vines, 1999). Another (controversial) approach is to propose multiple equilibria (Krugman, 1999), but this raises the question of how one moves between equilibria.

¹² Being a Bernoulli, the sample mean of these signals is an unbiased estimator for the chance of success on a single draw i.e. $E(\bar{S}) = x$. Bondarenko and Bossaerts (2000) have a somewhat similar set-up, but the parameter to be inferred is not part of the Data Generating Process for the signals.

¹³ Strictly speaking, the rejection region could be determined by Binomial rejection regions for a small sample. However, unless the null belief of x is 0.5, the Binomial is not symmetric, implying no uniformly most powerful test. Given a loss function, a full-blown optimization exercise could uncover an optimal rejection region, but we prefer the testing shortcuts in the text as they seem more like a 'fast and frugal' heuristic.

$$\frac{\frac{|\hat{s}_i - \hat{s}_k|}{i}}{\sqrt{\frac{\hat{s}_i(1-\hat{s}_i)}{i}}} > |z_\alpha|$$

If agents instead employ Chebyshev's inequality $P(|z| > k) \leq 1/k^2$, for a sample proportion we have

$$P\left(\frac{|\hat{s}_i - \hat{s}_k|}{\sqrt{\hat{s}_i(1-\hat{s}_i)/i}} > k\right) \leq \frac{1}{k^2}$$

which, by setting $\alpha = 1/k^2$, becomes

$$P\left(\frac{|\hat{s}_i - \hat{s}_k|}{\sqrt{\hat{s}_i(1-\hat{s}_i)/i}} > \frac{1}{\sqrt{\alpha}}\right) \leq \alpha$$

leading agents to reject the expectation \hat{s}_i corresponding to the null hypothesis H_0 if

$$\frac{|\hat{s}_i - \hat{s}_k|}{\sqrt{\hat{s}_i(1-\hat{s}_i)/i}} > \frac{1}{\sqrt{\alpha}}$$

In either case, only once the critical value is reached the mean jumps to the RE estimate \hat{s}_i .

To complete a stylized model of currency collapse, let tomorrow be the end period. The transversality condition is that the exchange rate is either unity with probability $1-x$, or zero with probability x . Let interest rates in both economies be the same, and agents in both economies be risk neutral (so that they only care about expected returns). The transversality condition is that the end-period exchange rate equals $1-x$, and, by backward induction, the current value of the exchange rate must also be $1-x$.

Finally, suppose agents have IE about x based on a given sample. A shift in their beliefs about x (and hence in the current exchange rate) can occur simply because one piece of information – the proverbial straw that broke the camel's back – takes the sample proportion into the rejection region. If the null is overturned, IE agents change beliefs according to RE. If the RE belief about x much greater than the previously held belief, a currency collapse may occur as $1-x$ shifts down. In the case of the 1997 Asian crisis, there could have been one

small bit of information that, for a given rejection region, had the capacity to dramatically alter beliefs.¹⁴

4 Inferential Expectations and Experimental Evidence

4.1 Introduction

In this section we describe an individual choice experiment designed to test whether IE has significantly greater explanatory power than RE or, put it differently, whether there are subjects for which assuming that $\alpha < 1$ provides a better fit. The individual choice design was thought to be best suited to test the idea of IE in its cleanest form, i.e. without having to worry about the strategic considerations that would arise from a strategic or market setting.

The basic structure of the experiment closely mapped the analytical framework of sections 2 within an individual choice setting. There were two urns reflecting two possible states of the worlds, namely different combinations of white and orange balls. The true state of the world was chosen randomly, and subjects received signals about its nature by the means of random ball draws with replacement from the 'chosen urn'. The prior probability of an urn being chosen was 0.5 at the start of the experiment, but should have then evolved differently according to the observed sequence of white and orange balls being drawn and, importantly, according to different models of expectation formation. We next describe the experimental design in more detail, and then move to the experimental predictions and results.

4.2 Experimental Design

The experiment was run in at the School of Finance and Economics, University of Technology Sydney, in September 2003.¹⁵ Recruitment was through lecture announcements, posters, and UTS Online (a local forum for electronic notices). Recruits were predominantly, though not exclusively, undergraduate students.

¹⁴ Two other factors may also be part of an IE-based explanation of the Asian crisis. First, agents may have become more inclined to change beliefs for a given number of signals (α became larger, reflecting less concern about Type I errors). Accordingly, investors could have suddenly discarded their conservatism about negative information coming out of Asia as α increased. Second, agents may have received a large number of new signals that caused the variance of the sampling distribution to shrink, so that the previously believed nulls were overturned. A deluge of information could have built up as the Asian crisis played out (n increased) collapsing the sampling distributions and leading to the overturning of null beliefs. While both these explanations have some merit in describing developments towards the end of 1997, they are less compelling as an explanation of the *onset* of the crisis in the middle of that year. Radelet and Sachs (1998) argue that there was not much new information about Asia at the onset of the crisis, and, MacLeod (1998) makes the same point about Indonesia.

¹⁵ It was approved by the UTS Ethics Committee. The experimental instructions can be found in Appendix A.

There were six experimental sessions, three for each of the two experimental conditions; all sessions had six subjects except the last one, which had seven, for a total of 37 subjects. The experiment lasted about two hours, and paid an average of 31.42 Australian dollars (AUS \$).¹⁶ The experiment was in two stages, structurally unrelated to one another; in this paper we focus only on the first stage, which had six periods of fifteen rounds each and took over 75% of the session time.

At the start of the session subjects faced a table on the top of which there were two identical urns, a set of white and orange balls in a basket, and a screen. In the 0.7 condition, the experimenter (a) showed subjects that both urns were empty, (b) in front of the subjects, he took seven white balls and three orange balls and placed them in one of the two urns (Urn 1 in what follows) (c) and he took three white balls and seven orange balls and placed them in the other urn (Urn 2 in what follows); (d) he then hid both urns behind the screen. The 0.6 condition was identical to the 0.7 condition, except that Urn 1 got six white balls and four orange balls, and Urn 2 got four white balls and six orange balls.

At the start of each period subjects were reminded about the period number and then one of the two urns was randomly chosen by the flip of a coin in front of the subjects, and put on display. Let us label this urn the 'chosen urn'. It was made clear to the subjects that the probability of Urn 1 being chosen was 50% at the start of each period, but they were not told which urn had actually been chosen.

At the start of each round the experimenter drew a ball from the chosen urn, showed it to the subjects and then put it back in; subjects were asked to write down the ball color in correspondence to the correct period and round in their answer booklet, and then had to make a probability *guess*, between 0% and 100%, on how likely it was that the chosen urn was Urn 1. Subjects were told not to change choices made in previous draws.¹⁷

Once a period was completed, the following period got started with a new flip of the coin, up to the end of the 6th period. It was made clear to the subjects that the probability an urn was chosen was entirely independent of the probability that it had been chosen in previous periods. A questionnaire administered to the subjects at the start of the experiment ensured that this, and other key points, were clear.¹⁸

¹⁶ This is roughly equal to 25 US dollars.

¹⁷ We shall return to this point towards the end of section 4.3.

¹⁸ The experimenters gave clarifications to the subjects who got answers wrong on the questionnaire.

Payment was based on the guess made in a randomly chosen period and round picked at the end of the experiment. A standard quadratic scoring rule (e.g., Davis and Holt, 1993) was used in relation to this round to penalise incorrect answers: if the chosen urn was Urn 1, then subjects got $25 - 25 \times (\text{guess} - 1)^2$ AUS \$; if the chosen urn was Urn 2, then subject got $25 - 25 \times \text{guess}^2$ AUS \$. Subjects were provided with a payment table detailing the payment for each level of error, without need of any computation on their part (see Appendix A). There was also a participation fee of 8 AUS \$.

4.3 Experimental Predictions

Rational Expectations. The prior probability was set at 0.5. As information flowed in, RE (or, equivalently, IE with $\alpha = 1$) predicted straightforward Bayesian updating depending on whether white or orange balls were drawn.

Inferential Expectations. IE for $\alpha < 1$ would predict the value of 0.5 to be retained until the rejection region was reached according to the two-sided p -value from a binomial distribution with the Bernoulli probability of a white being 0.5 under the null. If and when the null hypothesis of 0.5 should be rejected, a new null hypothesis corresponding to the RE prediction in that round would be formed, and so on for the following periods. The appropriate test here is a two-sided test, and we use both the Normal approximation method and the Chebyshev's inequality method (see section 2) to determine the rejection region. The Normal approximation method requires agents to maintain the belief corresponding to the null hypothesis until signal i is received such that

$$\frac{|\hat{s}_i - \hat{s}_k|}{\sqrt{\frac{\hat{s}_i(1-\hat{s}_i)}{i}}} > |z_\alpha|$$

If the Chebyshev's inequality method is instead used, it will be recalled from section 3 that Chebyshev's inequality predicts rejection of H_0 if

$$\frac{|\hat{s}_i - \hat{s}_k|}{\sqrt{\hat{s}_i(1-\hat{s}_i)/i}} > \frac{1}{\sqrt{\alpha}}$$

In what follows we label IE_N the predictions of IE complemented with Normal approximation and IE_C the predictions of IE complemented with Chebyshev's inequality. In both cases we estimate the value of α corresponding to each experimental subject by using a least squares method, i.e. by minimizing the sum of squared errors between predictions and observations. That is, we

consider all the choices made across rounds and periods by each subject (90 in the full sample) and we find the *subject-specific* value of α that mimimizes the sum of squared differences between IE (IE_C or IE_N) and such choices. These α values will be those used in comparing the performance of IE against RE and AE. We also employ the least squares method to estimate α values that best fit each period as played by each subject. That is, we consider the 15 choices made by a given subject in a given period, and we find the *period-specific* value of α that mimimizes the sum of squared errors between predictions and observations. Thus, for each subject there are six period-specific values of α , and one can analyze whether these period-specific values followed any particular dynamic pattern in the experiment. These period-specific values of α are estimated only for the purpose of testing Hypothesis 3 below; whenever we do not specify otherwise, we shall be referring to subject-specific α instead.

Adaptive Expectations. As the IE model has one degree of freedom relative to RE, we also tested IE against another hypothesis on belief formation also with one degree of freedom, in the form of a traditional adaptive expectations (AE) model of the form:

$$\hat{s}_i = \hat{s}_{i-1} + \beta s_i$$

that is, the expectation about the value of the variable after receiving i signals sequentially is equal to the expectation after receiving $i - 1$ signals plus the β -weighted value of the current signal. $\beta \in [0, 1]$ provides the required degree of freedom. The AE model can be solved recursively from the initial value $s_0 = 0.5$ for any given value of β . We determine the value of β corresponding to each experimental subject by using the least squares method.

Experimental Hypothesis. We can compute the expectations profile for RE, AE, IE_N , and IE_C agents in relation to each session, using the sequence of observed ball draws and the procedures described so far. All our experimental hypotheses are designed to test the performance and robustness of IE.

HYPOTHESIS 1. IE_N and IE_C perform significantly better than RE and AE.

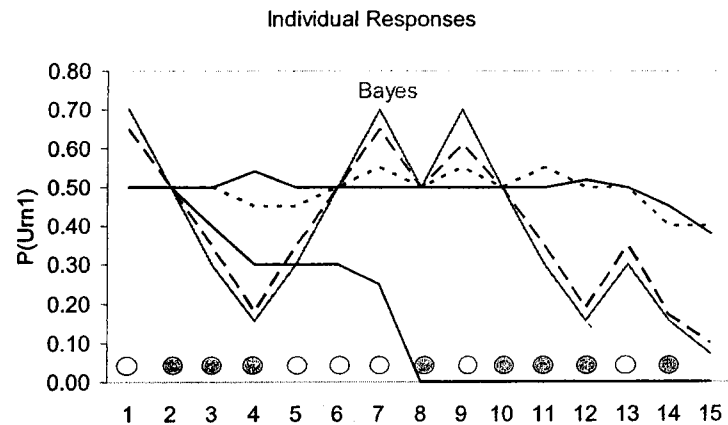
HYPOTHESIS 2. Mean α values do not significantly differ between the 0.6 and 0.7 conditions.

HYPOTHESIS 3. Mean period-specific α values tend to converge to 1 as the experiment progresses and the subjects have opportunities to learn about the nature of the task.

Hypothesis 2 is an obvious test of robustness in our α estimates to changes in the task they are estimated from. Hypothesis 3 is also a test of robustness, and aims to verify the absence of any obvious convergence towards greater RE play (i.e., IE with $\alpha = 1$) across the 90 rounds of the experiment. Relatedly, in testing Hypothesis 1 and 2 we use not only the ‘full’ sample from all six periods but also an ‘experienced’ sample which removes the observations from periods 1 and 2, thus allowing subjects to get some practice and experience about the nature of the task. We also considered a ‘restricted’ sample of observations where periods in which subjects altered their choices (notwithstanding our instructions to the contrary), and periods where some misperceptions occurred in the recording of the color of the balls, were removed.¹⁹ Overall, in order to check the robustness of our results, we considered four samples: the full sample, the experienced sample, the restricted sample, and the experienced restricted sample.

4.4 Experimental Results

The following statistical analysis asks which model of belief formation is best on a Least Squares criteria. However, it is worth noting that a number of respondents adopted a ‘hold-then-change-rapidly’ pattern, which is consistent with the IE stylization. The following four responses are taken from a particular group of draws; white balls are unshaded and orange balls are shaded.



¹⁹ A total of nine periods were removed in this way, six from choice alteration and three from apparent misperception. Five of the nine periods removed were in periods 1 or 2.

The selected profiles are not uncommon in the data; a number of respondents adopted a strategy of absorbing information, and then moving suddenly. This is important because it would be possible for IE to perform best on a Least Squares criteria, and yet be unconvincing as a micro-model for decision making. This would occur if all respondents changed beliefs with every draw of the ball, contrary to the basic intuition of a hypothesis test.

We now turn to the analysis of the least-squares alphas. Figure 2 provides histograms for the distribution of α for both IE_N and IE_C , in the various samples.

(Insert Figure 2 about here)

Mean α values were 0.636 for IE_N and 0.786 for IE_C . A non-negligible fraction of agents had $\alpha < 1$ in both cases: for example, in the experienced sample, 16 out of 37 subjects (0.432) seems to have employed $\alpha < 1$ for IE_C , a number rising to 25 out of 37 (0.676) for IE_N . The clear differences in the distributions of α s may suggest that, somewhat worryingly, IE_N and IE_C may bear little relation to one another. However, while there were differences in the distributions of the α s, Pearson r (IE_N , IE_C) is equal to 0.909, 0.814, 0.883 and 0.784 in the full, experienced, restricted, and experienced restricted samples respectively²⁰ ($P < 0.001$). Although IE_N values tend to be lower than IE_C values, they tend to follow each other closely.

Hypothesis 1 and related results. Figure 3 and Table 1 compare the mean sum of squares error, computed by period, of the various algorithms in the different samples.²¹ F tests can be used to test significance of the ratio between pairs: Table 1 illustrates the results for the various measures and samples.

(Insert Figure 3 and Table 1 about here)

Some results emerge:

- (1) both IE_N and IE_C outperform RE and AE, and robustly so in all samples;
- (2) we cannot reject the statement that IE_C and IE_N perform equally well;
- (3) we cannot reject the statement that RE and AE perform equally well in the experienced samples (whether restricted or not), but, once we move on to the full sample or the restricted sample, AE performs rather badly.

²⁰ Spearman ρ (IE_N , IE_C) is equal to 0.855, 0.820, 0.845 and 0.788 in the full, experienced, restricted, and experienced restricted samples respectively ($P < 0.001$).

²¹ We are forced to use mean sum of square errors computed by period for the F tests because of the non-independence of observations within each period. Appendix B contains details on the computation of the mean sum of square error by period, and notes how, while flawed, using mean sum of squares error computed by observation (i.e., the straightforward mean squared difference between each choice and the theoretical prediction) would only strengthen the case for Hypothesis 1.

However, the use of F tests is not ideal. The two groups of errors are highly correlated, making F tests not entirely appropriate; Table 2 exemplifies this correlation for the full sample.

(Insert Table 2 about here)

In the light of this, and as a further test of robustness, we also ran nonparametric sign tests comparing the goodness of fit of different algorithms in relation to each subject; the results of these tests are illustrated in Table 3.²²

(Insert Table 3 about here)

IE_N now outperforms IE_C: for example, in the experienced sample, out of 37 subjects IE_N performed worse than IE_C once, tied for twenty subjects, and performed better 16 times ($P = 0.001$, two-tailed). There is insufficient evidence for RE outperforming AE instead.

In relation to Hypothesis 1, because of the way that α is estimated, IE_N and IE_C will always outperform RE for any subject for which $\alpha < 1$.²³ The sign test enables us to see whether a *sufficiently large* number of subject has $\alpha < 1$ so as for the improvement in fit to be considered statistically significant. Table 3 shows that the answer is positive ($P < 0.001$). The following balance therefore emerges from the sign tests:

- (1) once again, both IE_N and IE_C outperform RE and AE across samples;²⁴
- (2) IE_N performs better than IE_C;
- (3) we cannot reject the statement that RE and AE perform equally well, in relation to all samples.

We conclude that the evidence from both parametric and nonparametric tests agrees in showing support for Hypothesis 1. An electronically available appendix²⁵ contains the mean choices and predictions according to each model of expectation formation by session and period; Figure 4 exemplifies the kind of aggregate dynamics observed by reproducing the graphs from session 4 (a 0.7 condition session).

²² There is no problem in determining goodness of fit for each subject by using mean sum of square errors by observation in relation to the sign tests, and so we do. However, results do not qualitatively change if they are computed by period in analogy to the F tests: see Appendix B for details.

²³ This is because $\alpha < 1$ implies that $\alpha = 1$ was not the coefficient minimizing the sum-of-squares error for a given subject.

²⁴ The comparison is unequivocal in the relationship between IE_N or IE_C and RE; in relation to AE, it is unequivocal for the full samples (restricted or not) while being less so ($P < 0.1$ in a one-tailed test) for the experienced samples.

²⁵ The appendix is at <http://www.economics.ox.ac.uk/Research/Ree/MZ/MenziesZizzoWebAppendix.pdf>.

(Insert Figure 4 about here)

In period 1 RE performs better in the first seven periods, but IE does better on average afterwards. In periods 2 and 6, with the exception of round 7, IE may be doing a better job in capturing the lower variability of choice relative to RE. In periods 3, 4 and 5 IE (especially IE_N for period 3) clearly does a better job at tracking mean choices than RE. Relative to RE, there appears to be a lower mean sensitivity of IE (with $\alpha < 1$) predictions to new information (though exceptions exist).

Hypothesis 2. As shown by Table 4, mean α values are stable between the two conditions, no matter the sample.

(Insert Table 4 about here)

For example, in the experienced sample, in relation to IE_C the mean α is 0.713 in the 0.6 condition and 0.764 in the 0.7 condition, with still greater stability in the full sample and the purified sample. Using t tests, we can never reject the hypothesis that mean α values differ between conditions. This is in contrast to the instability of β estimates in the calibration of the AE model.

Hypothesis 3. The fact that the evidence discussed so far is robust to the removal of the first two periods, as in the experienced sample, already suggests that IE may not tend to converge to RE as experience grows. This holds notwithstanding the fact that Figure 3 suggests a reduction in the noise in the data as one moves to the experienced sample. Figure 4 shows not only that no convergence to RE is observed, but also that if anything mean α values may tend to *decrease* with the progress of the experiment, especially in relation to IE_C, though the effect may be restricted to the initial acquisition of experience.

(Insert Figure 5 about here)

Nevertheless, while Pearson nor Spearman coefficients between period and either IE_C or IE_N are always negatively signed, they are never significant, ranging from -0.026 (Pearson r (*Period*, IE_N), full sample) to -0.109 (Pearson r (*Period*, IE_C), purified sample).

4.5 Experimental Evidence: Some Conclusions

The evidence we have presented is preliminary, being as it is from a small individual choice experiment. Nevertheless, we believe that the evidence is suggestive in a variety of ways. In a sample where AE does not fare well in

comparison to RE notwithstanding its extra degree of freedom, IE does perform significantly better. Over 40% had $\alpha < 1$ according to IE_C , and over 60% according to IE_N . Of course, these percentages would decrease if for the purpose of modeling parsimony we restricted the $\alpha < 1$ values to one of only a limited set of values, in the limit just one value. Even so, a non-negligible proportion of agents would still be classified as having $\alpha < 1$, even with IE_C : for example, if we allow only for one value of $\alpha < 1$, in the experienced and purified sample some 19% of the agents would still be classified with $\alpha < 1$ (specifically, with $\alpha = 0.25$).

We found that IE_N had the edge relative to IE_C in the nonparametric sign tests, though no statistically significant difference can be found using F tests; the sets of α estimates are highly correlated with one another. In any case, the design itself may have been biased *against* IE, since the only new thing happening between successive guesses was the provision of a piece of information, thereby potentially biasing subjects towards doing something with it in terms of their guesses.

5. Inferential Expectations and the Lucas Supply Curve

Notwithstanding the many incarnations of the Phillips curve, economists continue to believe in a short-run trade-off between inflation and unemployment (Mankiw, 2000). In this section we illustrate the impact of IE on a standard 'islands' set-up (Lucas, 1972). This is done purely to show how IE can be used as a modelling tool.

Log prices are decomposed into an average price p and a relative price.

$$p_i = p_i - p + p = r_i + p \quad p = \frac{\sum_i p_i}{n}$$

Using standard microfoundations, island labour supply is increasing in the inferred relative price, based on the revealed price p_i on island i . We make a simplifying assumption about the parameter in the labour supply function.

$$l_i = q_i = \frac{1}{\gamma - 1} E(r_i | p_i) = E(r_i | p_i), \text{ if } \gamma = 2$$

The problem is solved in Lucas (1972) by assuming a bivariate Normal distribution for the observed island price and the relative price, and finding the mathematical expectation for the relative price. The model was one of the first to

include Rational Expectations (RE) into a macroeconomic setting. The Lucas supply function appears as a result of aggregation across islands.

$$E(r_i | p_i) = \frac{V_r}{V_r + V_p} (p_i - E(p)) \quad (1)$$

$$y = \frac{\sum q_i}{n} = \frac{\sum l_i}{n} = \frac{V_r}{V_r + V_p} (p - E(p)) \quad (2)$$

Our point of departure comes in the expectations equation 1. We give the agents the same parameter information set.²⁶ However, we assume that discovering that the distribution is Normal and working out the signal extraction problem involves psychic (and/or pecuniary) costs that exceed the benefits. Therefore the agents resort to a fast and frugal heuristic, by conducting a preliminary hypothesis test: they take extreme values of the island price as evidence of a relative price change.

Informally, they do not use the information of the island price unless they have evidence that it has changed in an important way. It is as if they 'turn a blind eye' to the change in the island price unless it is large. Only if this is large do they bother to do the signal extraction problem.

Formally, the *cognitive target* is $p - E(p)$ (or p , since $E(p)$ is given), the *signal* is p_i , the *test statistic* is also p_i , and the *rejection region* is defined below.

The implicit null and alternative hypotheses are:

$$H_0: p_i = E(p), \text{ so } p_i - E(p) = 0 \\ H_1: p_i = p_i \text{ observed}, \text{ so } p_i - E(p) \neq 0.$$

One possible mechanism that justifies IE in this setting is that the benefits of obtaining information are positively correlated with the distance of the island price away from $E(p)$.²⁷

Aiming to find an information-cheap way to decide if it is worth calculating RE, we assume that agents use Chebyshev's inequality.

²⁶ We assume $E(p)$ is given as information, and does not have to be worked out.

²⁷ This is in keeping with the interpretation of 'sticky information' as analogous to 'sticky prices', with the costs of making an adjustment requiring to be compared to the marginal benefits, to which we already referred to in the introduction.

$$P\left(\left|\frac{p_i - E(p)}{\sqrt{\sigma_p^2 + \sigma_\eta^2}}\right| \geq k\right) \leq \frac{1}{k^2} \quad \text{and setting } \alpha = \frac{1}{k^2}$$

$$P\left(\left|\frac{p_i - E(p)}{\sqrt{\sigma_p^2 + \sigma_\eta^2}}\right| \geq \frac{1}{\sqrt{\alpha}}\right) \leq \alpha$$

That being the case, the rejection event R ('reject H_0 ') occurs when:

$$|p_i - E(p)| \geq \frac{\sqrt{\sigma_p^2 + \sigma_\eta^2}}{\alpha}$$

If they do reject, they then pay the costs of discovering the distribution and working out the signal extraction problem. This takes them back to the rational expectations solution.²⁸ Thus, when Inferential Expectations (IE) is applied to the expectations equation we obtain a modified Lucas supply curve.

$$IE(r_i | p_i) = \frac{V_r}{V_r + V_p} (\Phi_i - E(p))$$

$$\text{where } \Phi_i = \begin{cases} p_i & \text{if } R \quad (\text{i.e. } |p_i - E(p)| \geq \frac{\sqrt{\sigma_p^2 + \sigma_\eta^2}}{\alpha}) \\ E(p) & \text{if not } R \quad (\text{i.e. } |p_i - E(p)| < \frac{\sqrt{\sigma_p^2 + \sigma_\eta^2}}{\alpha}) \end{cases}$$

$$y = \frac{\sum q_i}{n} = \frac{\sum l_i}{n} = b(IE(p) - E(p)), \quad b = \frac{V_r}{V_r + V_p}$$

$$\text{where } IE(p) = \frac{\sum_{i=1}^n \Phi_i}{n}$$

The new supply function is the same as Lucas', except that $IE(p)$ has replaced the general price index p . The index $IE(p)$ replaces any statistically insignificant island price (i.e., in *not-R*) with the true-mean price.²⁹ The logic is simple; non-volatile prices are simply discounted by producers when setting output.

²⁸ The procedure is slightly different if the level of significance is unity. Given that a unit probability implies the null must be overturned, the RE solution is embraced, even if the rejection inequality evaluated with α equal to unity is not true.

²⁹ As an aside, this contrasts sharply with the practice of removing volatile items from price indexes. Here, the non-volatile items are removed.

The IE Lucas Supply Curve is a locus of aggregate-price aggregate-output pairs. For large (positive or negative) values of aggregate log price, the IE Supply curve will have less of an output response for a change in the aggregate price level, since $IE(p)$ usually lies between p and $E(p)$.³⁰

The IE Phillips curve will therefore be flatter for extreme values of p , giving less response in output for price changes; for values of p closer to $E(p)$, it is less clear what the function will look like (see Figure 6).

(Insert Figure 6 about here)

The interpretation is straightforward. Under IE, increases in prices in the neighbourhood of expected price do little for output because producers (who, in the monetary misperceptions story, are making incorrect output decisions) discount their relevance. However, beyond some point, producers start to take notice of the inflation and reductions in unemployment are possible. Monetary policy (or any aggregate demand policy) may have a 'band of inaction'.

We conclude this section with some comparative statics. The departures from the Lucas supply curve envisaged above become less pronounced as $IE(p)$ approaches p . This, in turn, occurs when the hypothesis test size α approaches unity. Furthermore, if some agents use IE and some RE (i.e. IE with $\alpha = 1$), $IE(p)$ approaches p as the proportion of rational agents rises.

To illustrate these implications of our model, we drew 200 Normal p_i samples of for 5 islands, and calculated the Lucas Supply function, together with the IE supply function. The sample was constructed so that $p \sim N(0, 0.25^2)$.

Initially, twenty percent of the agents (that is, one in five islands) had IE with a test size of 0.25. This is consistent with the calibration exercise referred to in section 4, where, if one assumes that agents use IE_C and forces α to have just one of two values, and one takes the experienced restricted sample as the most reliable, one obtains $\alpha = 0.25$ for 19% of the subjects. The results of this simulation are in the top graph of Figure 7, where y is the vertical axis and p is the horizontal axis.

(Insert Figure 7 about here)

³⁰ This intuition follows from the *sample* version of Chebyshev's inequality, which says that no more than $100/k^2$ per cent of the data can lie more than k *sample* standard deviations away from the *sample* mean. Since p is a sample average, the individual p_i s must be clustered around p , with the degree of clustering determined by Chebyshev's inequality. That being so, setting individual p_i draws equal to $E(p)$ will tend to drag $IE(p)$ closer to the true $E(p)$, since many of the individual p_i values are between p and $E(p)$. The simulations confirm this intuition.

The intuition outlined earlier is confirmed; for large values of p the IE points tend to have less response on y . Due to leverage, this implies that an OLS line through the IE points will be flatter.³¹ In the bottom graph of Figure 7 we increase the test size to 0.75. This must imply that nulls are overturned more easily for the given draw of data, and so there is less difference between the two lines.

Finally, we make 80 per cent of agents have an α equal to .25, and the other 20 per cent have RE, i.e. $\alpha = 1$.

(Insert Figure 8 about here)

Clearly, if most of the agents have IE, there is a considerable band of inaction for policymakers.³²

6. Conclusion

This paper has presented a new model of belief formation, which we labeled *inferential expectations*. The basic idea of inferential expectations is that beliefs are maintained or revised using a Neyman-Pearson hypothesis test. They are rejected in favour of the rational expectation prediction as the new null hypothesis only when the rejection region, determined by the test size α , is reached. This is congruent with the scientific practices of most scientists, including many economists, in forming and revising their beliefs in academic research: for example, the achievement of a 0.05 significance level is often assigned considerable relevance to be satisfied that a particular hypothesis is supported. It is also congruent with a view of decision-making as characterized by information-gathering and information-processing costs, and hence by the usefulness of fast and frugal heuristics.

In all IE models, there is a *cognitive target* (the variable or parameter that is believed to be in one of two states, described by the null and the alternative hypothesis), a *signal* (a model variable that provides information about the cognitive target), and a *test statistic* and *rejection region* which are defined conventionally. Rational expectations is a special case of inferential expectations

³¹ The impact on the slope of an individual OLS observation depends upon the distance from the mean of the independent variable; observations furthest away from the mean have the greatest impact.

³² We do not consider the other case when IE becomes RE. That is, when the number of pieces of information become infinite. This is so because the basic structure of the islands model prohibits it. Agents look at one piece of information only (the island price).

when the test size α is equal to 1, or when there is so much information available that the sampling distribution collapses to a point.

Our individual choice experiment showed a significant improvement of fit relative to rational expectations and a simple model of adaptive expectations with one degree of freedom. We used two ways of determining rejection regions, one based on a standard normal approximation and the other on Chebyshev's inequality. Both are closely related to one another, and, furthermore, α estimates are robust to the specific experimental condition subjects were faced with in different sessions. Across six periods of fifteen rounds each, there is no evidence suggesting a progressive convergence towards rational expectations. Depending on whether Chebyshev's inequality or a normal approximation is used, we find that respectively over 40% or over 60% of experimental participants used inferential expectations with $\alpha < 1$.

We illustrated the implications of inferential expectations with reference to currency collapse, offering a possible mechanism for the sudden currency movements in the 1997 Asian Crisis. The most parsimonious way in which IE may have explained the evolution of beliefs in the context of this crisis is that small bits of information may have acted as the proverbial straw that broke the camel's back, by bringing agents into the rejection region and hence to reject their null hypothesis forecast about the value of the exchange rate.

We also exemplified the use of inferential expectations as a modelling tool in the context of the Lucas (1972) islands model; we found that replacing rational expectations with inferential expectations causes monetary policy (or other aggregate demand policy) to have a 'band of inaction'.

We intend to develop IE further, both theoretically and empirically. The goal of this paper was simply to define IE, present experimental evidence in favour of it, and demonstrate its potential as a modelling device.

Acknowledgements

We wish to thank Malcolm Edey, Donald Hay, David Vines and Adrian Pagan for useful advice, Rebecca Chalmers for experimental assistance, Duncan Ford for Computer programming, and University of Technology Sydney for financial support. The usual disclaimer applies.

Appendix A. Experimental Instructions

Instructions for 0.7 Condition

Welcome to the experiment!

The experiment is divided into two parts, *Stage 1* and *Stage 2*. Your final winnings will be equal to the Stage 1 Payment, the Stage 2 Payment and a participation fee of 8 dollars. (All winnings will be rounded to the nearest 5 cents).

You are playing Stage 1 first. You can see two identical urns on the table, and a set of white and orange balls in a basket; you can also see a screen. The experimenter will shortly do the following:

- (a) show you that the urns are empty;
- (b) take seven white balls and three orange balls, and put them in one of the two urns; let us label this urn **Urn 1**;
- (c) take three white balls and seven orange balls, and put them in the other urn; let us label this urn **Urn 2**;
- (d) hide both urns behind the screen.

There are six *periods* in Stage 1. You have received an answer booklet with a sheet for each period.

At the start of each period, the experimenter announces the period number and writes it on the board. Then one of the two urns will be randomly chosen, by the flip of a coin, independently of what urns were chosen in previous periods. You will not be able to see whether this *chosen urn* is Urn 1 or Urn 2, but you will be asked to guess how likely you think it is that the chosen urn is Urn 1.

There are *sixteen draws* in each period. At the start of each draw the experimenter announces the draw number and writes it on the board. *In Draw 0*, which happens at the start of the period, your best probability guess that the chosen urn is Urn 1 would have to be 50%: this is because *at the start* of each period the chosen urn is picked randomly afresh. This Draw 0 probability guess has been printed into the answer booklet for you.

For Draws 1 through Draw 15 inclusive:

1. *first*, the experimenter draws a ball from the chosen urn and announces whether it is white or orange; *please write the ball colour on the answer sheet, in the line corresponding to the correct period and draw*; the experimenter then puts the ball back into the chosen urn;
2. *second*, you have to answer the following question: "how likely is it that the chosen urn is Urn 1? (Remember, Urn 1 is the urn with 7 white and 3 orange balls). Please choose a probability over the range 0% (definitely not) to 100% (definitely certain)"; *please put your guess in the line in the answer booklet corresponding to the correct period and draw*.

At the end of the period the experimenter hides the chosen urn again behind the screen. If you are in periods 1 through 5, you should move on to the answer sheet for the following period.

If you are in period 6, please wait until the sheets are collected and the material for Stage 2 is distributed.

Stage 1 Payment. It is important that you try to make your best probability guesses, both because it is important for the value of the experiment, and because your final winnings depend on it. At the end of the experiment the experimenter will randomly choose a winning draw to reward your performance. The experimenter will roll a die to choose the period, and pick randomly from a third urn (with balls numbered between 1 through 15) to choose the winning draw. Your Stage 1 Payment will depend on your choice in the draw corresponding to the number on the ball which has been picked. In relation to this draw, the experimenter will take your choice and compare it with the true chosen urn for that draw. If in the winning draw the chosen urn was Urn 1, then the correct probability of the chosen urn being Urn 1 is 100%; if the chosen urn was Urn 2, then the correct probability of the chosen urn being Urn 1 is 0%. Your Stage 1 Payment will then be equal to

$$25 - 25 \times (\text{guess} - \text{correct probability})^2$$

that is, to 25 dollars minus a penalty. The penalty will be equal to the square of the error, that is of difference between the guess and the correct probability, multiplied by 25. The Stage 1 Payment will be higher the more correct your guess is. The enclosed table provides Stage 1 Payment values corresponding to some possible error levels.

Please stay seated throughout the experiment. It is essential, for the scientific value of the experiment, that you (a) do **not** communicate in any way with other participants during the experiment; (b) do **not** change your guesses for previous draws. You are liable to be expelled from the experiment, and forfeit all winnings (including the participation fee), if you do not comply with these simple rules.

This is an individual choice experiment: your choices have no influence on the winnings of other participants, and similarly the choices of other participants have no influence on your winnings. If you have any question, please raise your hand until an experimenter comes close to you, and then ask with a low voice. This may be a good time to ask questions, but feel free to raise your hand to ask questions at any time.

Stage 1 Payment Table

$$\text{Payment} = 25 - 25 \times (\text{guess} - \text{correct probability})^2$$

Error	Stage 1 Payment	Error	Stage 1 Payment	Error	Stage 1 Payment
0%	25	34%	22.11	68%	13.44
1%	25	35%	21.94	69%	13.1
2%	24.99	36%	21.76	70%	12.75
3%	24.98	37%	21.58	71%	12.4
4%	24.96	38%	21.39	72%	12.04
5%	24.94	39%	21.2	73%	11.68
6%	24.91	40%	21	74%	11.31
7%	24.88	41%	20.8	75%	10.94
8%	24.84	42%	20.59	76%	10.56
9%	24.8	43%	20.38	77%	10.18
10%	24.75	44%	20.16	78%	9.79
11%	24.7	45%	19.94	79%	9.4
12%	24.64	46%	19.71	80%	9
13%	24.58	47%	19.48	81%	8.6
14%	24.51	48%	19.24	82%	8.19
15%	24.44	49%	19	83%	7.78
16%	24.36	50%	18.75	84%	7.36
17%	24.28	51%	18.5	85%	6.94
18%	24.19	52%	18.24	86%	6.51
19%	24.1	53%	17.98	87%	6.08
20%	24	54%	17.71	88%	5.64
21%	23.9	55%	17.44	89%	5.2
22%	23.79	56%	17.16	90%	4.75
23%	23.68	57%	16.88	91%	4.3
24%	23.56	58%	16.59	92%	3.84
25%	23.44	59%	16.3	93%	3.38
26%	23.31	60%	16	94%	2.91
27%	23.18	61%	15.7	95%	2.44
28%	23.04	62%	15.39	96%	1.96
29%	22.9	63%	15.08	97%	1.48
30%	22.75	64%	14.76	98%	0.99
31%	22.6	65%	14.44	99%	0.5
32%	22.44	66%	14.11	100%	0
33%	22.28	67%	13.78		

Answer Booklet: Content of the Sheet for Each Period

When the experimenter draws a ball, write down the colour of the drawn ball in the middle column (if you find it convenient, you can just write W for white and O for orange).

How likely is it that the chosen urn is Urn 1? (Remember, Urn 1 is the urn with 7 white and 3 orange balls). Please choose a probability over the range 0% (definitely not) to 100% (definitely certain). Write down your answer in the Probability Guess column.

Do not change probability guesses corresponding to previous draws. If you do, you are liable to be expelled from the experiment, and forfeit all winnings (including the participation fee).

If you discover that you have put your guesses in the wrong place (say, the wrong page or wrong row), please raise your hand.

Draw	Drawn Ball Colour	Your Probability Guess
0		50%
1		
2		
3		
4		
5		
6		
7		
8		
9		
10		
11		
12		
13		
14		
15		

Instructions for 0.6 Condition

These were identical to those for the 0.7 condition, except that 'six balls' ('6 balls') were replaced for 'seven balls' ('7 balls'), and 'four balls' ('4 balls') for 'three balls' ('3 balls').

Appendix B. Computation of mean sum of squares error and robustness analysis.

The mean sum of squares error (MSE) is equal to the sum of squares error (SSE) divided by the number of relevant datapoints. There are two procedures to compute the MSE.

MSE by observation. It is possible to compute, in relation to each observation, the mean squared difference between prediction and observation. Then, in relation to each subject making a prediction p_i^{actual} in round i and to each corresponding theoretical prediction p_i^{theory} , it is possible to compute:

$$MSE = \frac{SSE}{n} = \sum_{i=1}^n (p_i^{actual} - p_i^{theory})^2$$

n obviously depends on the experimental sample under consideration; it is always equal to 90 in the full sample and to 60 in the experienced sample, but these numbers can be slightly lower in the restricted samples. MSE by observation provides the natural measure in relation to which to estimate sign tests.

In addition, if this procedure is used to run F tests for the ratio of MSE values, we find increased support for Hypothesis 1. However, these F tests are substantively incorrect, because independence across observations in the same period does not hold, spuriously enhancing the significance of all the F ratios.

MSE by period. It is possible to compute, in relation to each period, the mean squared difference between predictions and observations. The relevant test statistic, in relation to each subject, here is

$$MSE = \frac{SSE}{n} = \sum_{i=1}^q \left(\sum_{j=1}^{r(i)} p_{ij}^{actual} - \sum_{j=1}^{r(i)} p_{ij}^{theory} \right)^2$$

where q is equal to the number of periods in the sample, i.e. either 6 (in the full sample and the restricted sample) or 4 (in the experienced and the experienced restricted sample), whereas r is normally equal to 15, i.e. to the number of rounds per period, which is usually the same regardless of the period. However, in the restricted samples, on occasion r can be less than 15 in specific periods.

It is possible to run sign tests using MSE values by period rather than by observation. The results are shown in Table 5, and essentially mirror those if MSE by observation is used (Table 3), as discussed in the main text.

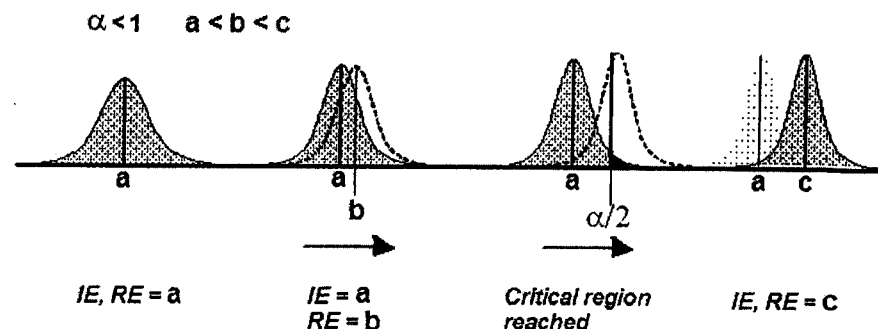
(Insert Table 5 about here)

There is stronger evidence for Hypothesis 1 than if MSE by observation is used: in Table 5 IE_N and IE_C always outperform RE at $P = 0.002$ or better. IE_N and IE_C also still outperform AE, whereas the evidence on RE vs. AE remains mixed. IE_N outperforms IE_C at the $P < 0.07$ level or better.

References

- Agenor, P., Miller, M., Vines, D., & A. Weber (1999) *The Asian Crisis: Causes, Contagion and Consequences*, Cambridge: Cambridge University Press.
- Akerlof, G. A. and J. Yellen (1985) Can Small Deviations from Rationality Make Significant Differences to Economic Equilibria?, *American Economic Review* 75, 708-721.
- Beckman, S. R. & D. Downs (1997) Forecasters as Imperfect Information Processors: Experimental and Survey Evidence, *Journal of Economic Behavior and Organization* 32, 89-100.
- Blomqvist, H. C. (1989) The 'Rationality' of Expectations: An Experimental Approach, *Journal of Economic Psychology* 10, 275-299.
- Bondarenko, O. & P. Bossaerts (2000) Expectations and Learning in Iowa, *Journal of Banking and Finance* 24, 1535-1555.
- Camerer, C. (1995) Individual Choice, in J.H. Kagel & A.E. Roth (eds.), *Handbook of Experimental Economics*, Princeton: Princeton University Press, 587-703.
- Cochrane, J.H. (1989) The Sensitivity of Tests of the Intertemporal Allocation of Consumption to Near-Rational Alternatives, *American Economic Review* 79, 319-337.
- Corbett, J. & D. Vines (1999) The Asian Crisis: Lessons from the Collapse of Financial Systems Exchange Rates and Macroeconomic Policy, in P. Agenor, M. Miller, D. Vines, & A. Weber (eds.), *The Asian Crisis: Causes, Contagion and Consequences*, Cambridge: Cambridge University Press, 67-111.
- Davidson, R. & J. G. MacKinnon (1993) *Estimation and Inference in Econometrics*, New York and Oxford: Oxford University Press.
- Davis, D. D. & C. A. Holt (1993) *Experimental Economics*, Princeton: Princeton University Press.
- Dornbusch, R. (1976) Expectations and Exchange Rate Dynamics, *Journal of Political Economy* 84, 1161-1176.
- Dwyer, G. P. Jr., Williams A. W., Battalio R. C. & T. I. Mason (1993) Tests of Rational Expectations in a Stark Setting, *Economic Journal* 103, 586-601.
- Foster, D.P. & H. Peyton Young (in press) Learning, Hypothesis Testing, and Nash Equilibrium, *Games and Economic Behavior*.
- Frankel, J. A. & A. K. Rose (1995) Empirical Research on Nominal Exchange Rates, in G. Grossman & K. Rogoff, *Handbook of International Economics*, Amsterdam: Elsevier, 1689-1729.
- Frydman, R. & M. D. Goldberg (2003) Imperfect Knowledge Expectations, Uncertainty-Adjusted Uncovered Interest Rate Parity, and Exchange Rate Dynamics, in P. Aghion, R. Frydman, J. Stiglitz & M. Woodford (eds.), *Knowledge, Information and Expectations in Modern Macroeconomics*, Princeton: Princeton University Press, 145-182.
- Gigerenzer, G., P. M. Todd & the ABC Research Group (1999) *Simple Heuristics That Make Us Smart*, Oxford: Oxford University Press.
- Goldberg, M. D. & R. Frydman (1996), Imperfect Knowledge and Behavior in the Foreign Exchange market, *Economic Journal* 106, 869-893.
- Gruen, D. & G. D. Menzies (1995) Forward Discount Bias: Is it Near Rationality in the Foreign Exchange Market?, *Economic Record* 71, 157-166.

- Hall, R. (1978) Stochastic Implications of the Life Cycle Permanent Income Hypothesis: Theory and Evidence, *Journal of Political Economy* 85, 971-987.
- Hamilton, J. (1989) A New Approach to the Economic Analysis of Non-stationary Time Series Subject to Changes in Regime, *Econometrica* 57, 357-84.
- Hoel, P. G. (1984) *Introduction to Mathematical Statistics*, 3rd Ed., New York: Wiley.
- Kuhn, T. S. (1970) *The Structure of Scientific Revolutions*, 2nd Ed., Chicago: University of Chicago Press.
- Krugman, P. (1999) Balance Sheets, the Transfer Problem and Financial Crises, *International Finance and Financial Crises; Essays in honour of Robert P. Flood Jr.*
- Kydland, F. & E. C. Prescott (1982) Time to Build and Aggregate Fluctuations, *Econometrica* 50, 1345-1370.
- Lucas, R. E., Jr. (1972) Expectations and the Neutrality of Money, *Journal of Economic Theory* 4, 103-124.
- Maddock, R. & R. Carter (1982) A Child's Guide to Rational Expectations, *Journal of Economic Literature* 20, 39-51.
- Mankiw, N. G. (2000) The Inexorable and Mysterious Tradeoff Between Inflation and Unemployment, HIER Discussion Paper 1905, Harvard University.
- Mankiw, N. G. & R. Reis (2002) Sticky Information versus Sticky Prices: A Proposal to Replace the New Keynesian Phillips Curve, *Quarterly Journal of Economics* 117, 1295-1328.
- Mankiw, N. G. & R. Reis (2003) Sticky Information: A Model of Monetary Nonneutrality and Structural Slumps, in P. Aghion, R. Frydman, J. Stiglitz & M. Woodford (eds.), *Knowledge, Information and Expectations in Modern Macroeconomics*, Princeton: Princeton University Press, 64-86.
- MacLeod, R. (1998) Indonesia, in R. MacLeod & R. Garnaut (eds.), *East Asia in Crisis: from being a miracle to needing one?*, London: Routledge, 31-48.
- Radelet, S. and J. Sachs (1998) The Onset of the East Asian Financial Crisis, NBER Currency Crisis Conference, 6-7 February 1998.
- Rappoport, P. (1985) Unfalsified Expectations: An Alternative Perspective on Modelling Expectations in Macroeconomics, C.V. Starr Center for Applied Economics research report 85-16, New York University.
- Romer, D. (1996) *Advanced Macroeconomics*, 2nd Ed., New York: McGraw Hill.
- Rotheli, T. E. (1998) Pattern Recognition and Procedurally Rational Expectations, *Journal of Economic Behavior and Organization* 37, 71-90.
- Schmalensee, R. (1976) An Experimental Study of Expectation Formation, *Econometrica* 44, 17-41.
- Swenson, C.W. (1997) Rational Expectations and Tax Policy: Experimental Market Evidence, *Journal of Economic Behavior and Organization* 32, 433-455.
- Woodford, M. (1991) Self-Fulfilling Expectations and Fluctuations in Aggregate Demand, in N. G. Mankiw & D. Romer (eds.), *New-Keynesian Economics*, vol. 2, Cambridge (MA): MIT Press, 77-110.
- Zellner, A. (1988) Bayesian Analysis in Econometrics, *Journal of Econometrics* 37, 27-50.
- Ziliak, S. T. & D. McCloskey (in press) Size Matters: The Standard Error of Regressions in the American Economic Review, *Journal of Socio-Economics*.

Figure 1. Example of IE dynamics (for $\alpha < 1$).

The example assumes that parcels of information having values above c (where $a < b < c$) flow in sequentially. The IE ($\alpha < 1$) agent sticks with her IE = a belief until the rejection region is reached. When the rejection region is reached, she switches to the RE belief c .

Figure 2. Histograms of α values for IE.

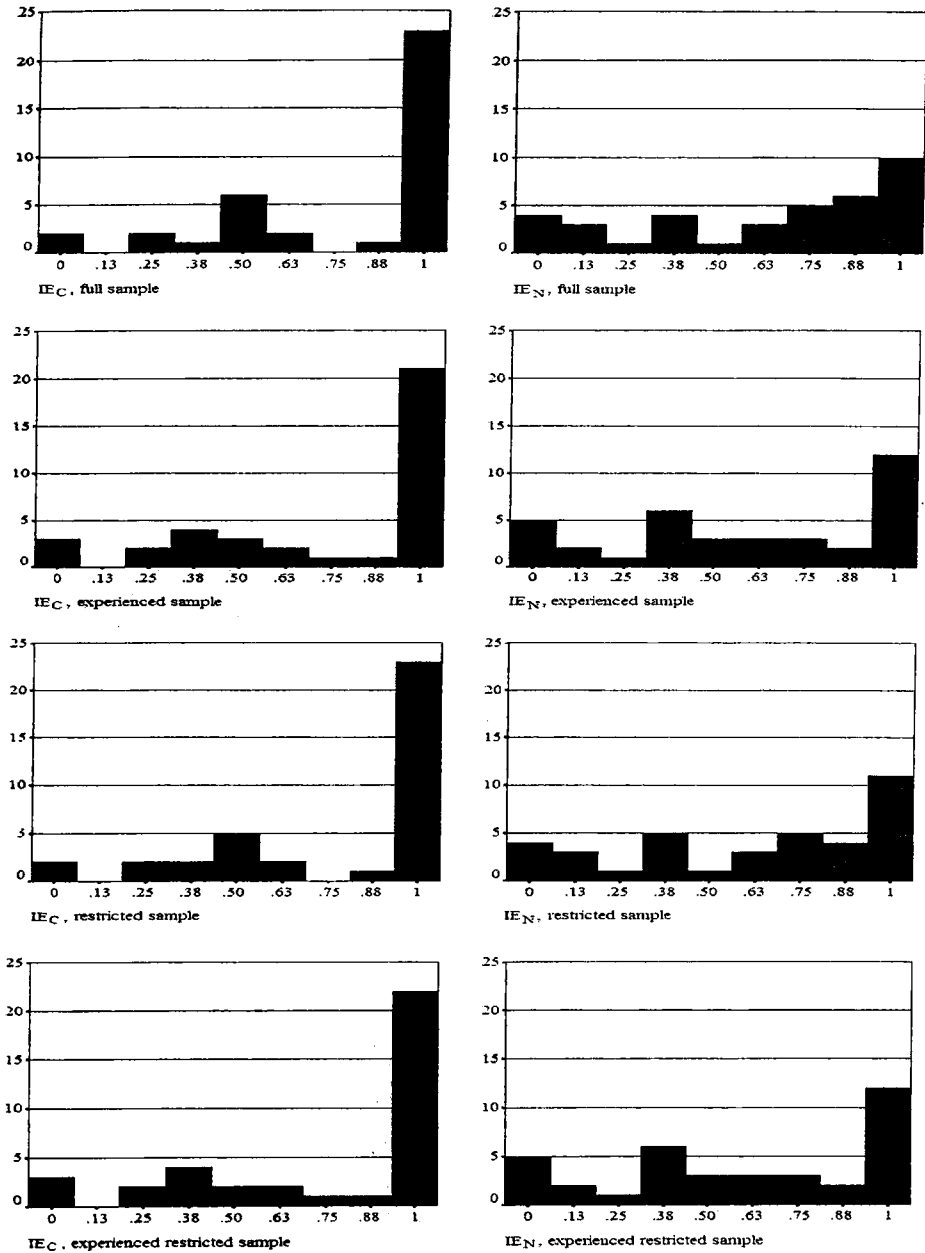
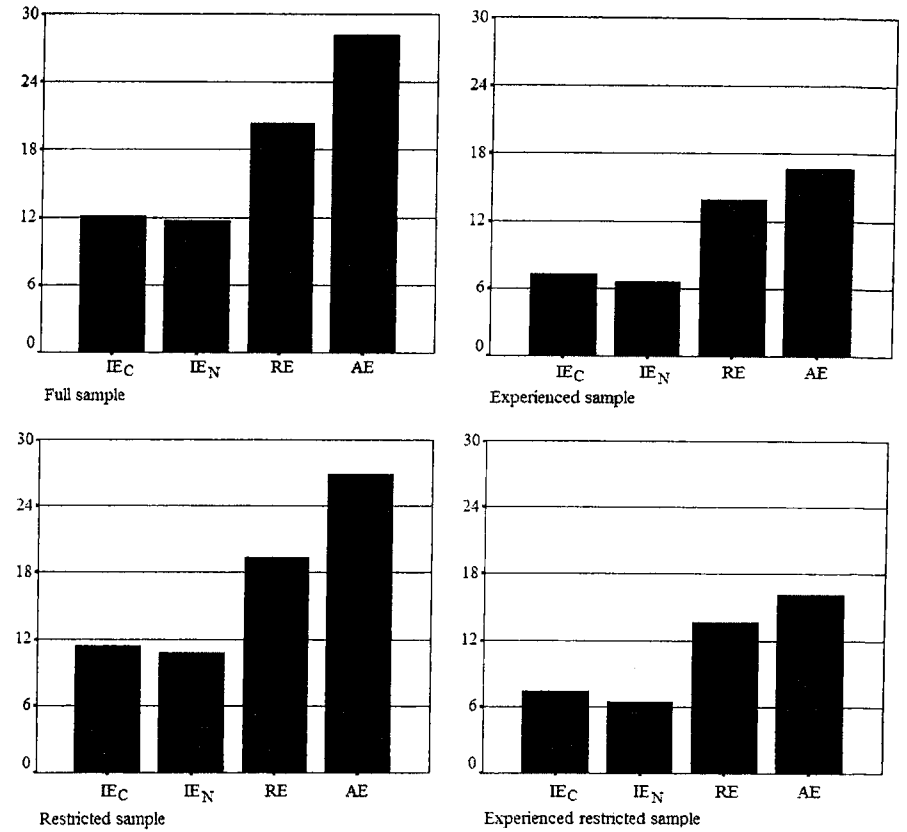


Figure 3. Goodness of fit of belief formation models: mean squared error.



Mean sum of square error (MSE) values are computed by period (see Appendix B).

Figure 4. Full sample mean choices and predictions: examples.

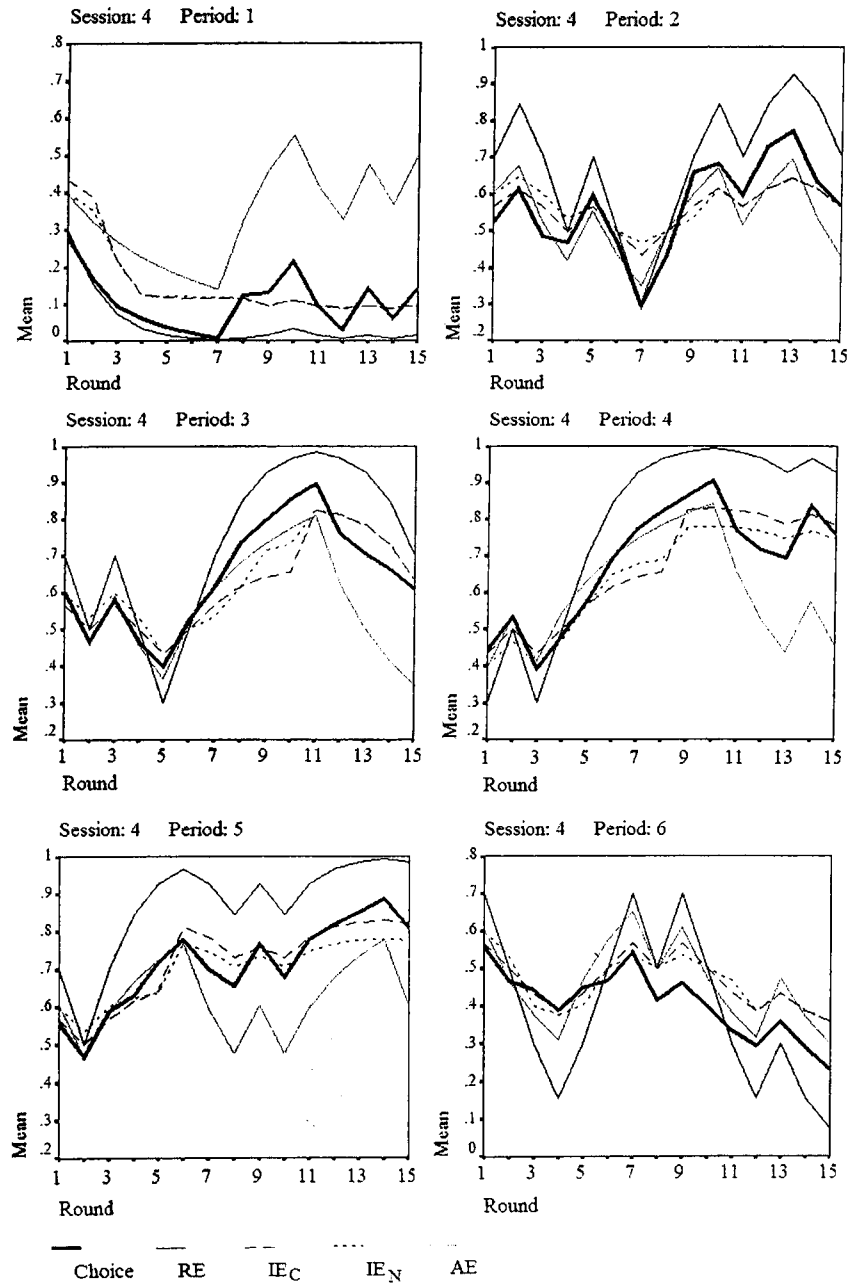


Figure 5. α estimates by experimental period.

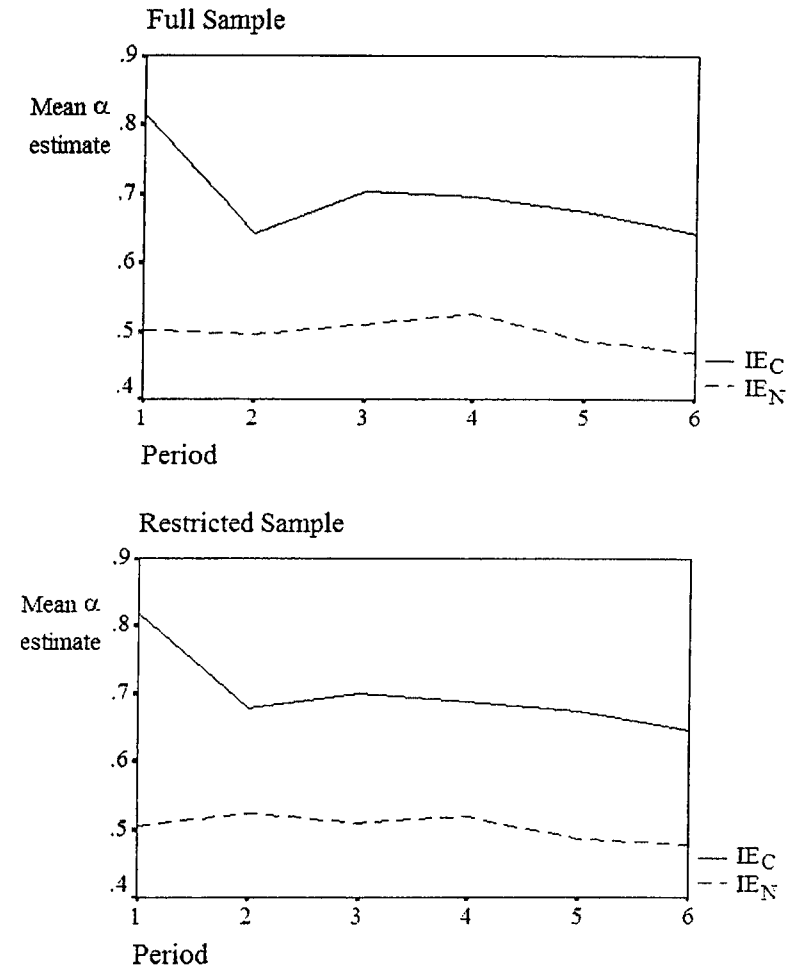


Figure 6. Comparison of Lucas with the IE Supply curve

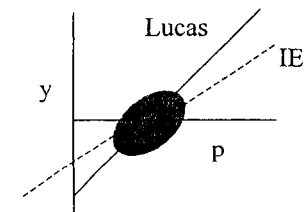


Figure 7. Computer simulations with 20% IE agents.

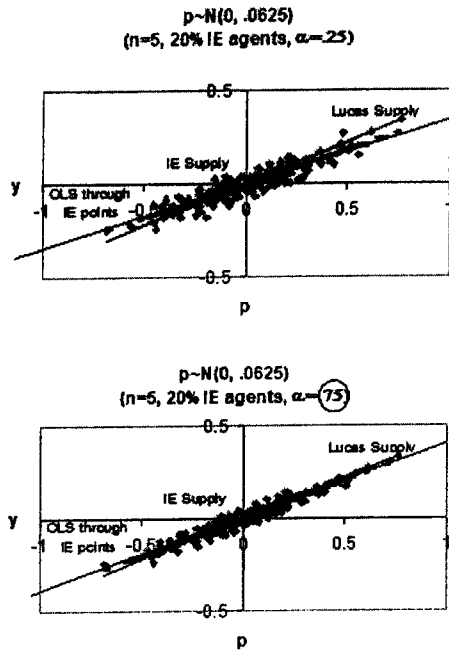


Figure 8. Computer simulation with 80% IE agents.

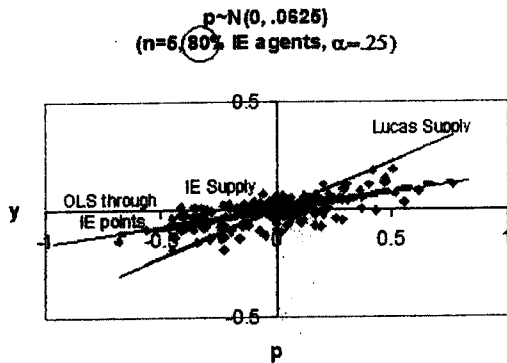


Table 1. Goodness of fit of expectational models.

Full Sample					
	SSE	RE	IE _C	IE _N	AE
RE	754.606	1.000	1.685***	1.744***	0.724
IE _C	447.907	0.594	1.000	1.035	0.430
IE _N	432.810	0.574	0.966	1.000	0.415
AE	1041.742	1.381**	2.326***	2.407***	1.000
Experienced Sample					
	SSE	RE	IE _C	IE _N	AE
RE	514.228	1.000	1.904***	2.111***	0.834
IE _C	270.045	0.525	1.000	1.108	0.438
IE _N	243.646	0.474	0.902	1.000	0.395
AE	616.705	1.199	2.284***	2.531***	1.000
Restricted Sample					
	SSE	RE	IE _C	IE _N	AE
RE	714.673	1.000	1.698***	1.803***	0.719
IE _C	420.789	0.589	1.000	1.061	0.423
IE _N	396.434	0.555	0.942	1.000	0.399
AE	994.479	1.392**	2.363***	2.509***	1.000
Experienced Restricted Sample					
	SSE	RE	IE _C	IE _N	AE
RE	502.767	1.000	1.84***	2.128***	0.844
IE _C	273.247	0.543	1.000	1.156	0.459
IE _N	236.310	0.470	0.865	1.000	0.397
AE	595.631	1.185	2.18***	2.521***	1.000

The first data column contains the sum-of-squares error (SSE) in relation to each expectation model. The values in the 2nd to 5th column are equal to $F = a / b$, where a is the mean SSE (MSE) corresponding to the row expectation model and b is the mean SSE (MSE) corresponding to the column expectation model. MSE values are computed by period: see Appendix B. ** : significant at the 0.01 level; ***: significant at the 0.001 level.

Table 2. Correlation matrix between mean square errors.

MSE	Pearson r			Spearman ρ		
	RE	IE _C	IE _N	RE	IE _C	IE _N
IE _C	0.88***			0.798***		
IE _N	0.897***	0.994***		0.883***	0.965***	
AE	0.369*	0.64***	0.634***	0.086	0.374*	0.342*

Correlation coefficients between row and column mean sum-of-squares error. *: significant at the 0.05 level; **: significant at the 0.01 level; ***: significant at the 0.001 level.

Table 3. Goodness of fit of expectational models using sign tests.

		Full Sample				Experienced Sample			
x	y	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail P	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail P
IE _C	IE _N	2	17	18	0	1	20	16	0
IE _C	RE	14	23	0	0	16	21	0	0
IE _C	AE	26	0	11	0.021	23	0	14	0.188
IE _N	RE	26	11	0	0	25	12	0	0
IE _N	AE	26	0	11	0.021	23	0	14	0.188
RE	AE	24	0	13	0.1	22	0	15	0.324

		Restricted Sample				Restricted Experienced Sample			
x	y	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail P	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail P
IE _C	IE _N	1	17	19	0	1	20	16	0
IE _C	RE	14	23	0	0	15	22	0	0
IE _C	AE	26	0	11	0.021	23	0	14	0.188
IE _N	RE	26	11	0	0	25	12	0	0
IE _N	AE	26	0	11	0.021	23	0	14	0.188
RE	AE	24	0	13	0.1	22	0	15	0.324

MSE: mean sum of squares error (computed by observation: see Appendix B; Table 5 contains the corresponding analysis in relation to MSE by period). P values are approximated to three decimal places.

Table 4. Mean α and β value by condition.

	Full Sample		Experienced Sample	
	Condition		Condition	
Mean α or β	0.6	0.7	0.6	0.7
IE _C mean α	0.793	0.78	0.713	0.764
IE _N mean α	0.653	0.619	0.545	0.628
AE mean β	0.136	0.226	0.121	0.229
	Restricted Sample		Restricted Experienced Sample	
Mean α or β	0.6	0.7	0.6	0.7
IE _C mean α	0.78	0.78	0.713	0.79
IE _N mean α	0.616	0.626	0.545	0.628
AE mean β	0.136	0.226	0.121	0.23

The table contains the mean of the α (β) values estimated for each subject and sample in relation to IE_C and IE_N (AE), by experimental condition. † tests estimate the significance of the difference between each pair of α or β estimates. **: significant at the 0.01 level.

Table 3. Further sign tests of the goodness of fit of expectational models.

		Full Sample				Experienced Sample			
x	y	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail P	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	P
IE _C	IE _N	7	13	17	0.064	4	17	16	0.12
IE _C	RE	13	23	1	0.002	15	21	1	0.001
IE _C	AE	28	0	9	0.003	27	0	10	0.009
IE _N	RE	27	7	3	0	26	9	2	0
IE _N	AE	28	0	9	0.003	26	0	11	0.021
RE	AE	25	0	12	0.049	21	0	16	0.511
		Restricted Sample				Restricted Experienced Sample			
x	y	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail P	MSE(x)<MSE(y)	MSE(x)=MSE(y)	MSE(x)>MSE(y)	2-tail P
IE _C	IE _N	6	14	17	0.035	4	17	16	0.012
IE _C	RE	13	23	1	0.002	14	22	1	0.001
IE _C	AE	28	0	9	0.003	27	0	10	0.009
IE _N	RE	26	8	3	0	26	9	2	0
IE _N	AE	28	0	9	0.003	26	0	11	0.021
RE	AE	25	0	12	0.049	21	0	16	0.511

MSE: mean sum of squares error by period. *P* values are approximated to three decimal places.